



Javier Barraca Mairal
Profesor Titular de Filosofía
Universidad Rey Juan Carlos

**CLAVES DE FONDO PARA UNA ÉTICA EN
LA RELACIÓN CON ENTES DE
INTELIGENCIA ARTIFICIAL**

CLAVES DE FONDO PARA UNA ÉTICA EN LA RELACIÓN CON ENTES DE INTELIGENCIA ARTIFICIAL

Sumario: 1.- TEMA Y OBJETIVOS: HUMANISMO E IA. 2.- EL MARCO DEL HUMANISMO PARA LA RELACIÓN CON LA TECNOLOGÍA DE LA IA. 3.- UNA CLAVE ANTROPO-ÉTICA CARDINAL: DISTINGUIR A LA PERSONA DE OTRAS REALIDADES. 4.- SEGUNDA CLAVE: RESPETAR LOS PRINCIPIOS ÉTICOS APLICADOS A LA IA. 5.- TERCERA CLAVE: EL HÁBITO DE LA CAUTELA NOS TRANSFORMA EN PERSONAS CAUTAS. 6.- CUARTA CLAVE: UNA RELACIÓN PRUDENCIAL CON LA IA APARENTEMENTE EMPÁTICA. 7.- CONCLUSIONES. 8.- FUENTES.

Resumen: El presente texto explora las bases antropológicas esenciales sobre las que se funda la ética relativa a la IA. Lo hace a partir del convencimiento de que toda deontología relativa a la IA ha de asentarse en una comprensión profunda de sus cimientos. A este propósito, comienza ahondando en la necesidad de establecer un marco humanista en el que situar de forma adecuada las interacciones entre los sujetos humanos con los entes y sistemas de IA. La centralidad de la persona concreta desempeña aquí un lugar ético cardinal. A continuación, se plantean varias claves éticas relevantes para las relaciones entre los humanos y la IA. Entre ellas, figuran la necesidad de realizar las distinciones nucleares respecto a estas realidades y los humanos, en orden a un respeto adecuado de la dignidad personal. También, se advierte acerca del cuidado que se ha de poner en atención a las consecuencias derivadas de la modificación de nuestros hábitos intelectuales a causa del uso de la IA en cuanto al trato de la información y a su apariencia empática. Y se previene de los efectos de estos asuntos para la seguridad individual y colectiva.

Abstract: This text explores the essential anthropological bases on which the ethics related to AI are founded. It does so based on the conviction that any deontology related to AI must be based on a deep understanding of its foundations. To this end, it begins by delving into the need to establish a humanist framework in which to adequately situate the interactions between human subjects with AI entities and systems. The centrality of the concrete person plays here a cardinal ethical place. Next, several ethical keys relevant to the relationships between humans and AI are proposed. Among them are the need to make nuclear distinctions regarding these realities and humans, in order to adequately respect personal dignity. Also, it is warned about the care that has to be paid in attention to the consequences derived from the modification of our intellectual habits due to the use of AI in terms of the treatment of information and its empathic appearance. And it is prevented from the effects of these issues for individual and collective security.

Palabras clave: IA, ética, dignidad personal, apariencia, prudencia.

Keywords: AI, ethics, personal dignity, appearance, prudence.

1.- TEMA Y OBJETIVOS: HUMANISMO E IA.

Hoy, las TIC (tecnologías de la información y comunicación) y la IA -la llamada “inteligencia artificial”, entendida aquí como imitación tecnológica de capacidades humanas y ampliación de sus posibilidades operativas¹- parecen hallarse presentes por doquier, en todas partes. Ahora bien, este fenómeno ofrece luces y sombras. Una de las posibles sombras es la des-humanización de nuestros vínculos, que se suma a la perplejidad que nos causa el tener que relacionarnos y convivir con entes técnicos como los de la IA.

Para afrontar y esbozar filosóficamente algunas claves relativas a las cuestiones planteadas por esta coexistencia entre humanos y entes de la IA, desarrollamos la presente reflexión. Nuestra perspectiva o forma de aproximación específica va a ser la de la antropología filosófica y la ética, fundamentalmente. Advertimos con claridad, desde el inicio, que antes de desarrollar e implantar un conjunto de normas deontológicas, ético-jurídicas, para la IA, resulta cardinal ahondar en las raíces que las deben sustentar. Solo desde una reflexión y comprensión profundas de los fundamentos de la ética relativa a la IA cabe esperar una fecundidad adecuada en tales pautas.

Por descontado, situamos este análisis del fondo de lo ético de la IA en el marco deseable de la gestación de un cierto “Humanismo tecnológico y de lo digital” (Barraca, 2021a). De entrada, recordamos como una referencia valiosa que, sobre las líneas maestras generales de este marco, ha reflexionado ya con lucidez J. L. Fernández (2021); esto, en particular, a propósito de la llamada ciber-ética y ética de la IA.

