



Alfonso Moraleja Juárez (Universidad Autónoma de Madrid)

Parece más fácil determinar qué debemos hacer con respecto a los problemas que plantean las nuevas tecnologías y la IAG, que establecer cómo llevar a cabo las posibles soluciones. Lamentablemente es necesario tener una perspectiva realista de la situación política y social actual, lo que significa que no podemos esperar desenlaces significativos al respecto. En cualquier caso, es preciso a su vez negar por principio el determinismo tecnológico que subyace en muchos planteamientos asociados a la autonomía de la tecnología. Si algo nos enseña la historia es que las inercias tienen un recorrido limitado y que la imprevisibilidad en la historia es también una constante.

El homo sapiens es también es un *homo insipiens*, y no por su ignorancia, sino debido a su ingenua creencia en su invulnerabilidad, lo que evidentemente le hace más vulnerable. Al parecer, solo un desastre cuasi irreversible podría propiciar la supeditación de los intereses económicos y políticos de las respectivas potencias y superpotencias a intereses de seguridad conjunta. Ciertamente, la división aparentemente irreconciliable entre los distintos bloques de poder nos remite a un futuro próximo poco halagüeño, lo que

no significa que la inacción y la pasividad sean posibilidades lógicas.

Los problemas tecnológicos y científicos pueden tener soluciones sociales, culturales y políticas, en la misma medida en que un determinado tipo de sociedad y política ha propiciado que la tecnología y la ciencia se presenten como un problema. El fracaso de organizaciones internacionales como la ONU o el idealismo pretencioso de la Declaración de los Derechos Humanos nos remiten de nuevo a un futuro poco esperanzador. Sin embargo, la paradoja reside en que el problema es parte de la solución: sólo una ONU fortalecida (o una nueva organización supranacional a nivel global), una autentica declaración no sólo de derechos sino de obligaciones y unos protocolos tecnológicos conjuntos de actuación realmente vinculantes podrían crear las condiciones de posibilidad de un futuro mejor. Dichos instrumentos no pueden ser diseñados de manera exclusiva por Occidente o por un par de continentes. Los Derechos Humanos, a pesar pretender mediar entre la espera ética y jurídica, son máximas tan genéricas como poco vinculantes: por mucho que cualquier Estado los acepte, no tiene por qué garantizar su cumplimiento. Para colmo, el propio iusnaturalismo del que parten los DDHH será cada vez más puesto en tela de juicio por los avances tecnológicos venideros. Al proceso evolutivo de la naturaleza humana se podrían incorporar tales cambios biotecnológicos que harían de esa supuesta naturaleza una bella alegoría del pasado. Es evidente que la fundamentación última no podría ser teológica. Podríamos pensar que serían las necesidades humanas (Bobbio) quienes establecieran las respectivas líneas a seguir; pero de nuevo Rorty nos recuerda que cada tradición cultural podría establecer al respecto (como de hecho sucede hoy) líneas de actuación en absoluto coincidentes. Es preciso recordar que la propia democracia constituye tan solo un sistema político perteneciente a una tradición determinada.



El mundo del mañana, si es posible, deberá contar con la participación de todos los miembros y Estados de manera activa. La prueba de que estamos lejos de una solución consensuada es el numeroso catálogo de principios y declaraciones que pretenden limitar los usos y abusos de la IA: Los principios de Asilomar, la Declaración de Montreal, El reglamento de la UE del 2024; amén de las iniciativas o marcos de gobernanza de países como EEUU, China, Japón, Brasil, Australia o Singapur. La pregunta es: ¿por qué nuestra línea de actuación debe ser conjunta? Porque cualquier actuación particular sería como intentar poner puertas al campo: estas nuevas tecnologías no saben de fronteras, los distintos Estados y los propios usuarios impiden una delimitación real de los nuevos desarrollos tecnológicos. Los “paraísos regulatorios”, con regulaciones más laxas, impedirían soluciones reales.

Por si esto fuera poco, creo que los intereses económicos y políticos seguirán siendo la prioridad generalizada para casi todas las potencias y superpotencias. La sociedad civil y la comunidad científica seguirán estando supeditadas a los intereses nacionales e internacionales. A pesar de nuestro escepticismo, nos atreveremos a establecer una propuesta no tecnológica (y bastante idealista) asociada a la transformación del discurso, mirando siempre de reojo a la teoría de la acción comunicativa de Habermas. Ya que el discurso de la clases política parece haber perdido la carga de ejemplaridad que debería tener para la ciudadanía, esperemos que, en este ámbito al menos, la ciudadanía pueda

ejercer cierta influencia en esa decepcionante clase que supuestamente nos representa. Nuestra atención por el lenguaje pretende también influir en los posibles lenguajes de programación. Al igual que los distintos códigos generan distintas Inteligencias Artificiales y distintas tecnologías, es necesario reivindicar que una utilización del lenguaje bien distinta también pueda ser parte de la solución.