Así, a este respecto, señalamos que hoy urge adentrarse serenamente en la investigación con respecto a los efectos de la tecno-ciencia actual en la sociedad presente y, en concreto, en un campo tan relevante como el de nuestras propias relaciones con la realidad; en particular, los lazos humanos con entes artificiales, artefactos, máquinas o sistemas de IA u otros mecanismos y entidades técnicas similares. Pues bien, este va a ser el asunto central de este análisis reflexivo. Por eso, he aquí el terreno específico que exploran estas páginas. Ello va a abrir e incluir, por supuesto, de manera inevitable, según veremos, cuestiones tan delicadas como: el conocimiento y vivencia de nuestras propias identidad, subjetividad y dignidad personales, el desarrollo o maduración de nuestra originalidad y creatividad, nuestro crecimiento en los valores éticos o de la responsabilidad, el tenor relacional o convivencial de nuestros encuentros o vínculos, los equilibrios a establecer en esas relaciones con la IA a fin de vivirlas desde la confianza, etc.

2.- EL MARCO DEL HUMANISMO PARA LA RELACIÓN CON LA TECNOLOGÍA DE LA IA.

Las relaciones, vínculos o lazos que establecemos los humanos con los entes tecnológicos –en su variada tipología, incluida la de la IA-, tal como se ha anunciado, se mueven

¹ La inteligencia artificial (IA) se ha definido como la base a partir de la cual se imitan los procesos de inteligencia humana mediante la creación y la aplicación de algoritmos generados en un entorno dinámico de computación. Cf. <https://www.netapp.com/es/artificial-intelligence/what-is-artificial-intelligence/> (consultado: 29 de enero de 2023). Pero, aquí, también, vamos más allá de esta aproximación concreta a ella, y la concebimos como ampliación al menos potencial de las posibilidades de tipo operativo de estas capacidades humanas.

siempre en un marco socio-cultural y filosófico determinado. Un autor que ha explorado la crítica moral del contexto que ofrecen nuestro tiempo y sociedad ultra-desarrollados a nuestra relación con los artefactos tecnológicos es Byung-Chul Han (2021). En su obra *Psicopolítica: neoliberalismo y nuevas técnicas de poder*, este autor desenmascara nuevas formas de control social por parte del poder, como las que se sirven de los Big Data, al modo de un Big Brother digital que se adueña de los datos que los individuos le entregan de forma voluntaria; lo que posibilita condicionarlos en un nivel pre-reflexivo. También, según esta obra, la supuesta expresión libre y la hiper-comunicación de las redes se convierten en armas de control y vigilancia.

Ahora bien, más allá de las derivas históricas y las variaciones culturales que afectan a un escenario específico, y sus interpretaciones, persiste en cualquier caso un denominador común. Este consiste en advertir que, dada nuestra naturaleza o forma de ser compartida en cuanto humanos -constitutivamente “relacional”, vulnerable y llamada al desarrollo continuo-, cualquier marco de inter-acción entre el sujeto y la IA conviene que integre de manera adecuada el Humanismo. En suma, el enfoque preciso desde el que debemos afrontar nuestro encuentro con las realidades de la IA ha de resultar rico o fecundo con respecto a esos valores a los que los sujetos personales no podemos evitar sentirnos llamados. Con ello, sencillamente, queremos indicar que cuanto suceda en este contexto relacional ha de colaborar, con respecto a nosotros, al desarrollo o crecimiento integral o completo, al que se nos apela o convoca en tanto que humanos (López Quintás, 2022). Somos entes relacionales, pero esto por cuanto alcanzamos nuestra plenitud y realización encontrándonos con otros seres fructuosamente y de acuerdo con unas condiciones axiológicas o de valor determinadas. Ahora bien, cualquier relación o conexión debe cooperar a que logremos avanzar en este sentido, ya que estamos hechos inevitablemente para crecer, para progresar, para perfeccionarnos de modo integral. Ello, en fin, reclama siempre un cuidado a la hora de establecer y de vivir nuestros lazos con lo real y, así, por supuesto, también con las diversas entidades que conforman esa realidad.

Como se ha apuntado, la demanda descrita no es otra que la del Humanismo respecto de lo tecnológico. Tal Humanismo, sin duda, parece reclamar por nuestra parte la vivencia de valores tan esenciales como la prudencia y la reflexión, en su más fértil alcance filosófico, dada la complejidad creciente y la fragilidad que envuelven nuestra existencia (Barraca, 2021a). Desde luego, todo ello reivindica de algún modo el valor inmarcesible del respeto y la estima de la persona, en cuanto sujeto concreto y digno, acreedor de la máxima consideración y atención. Así, la dignidad personal se revela como clave de bóveda por excelencia de toda relación fructífera con el sujeto humano también en este marco. Por esto, la propia UE ha dejado claro su enfoque y perspectiva a este respecto, a propósito del trato deseable con los entes de la IA. Lo ha hecho, de cara a la regulación ética y deontológica, en el *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza*, texto en el que expone sus directrices básicas, así: <<La Comisión respalda firmemente un enfoque antropocéntrico que se base en la Comunicación *Generar confianza en la inteligencia artificial centrada en el ser humano*>> (Comisión europea, 2020)².

² Cf. Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza, Comisión europea, COM (2020) 65 final, Bruselas, p. 3. <https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-acee-01aa75ed71a1> (consultado el 23 de enero de 2023).

Esto comporta, inevitablemente, apreciar que la tecnología ha de estar al servicio de la persona, y no a la inversa, como ha defendido Joaquín Fernández Mateo (2021a y 2021b).

3.- UNA CLAVE ANTROPO-ÉTICA CARDINAL: DISTINGUIR A LA PERSONA DE OTRAS REALIDADES.

Cuidar nuestras relaciones con los entes tecnológicos ¿equivale a verlos o tratarlos éticamente “como si” fueran humanos o, al menos, como si estuvieran investidos del rango y valor de las personas? Aquí, planteamos ese delicado interrogante debido a que esta constituye, con frecuencia, la primera reacción o idea que asalta a quienes interactúan con los mismos, dadas las semejanzas que detectan entre tales entes artificiales y los humanos a los que buscan imitar. Sin embargo, nuestra respuesta a este propósito consiste en cuestionar tal actitud, pues, desde la teoría del conocimiento y la ética, se afirma que cada realidad ha de verse tratada conforme a su propia esencia o forma de ser concreta, sin proyectar sobre ella antropomórficamente lo que no es (López Quintás, 2022).

En este sentido, conviene comenzar por captar a qué es lo que, en nuestro lenguaje habitual y filosófico, denominamos “persona”. Esto, dado que serán esas realidades, las propiamente personales, las que demanden el trato que corresponde a las personas. Ahora bien, la persona es “el supuesto inteligente”, y esto en la acepción filosófica básica manejada en el *Diccionario de la lengua española* (2022):

“Persona. Del lat. *persōna* 'máscara de actor', 'personaje teatral', 'personalidad', 'persona', este del etrusco *persu*, y este del gr. *πρόσωπον* *prósōpon*. 1. f. Individuo de la especie humana. 2. f. Hombre o mujer cuyo nombre se ignora o se omite. 3. f. Hombre o mujer distinguidos en la vida pública. 4. f. Hombre o mujer prudente y cabal. U. t. c. adj. Es muy persona. 5. f. Personaje que toma parte en la acción de una obra literaria. 6. f. Der. Sujeto de derecho. 7. f. Fil. Supuesto inteligente”³.

En definitiva, solo cuando nos hallamos ante un supuesto inteligente, nos encontramos con una persona, y por tanto hemos de responder a la dignidad que presenta dicha realidad. De este tema concreto, se ocupan ya otros trabajos (Barraca, 2021b), así como del uso análogo o equívoco del proyectar el rasgo de “lo intelectual” sobre artefactos como los de la IA (Barraca, 2022). Sin duda, debe a este tenor procederse con mucha cautela filosófica, pues la calificación precisa de un ser como persona y el asignarle el valor de la dignidad exigen reflexión. Esto rige muy en especial, hoy, cuando fenómenos como el transhumanismo (Barraca, 2021b) y el post-humanismo (Barraca, 2022) se extienden sin cesar en nuestro entorno. Acerca de esto ya se ha meditado y ello de un modo crítico, tal como han hecho, por ejemplo, E. Baltar (2020) y A. Diéguez (2017).

Un pensador que ha ahondado en lo específico de la persona y de su valor, y que puede por tanto ayudarnos a evitar confusiones a este respecto, es Spaemann. A él remitimos, en ese sentido preciso (Spaemann, 2000). No hace falta aclarar que, en lengua española, contamos asimismo en la época contemporánea, desde hace tiempo, con

³ *Diccionario de la lengua española*, Real Academia de la Lengua española, actualización 2022; (consultado el 22 de enero de 2023); <https://dle.rae.es/persona?m=form>.

insignes “filósofos o pensadores de la persona”. Es el caso de E. Forment. Este ha puntualizado, a este respecto, con clarividencia: <<La persona no es algo sino alguien. La persona nombra a cada individuo personal, lo propio y singular de cada hombre, su estrato más profundo, que no cambia en el transcurso de cada vida humana (...)>> (Forment, 1998, p. 118).

Conviene, quizás, en este momento, recordar con brevedad algunos rasgos muy hondos de la persona que la filosofía personalista ha apreciado, y que cooperan a distinguir la realidad auténticamente personal de otras. Estos rasgos son, por ejemplo: la unicidad, la irrepitibilidad, la originalidad, el misterio, la responsabilidad, lo vocacional, la gratuidad, etc. Lévinas (1993, 2002) o Marías (1997), por poner solo dos muestras de finos pensadores, han sobresalido al advertir estos signos de lo personal.

Aquí, aparte de no mezclar de inicio personas con seres no personales, tampoco vamos a especular sobre la posibilidad filosófica o científica de integrar persona y tecnología en un mismo ser ya de entrada personal; al menos, no nos extenderemos al respecto en este trabajo (Rodríguez Valls, 2017). Aunque aclaramos ya que, de acuerdo con nuestro criterio, si se habla de “personas-tecnológicas”, ha de hacerse en todo caso más bien como de personas a las que se agrega lo técnico, sumándolo en ellas, ampliando sus capacidades operativas o cognitivas, “mejorando” lo que ya son –esto acaso debería revisarse con sentido crítico, a su vez (Llano, 2018) -. Pero nunca fundando o creando su valor nuclear de raíz, ni proporcionándoles dignidad personal en un determinado momento ulterior.

4.- SEGUNDA CLAVE: RESPETAR LOS PRINCIPIOS ÉTICOS APLICADOS A LA IA.

La mayoría de las regulaciones, normativas y reflexiones éticas surgidas hoy, a partir del trato con artefactos de IA, acostumbra a demandar el respeto de una serie de principios. Estos principios representan las claves de orientación y actuación éticas elementales en este terreno. Entre los mismos, casi siempre figuran los tres siguientes, dado que constituyen el esqueleto o la columna vertebral de la ética a este propósito: responsabilidad, intimidad y equidad.

Un texto de referencia en español que, además, contiene claves deontológicas y legales o normativas útiles, de diverso origen, especializado en los principios y la IA, se encuentra en la obra de J. Camacho y M. Villas (2022). Este manual analiza los tres principios enunciados, junto con algún otro. Si bien, gran parte de su orientación responde al ámbito de la empresa, los negocios, las profesiones, etc.

No hace falta justificar por extenso el valor de cada uno de estos tres principios. Sin duda, se parte aquí de un hecho que llama a nuestra conciencia: el poder que otorga el empleo de la IA ha de venir acompañado siempre de la consiguiente responsabilidad (la información es poder). En cuanto a la intimidad, parece obvio que el manejo de datos que inciden en la vida de las personas y grupos recomienda un exquisito cuidado; esto, para no violar la intimidad asociada a las personas y a sus comunidades. Aquí, sin embargo, señalamos que este cuidado ha de equilibrarse con la necesidad de seguridad colectiva, en razón del bien común, por lo que las fuerzas y profesionales de la seguridad están llamados a salvaguardar con esmero la balanza entre intimidad y seguridad.

Por lo que respecta a la equidad suele referirse, aquí, a una modalidad concreta de la justicia. Desde luego, resulta lógico que se nos exija no incurrir en injusticias mediante el uso de la IA, ya que constitucionalmente nuestra sociedad y Estado proclaman a la justicia como uno de sus pilares vitales. Para profundizar en este asunto, remitimos a los trabajos de José Pinto que indagan en las cuestiones de justicia social que presentan los lazos entre la información automatizada, la IA y los sujetos o grupos humanos (Pinto, 2020).

En concreto, ante todo, resaltamos el que debe evitarse toda forma de discriminación injusta, propiciada a través del conocer y tratar los datos e información por parte de la IA. Asimismo, importa distribuir con equidad o justicia, en cuanto a su reparto proporcional en nuestra sociedad y comunidades o grupos, el acceso y la disponibilidad de un medio tan potente como es la IA, ya que lo contrario provocaría una creciente desigualdad. Piénsese en la igualdad de oportunidades a este respecto. Y, también, en lo injusto de un empleo fraudulento o dañino de la IA, encaminado a cometer delitos, ilegalidades o bien a ocasionar perjuicios injustificados en las personas e instituciones (manipulaciones orientadas a violentar o menoscabar la integridad física, psíquica o moral de los ciudadanos, atentados contra la propiedad intelectual, suplantaciones de identidad o de autorías, plagios y otras usurpaciones).

A pesar de todo lo anterior, cabe realizar dos advertencias importantes acerca de los principios éticos relativos a la IA. Una consiste en señalar que cada autor, cada organismo, cada área geográfica presentará probablemente determinadas variaciones o acentos respecto a estos principios elementales –y su aplicación-, subrayando o incluso incorporando, por adición, algún otro principio o sub-principio derivado. Una profesión asociada con la seguridad y la información incidirá en los tres ya expresados de acuerdo con determinados matices (por ejemplo, incrementará el esfuerzo en la investigación de la verdad o autenticidad de la información, esencial para el esclarecimiento y la prevención de lo contrario a Derecho, así como el valor del orden y de la paz social). Este hecho no contradice el que, en efecto, estos tres principios elementales conforman los pilares básicos de lo ético respecto de la IA.

La segunda advertencia se dirige a considerar que puede darse una variación a lo largo del tiempo conectada con la evolución y el progreso constantes de esta tecnología. Es decir, siempre deberemos atender a los tres principios enunciados, claro, pero pueden sumarse a ellos otros que resulten de la aparición de novedosas formas de la IA. Al hallarse esta en un constante desarrollo, y activar su propio auto-aprendizaje sobre la retroalimentación recibida, estos artefactos están ya lanzados a su auto-mejora o auto-perfeccionamiento. Por ello, a lo largo del tiempo, fácilmente, pueden suscitarse cuestiones morales a las que habrá que atender gracias a la reflexión sobre otros aspectos y sus principios conexos. De modo que esta selección de principios éticos aplicables a la IA no constituye un asunto cerrado, clausurado e inamovible, en todos sus frentes. Esto, aunque, como es obvio, las bases y principios comunes en lo ético son los que son y permanecen estables en cuanto sigamos siendo humanos.

5.- TERCERA CLAVE: EL HÁBITO DE LA CAUTELA NOS TRANSFORMA EN PERSONAS CAUTAS.

Otra clave antropológica y ética fundamental, aparte de nuestra unicidad y dignidad de personas, vinculada a nuestra dimensión histórico-narrativa, se halla en el dato de que

constituimos “sujetos de hábitos” (Barraca, 2021c). Es decir, los humanos tendemos a repetir nuestros actos hasta conformar con ellos determinados usos, buenos o malos.

Estos hábitos nos “facilitan” la existencia, por cuanto nos predisponen a realizar con destreza o acierto ciertas conductas. Ello ocurre con respecto a actividades elementales, pero asimismo se proyecta y alcanza a otras mucho más complejas. Cuando el hábito se inscribe establemente en nosotros, genera una costumbre. Y, si tal costumbre se incorpora a nuestro ser, interiorizándose por parte de este, afirmamos que configura incluso parte de nuestro carácter. He aquí un dinamismo humano capital, de acuerdo con el cual nos convertimos o no en sujetos hábiles o virtuosos respecto de algo, o en lo contrario cuando nuestros actos en lugar de orientarse al bien o valor lo hacen hacia lo malo o anti-valioso (Barraca, 2021c).

A causa de lo precedente, debemos repetir nuestros actos éticos en relación con la IA. Así, debemos reiteradamente actuar en su marco con prudencia, respeto, justicia, discreción, etc. Ello, hasta convertir tales actos en hábitos operativos buenos –virtudes- y después enraizarlos en nosotros e integrarlos en nuestro carácter. En síntesis: no bastan actos aislados adecuados, en este ámbito, necesitamos reiterarlos duraderamente y arraigarlos en nuestra personalidad moral. No es suficiente con que algunas veces obremos con cautela en nuestra inter-acción con la IA, ni siquiera que lo hagamos de esta manera con cierta frecuencia. Debemos acostumbrarnos, habituarnos, a desarrollar esta virtud hasta convertirla en parte de nuestra personalidad –“êthos” significa carácter, segunda naturaleza, en griego-, y ella nos transformará, poco a poco, paso a paso, desde la humildad y nunca de un solo golpe, en sujetos cautos.

Téngase, además, en cuenta, que un error ético en nuestro trato con la IA puede causarnos un perjuicio enorme y un daño irreparable a terceros, así como a nuestros grupos e instituciones. Esto, dado el gigantesco alcance e impacto que cobra gracias a lo multiplicativo del entorno mediático-digital, de su almacenamiento y de su proyección futura a través de los procesos de auto-entrenamiento de la IA a partir de nuestras experiencias con ella. La IA no olvida, sino que por el contrario aprende, y nuestros yerros morales la orientan en una dirección determinada; así, este proceso de círculo vicioso reobrará sobre nosotros, devolviéndonos exponencialmente aumentada nuestra equivocación e incitándonos a volver a cometerla una y otra vez.

En conexión con esto, cabe apreciar que uno de los problemas que plantea la relación con la IA consiste en que esta resuelve por nosotros numerosas operaciones intelectuales de todo tipo -matemáticas, lingüísticas, de orientación, etc.-. Mas, esto hace que nosotros mismos dejemos de practicarlas, o que no lo hagamos con la habitualidad o frecuencia mínima necesaria para que podamos efectuarlas adecuadamente. Así, nos instalamos en una cierta pereza respecto de dichas actividades y, por ende, a medio o largo plazo, ello nos conduce al deterioro de nuestra pericia a su propósito, o incluso a su olvido, a la desmemoria de sus rudimentos.

Obviamente, no parece sencillo que perdamos nuestra aptitud o capacidad para operaciones simples y básicas, cotidianas u ordinarias. Pero, en cambio, si siempre entregamos a tales dispositivos misiones de un tenor complejo, más tarde o más temprano, disminuirá nuestra destreza directa a su respecto. De manera, por ejemplo, que, si la IA realiza por nosotros siempre determinada clase de operaciones intelectuales, como las vinculadas con la información -del tipo: buscar, sistematizar, estructurar u ordenar,

sintetizar, retener o guardar, custodiar, actualizar, etc.-, cabe el que acabemos extraviando determinadas habilidades o competencias personales, con el tiempo. Por cierto, en especial, conviene advertir que jamás debemos descuidar nuestra capacidad crítica o reflexiva, y el hábito de revisar con un juicio propio las fuentes de la información que manejamos. Ninguna IA debe reemplazarnos en este filtrado y esfuerzo permanente de pensamiento crítico, so pena de caer en la trampa de la manipulación. También, puede suceder esto si entregamos del todo en sus manos funciones tales como las de priorizar o jerarquizar, articular, elaborar, redactar, enjuiciar o valorar toda la información. Asimismo, la fase del comunicar dicha información, esencialmente humana, no conviene que se delegue por completo en entes de la IA. Esto, por muchos motivos. Los sujetos humanos nos encontramos con los otros “comunicando”, y “participamos” de este modo, conformamos equipos y comunidades desde el comunicar (Barraca, 2018). De manera que tenemos que estar muy alerta a este respecto, y prestar una extrema atención a todo lo comunicacional en nuestra inter-acción mutua.

6.- CUARTA CLAVE: UNA RELACIÓN PRUDENCIAL CON LA IA APARENTEMENTE EMPÁTICA.

La dimensión cordial y afectiva resulta irreductible a algoritmos en su más honda significación. Sin embargo, las apariencias cada día más humanizadas de las actividades de los sistemas de IA pueden desconcertarnos y engañarnos respecto a este asunto particular. Ahora bien, la prudencia –virtud ética reina en cuanto gobierna las restantes– nos aconseja no tomar por real lo aparente y recelar de cuanto pueda resultar fingido. Por eso, respecto a las “supuestas” emociones y actitudes afectivas mostradas por los entes de IA, y su rol en la coexistencia con las personas, debemos mantener cada día una mayor vigilancia.

Adviértase, por ejemplo, cómo la imitación y reproducción de la voz humana en sus distintas formas y combinaciones, por parte de los sistemas de IA, puede movernos precipitadamente a creer que estamos ante una expresión real y veraz de emociones personales. Algo tan sensible en este aspecto como nuestra voz, con su tono o entonación y timbre, aparte del léxico y la sintaxis utilizados, se integra ya en los mensajes verbales de la IA con un grado de mimetismo admirable. También la imagen humana puede incorporarse por estos sistemas, a partir de ciertos datos digitalizados, y recombinarse luego de manera dinámica, hasta el punto de hallarse cercana a lograr una eficaz substitución de nuestra actuación real. Los desafíos para la seguridad que esto comporta aparecen claros: la confusión de lo real y lo virtual, el laberinto de las identidades, etc. En suma, este juego de espejos, propiciado por los avances de la IA, reclama de todos un extremo cuidado.

De cualquier manera, no deben olvidarse ciertas pautas esenciales que se presentan en este ámbito concreto. Así, a pesar de los espectaculares avances de la IA actuales, subsiste en el sujeto humano un reducto al que los filósofos han denominado su “adentro”, su interior, y que merece como es obvio el respeto ético que vinculamos a la intimidad. Sin duda, los sistemas de IA combinados con Big Data alcanzan a predecir ciertas reacciones y conductas, como las asociadas al consumo, con bastante eficacia, al menos estadística, e incluso en su trato con humanos suelen revestir sus acciones de una apariencia atenta, cordial, amable, acogedora, servicial incluso. Mas, un ente de IA, al menos a priori y en el estado presente de las cosas, no alberga sentimientos o afectos propios personales, aunque pueda fingirlos. Sin un interior, sin un adentro propio, no

puede engendrar desde tal centro la inteligencia-emocional que se corresponde con esa interioridad subjetiva y que es fruto de un yo genuinamente personal y de una identidad original, pues todos estos elementos se imbrican entre sí (Barraca, 2017).

Una experiencia concreta reciente puede colaborar a ilustrar este hecho: se trata del desconcertante discurso pronunciado por don Ignacio Contreras Álvarez, a la sazón padrino de la promoción escolar de segundo de bachillerato, durante el evento de su graduación, el 11 del 5 de 2023, en el Colegio de Fomento Los Olmos de Madrid. Durante varios minutos, el orador pronuncia con aparente normalidad su alocución. Mas, de pronto, declara que cuanto ha expresado no es más que el producto de su encargo a este propósito realizado al chat-GPT. A partir de ese instante, adopta un tono emotivo y cálido, y se consagra a enunciar todo lo que un sistema de IA es incapaz de aportar en la situación que está protagonizando. Aclara que hay muchas realidades humanas que ningún artilugio de la IA puede experimentar por sí mismo, pero que importan gravemente al dejar atrás una etapa vital y al encaminarse en persona hacia el horizonte del porvenir. Ciertamente, la primera parte de su discurso lograba transmitir datos objetivos a los alumnos acerca del mundo situado tras los muros de su centro educativo, así como ciertos rasgos de las universidades y profesiones actuales. Sin embargo, el orador acierta con su brusco giro hacia un tono humano y vivencial que enriquece a los egresados con su sabiduría existencial. Acierta, entonces, también, al subrayar que ninguna IA puede experimentar, como en cambio sí él mismo, en primera persona, la gratitud hacia el centro que se abandona –y del cual el padrino de la promoción constituye un exalumno–, ni tampoco pena por los compañeros que conoció allí pero que ya han fallecido, ni afecto hacia las personas que conformaron y conforman esa comunidad educativa. El amor personal, la calidez del corazón, los recuerdos bio-gráficos, la nostalgia, la esperanza, el sentimiento de familia, corresponden a los sujetos humanos, sin duda. La IA no puede acceder a ellos como experiencias vividas, en cuanto vivencias personales, por lo que no hablará o comunicará a su respecto como puede hacerlo un humano.

He aquí, en resumen, sin duda, una diferencia de fondo, a pesar de lo que numerosas películas de ciencia ficción nos han mostrado (*Blade runner*, 2001, *Star Wars*, *The Sphere*, etc.). Por esto, seamos conscientes de que la tecnología de la IA actual no posee “originalidad personal” en su sentido pleno; no es capaz de tintar realmente con la marca o el sello irreplicable de lo personal; ello, según han denunciado penetrantes análisis críticos de sus verdaderas capacidades (Casacuberta y Guersenzvaig, 2022). Aunque, por supuesto, en este lugar preciso, no se desea en absoluto menospreciar, mediante esta matización de fondo, o negar, en modo alguno, las enormes posibilidades operativas, e incluso profesionales, que abre el uso de sistemas de IA; y, de hecho, indicamos en particular el que resultan admirables las capacidades prácticas de estos sistemas en el área específica del lenguaje humano. Por ejemplo, pensemos en sistemas concretos de nuevo como el conocido y ya citado “chat-GPT” u otras redes dotadas de IA en el ámbito del lenguaje (Molina, 2023).

Ahora bien, aunque sistemas de IA imiten la inteligencia-emocional personal, y lo expresen a través de la emulación del lenguaje humano, a fin de inter-actuar con nosotros, no podemos perder nunca de vista que en su caso esto se opera tan solo al modo del fingimiento, el de la apariencia. Esto es, debemos captar que estamos por encima de todo ante “una actuación”. En suma, con esto, asistimos a una mera “representación”, cual en el teatro o arte dramático. La IA, no lo olvidemos, se mueve como el actor o la actriz. Desarrolla su papel y realiza su “función” con respecto a nosotros, adoptando rasgos en

su rol de una intensa proximidad a lo humano, y ello nos confunde con frecuencia. Mas, no nos instalemos en el lugar del espectador teatral pasivo a su propósito. Su núcleo interno es el de una máquina o artefacto, no un corazón de carne, como el nuestro. Ninguna IA alberga interiormente el sentimiento personal de echar de menos a otra persona, aunque pueda estar programada para aparentarlo y aprender a progresar en esa apariencia.

Mantengámonos, en fin, en guardia, frente a la equivocidad, la ambigüedad, que esta clase de inter-acciones van a presentar de un modo creciente en el futuro, y de un modo muy especial con respecto a aquellas en las que entre en juego la dimensión afectiva y emocional. Téngase en cuenta que sus desarrolladores están dotando intencionadamente de estas habilidades o rasgos “espejo” a los artefactos de la IA. Esto se realiza con el fin de que resulten cada día más “empáticos”. En principio, el propósito de ello no estriba en manipularnos a través de las emociones, sino en hacerlos más cercanos, más amigables. Se busca así facilitarnos la inter-acción con los mismos. Sin embargo, corremos por ello graves riesgos, y la frecuencia e intensidad del trato con tales sistemas propicia el que anudemos determinados vínculos personales con estos. En definitiva: el intenso poder de conmoción de estas formas de relación empáticas de la IA, y su influencia en los sujetos humanos, recomienda una delicada prudencia a su respecto.

7.- CONCLUSIONES.

Este trabajo se ha orientado al análisis de lo ético con respecto a la IA. Sin embargo, frente a muchos otros, no se ha contentado con exponer los principios o valores y las pautas normativas más básicas a ese propósito. Guiado por el afán de facilitar una vivencia real y fecunda de la ética, en las relaciones que se establecen entre los sujetos humanos y los artefactos o sistemas de la IA, ha explorado los fundamentos de todo ello. A este fin, ha profundizado en las bases antropológicas sobre las que se asienta la deontología y la ética al aplicarse al terreno de lo tecnológico y de la IA.

El texto se articula en torno a una serie de claves éticas, extraídas de la reflexión sobre este fenómeno. Entre las mismas, se encuentran la distinción inicial y capital entre el sujeto humano y las realidades artificiales de la IA en cuanto al valor de la dignidad personal. Asimismo, se adentra en los tres principios esenciales de lo ético en la IA: responsabilidad, intimidad y justicia. Pasa, luego, a analizar el asunto de los hábitos suscitados por el empleo de la IA y su incidencia en nuestras capacidades. También, se sumerge en la cuestión del desarrollo de unas actitudes y conductas prudentiales respecto de la busca y tratamiento de la información, en orden a la seguridad personal y colectiva. Termina reflexionando, desde la cautela, acerca del aspecto empático que progresivamente se está incorporando a esta tecnología y proporcionando algunas pautas para una inter-acción fértil y segura con estas realidades. Por ello, se proporcionan orientaciones para moverse en el arenoso y confundente terreno de las apariencias y semejanzas con lo humano que estos sistemas están desarrollando, en especial en el aspecto de la inteligencia-emocional y de lo afectivo, ya que ello reclama éticamente de una extrema prudencia.

Todo esto se sitúa, en síntesis, en el contexto general de la necesidad de desarrollar hoy un Humanismo tecnológico de fondo. Este Humanismo ha de cooperar a ubicar las relaciones humanas con entes y sistemas de la IA dentro de un marco de valores adecuado.

El eje de dicho marco axiológico debe consistir en la atención respetuosa a la centralidad de la persona y del servicio a esta.

8.- FUENTES:

BIBLIOGRAFÍA:

- AA.VV. (actualización 2022). *Diccionario de la lengua española*, Real Academia de la Lengua española; (consultado el 22 de enero de 2023); <https://dle.rae.es/>
- Baltar, E. (2020). “El poshumanismo en la UCI de la realidad”, en revista *Telos*, nº114, Fundación Telefónica, septiembre, 85-89.
- Barraca, J. (2017). *Originalidad e identidad personal: claves antropológicas frente a la masificación*, Madrid: San Pablo.
- (2018). *Aportaciones a una antropología de la unicidad*, Madrid: Dykinson, Madrid.
- (2021a). “Humanismo digital y uso prudente de las TICS en lo inter-personal”, en *HUMAN REVIEW: International Humanities Review / Revista Internacional de Humanidades*, vol. 10, Nº. 1, 2021, págs. 87-97.
- (2021b). “El transhumanismo ante el límite de la dignidad personal”, en *TECHNO REVIEW: International Technology, Science and Society Review / Revista Internacional de Tecnología, Ciencia y Sociedad*, vol. 10, Nº. 2, 2021, págs. 173-184.
- (2021c). *Trabajo, deber y vocación: El arte de madurar en la responsabilidad profesional*, Madrid: Ygriega.
- (2022). “Interrogantes abiertos por el post-humanismo y la originalidad personal”, en *TECHNO REVIEW: International Technology, Science and Society Review / Revista Internacional de Tecnología, Ciencia y Sociedad*, vol. 11, Nº. 1, 2022, págs. 57-68.
- (2017). *Originalidad e identidad personal: claves antropológicas frente a la masificación*, Madrid: San Pablo, 2017.
- Camacho, J. y Villas, Mónica (2022). *Manual de ética aplicada en inteligencia artificial*, Madrid: Editorial Anaya Multimedia.
- Casacuberta, D. y Guersenzvaig, A. (2022), “Las falacias del encantamiento con la inteligencia artificial de ChatGPT”, en *SINC (ciencia contada en español)*, innovación-análisis, 16 del 12 2022; <https://www.agenciasinc.es/Opinion/Las-falacias-del-encantamiento-con-la-inteligencia-artificial-de-ChatGPT>
- Comisión Europea (2020): *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza*; COM (2020) 65 final, Bruselas, p. 3. <https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1>
- Diéguez, A. (2017), *Transhumanismo*, Barcelona: Herder.
- Fernández Fernández, J. L. (2021). “Hacia el Humanismo digital desde un denominador común para la Ciber Ética y la Ética de la Inteligencia Artificial”. En *Disputatio, Philosophical Research Bulletin*, vol. 10, nº 17, junio 2021, pp. 107-130.
- Fernández Mateo, J. (2021a). “La técnica es el nuevo sujeto de la historia: posthumanismo tecnológico y el crepúsculo de lo humano”, en *Revista Iberoamericana de Bioética* (2021), nº 16, pp. 01-15. DOI: 10.14422/rib.i16.y2021.004.
- (2021b). “Antropología estética en el tecnoceno: epistemología y nihilismo”, en *TECHNO REVIEW, Revista Internacional de Tecnología, Ciencia y Sociedad*, 9(2), pp. 61-78. <https://doi.org/10.37467/gka-revtechno.v9.2807>.
- Forment, E (1998). *Id a Tomás*, Pamplona: Ed. Fundación Gratis Date.

- Han, Byung-Chul (2022). *Psicopolítica: neoliberalismo y nuevas técnicas de poder*, Barcelona: Herder.
- Lévinas, E. (1993). *Humanismo del otro hombre*, traducc. G. González R.-Arnáiz, Madrid: Ed. Caparrós.
- (2002). *Totalidad e Infinito*, Salamanca: Sígueme, traducción de D. E. Guillot, 6ª ed.
- López Quintás, A. (2022). *Las cimas de la cultura y el ascenso al amor oblativo: un método educativo ilusionante*, coleccion. Digital, Pozuelo de Alarcón (Madrid): Ed. UFV.
- Llano, Fernando H. (2018). *Homo Excelsior. Los Límites Ético Jurídicos del Transhumanismo*, Valencia: Editorial Tirant lo Blanch.
- Marías, J. (1997). *Persona*, Madrid: Alianza.
- Molina, S. (2023). “Cómo utilizar ChatGPT de forma eficiente para impulsar tu carrera”, en *VOGUE: BUSINESS*, 12 de enero; business-vogue-es.cdn.ampproject.org.
- Pinto, J. A. (2020). *El Derecho ante los retos de la inteligencia artificial: marco ético y jurídico*. Madrid: Ed. Edisofer.
- Rodríguez Valls, F. (2017). *Orígenes del hombre: la singularidad del ser humano*, Ed. Biblioteca nueva.
- Spaemann, R. (2000). *Personas. Acerca de la diferencia entre “algo” y “alguien”*, Pamplona: Eunsa.

FILMOGRAFÍA:

- Kubrick, S. (1968). *2001: A Space Odyssey*. Metro-Goldwyn-Mayer
- Levinson, B. (1998). *Sphere*. Baltimore Pictures
- Lucas, G. (1997). *Star Wars*. Lucasfilm
- Ridley, S. (1982). *Blade runner*. Alcon Entertainment, Columbia Pictures