La política de la hipérbole: radicalismo y polarización

¿Sigue siendo válida la afirmación de Adorno de que “solo la exageración es verdadera”? Lo que parece evidente es que la crítica de Adorno a la presente “industria cultural” sigue teniendo sentido: más que nunca, nuestra forma de razonar es claramente instrumental y los nuevos mercados siguen ofertando un entretenimiento superficial en el que sus productos solo son mercancías y fetiches; no obstante, junto al cine, a la televisión y a la radio se añaden hoy en día las redes sociales, internet, la industria del videojuego y las potentes prestaciones de la IA. El salto ha sido cualitativo si lo comparamos con los efectos de los medios precedentes. Adorno critica también la tendencia a reducir la complejidad y la diferencia a categorías fijas e identitarias; pero ahora, el pensamiento único se ha convertido en pensamiento polarizado. Creo que la polarización a la que hemos llegado no tiene precedentes en la historia de la humanidad. Es necesario reconducir nuestros pensamientos y nuestras gramáticas a zonas de analogías que superen las dicotomías irreconciliables, sin caer en el lamentable posmodernismo del “todo fluye” tan próximo al “todo vale”. La filosofía de Séneca o Gracián nos aconseja “nunca exagerar”, a pesar de que esta máxima también puede ser exagerada, parece que el momento en el que nos encontramos demanda una mayor proximidad a este prudencialismo. La política de la hipérbole en la que estamos inscritos ha perdido su intencionalidad, ha olvidado su conciencia de exageración y el énfasis se ha adueñado de la literalidad. La exageración se ha instaurado en la cultura del espectáculo: su carácter llamativo e incendiario la hacen paradójicamente más creíble y contagiosa. El tremendismo aplicado a la política convierte a toda reflexión en algo grotesco y vehemente. La deformidad y la visceralidad nos alejan de toda posible solución. La violencia que se ejerce en la descripción y en el análisis se paga con la reproducción y multiplicación de la propia violencia. La exageración, al perder toda medida y equilibrio, pierde su capacidad crítica. Los discursos tienden a ser arengas, proclamas y peroratas en las que la vehemencia sustituye a una razonable razón. Es preciso diseñar un nuevo prudencialismo político que establezca territorios comunes entre posiciones cada vez más polarizadas, y por lo tanto cada vez más extremas.

A lo largo de nuestra historia, la polarización ha podido estar propiciada por

múltiples factores biológicos, culturales y sociales. La propia estructura psíquica binaria y las estructuras del parentesco (Lévi-Strauss), la polaridad gramatical y clasificadora (Lloyd), el binarismo matemático (Boole), los sexos y los géneros (Durkheim), la asimetría física (Hertz), la religión (Berger) y en especial las religiones monoteístas (Armstrong), la lucha de clases (Marx), el conflicto intergeneracional (Ortega), el choque de culturas (Huntington) y, por supuesto, entender la política desde la distinción “amigo/enemigo” (Schmitt) han influido considerablemente en nuestra forma polarizada de pensar y de relacionarnos. No obstante, parece indudable que, en nuestra época, las nuevas tecnologías han potenciado y multiplicado estos procesos, lo que a su vez conlleva un mayor intento de adoctrinamiento y una recaída en la intolerancia. Señalamos a continuación algunas posibles causas:

1. El filtro burbuja de las redes sociales: el mayor aislamiento de los sujetos y una vinculación personificada de contenidos que tiende a reforzar los gustos e ideas del consumidor de datos.
2. El “sesgo de confirmación”: el “eco” de las informaciones se amplifica y se radicaliza para llamar cada vez más la atención del espectador u oyente.
3. El carácter cada vez más gregario y tribal de un nuevo sujeto cada vez más individualista. El individualismo no busca discutir ideas, tan solo compartirlas con otros de manera acrítica.
4. Las redes sociales y el contacto no personal y directo no propicia la empatía ni la comprensión. Oponerse es más fácil e inmediato.
5. El aumento desproporcionado de las desigualdades sociales y económicas afianza una ideología de clases cada vez más extrema, más aún que la que existe entre las propias clases.
6. El descontento social alimenta la radicalización.
7. El radicalismo, aunque ya no es algo extraordinario, sigue siendo llamativo y fascinante: lo espectacular vende.
8. La desinformación y la manipulación informativa busca la polarización, no el cuestionamiento crítico y prudente.
9. La polémica polarizada de la política incentiva la propia polémica: crea la sensación de que ese tipo de discurso es tan inevitable como necesario.

De la realidad del “qué” a la filosofía política del “cómo”. Debate versus deliberación

Mandela era consciente de que cuando demonizamos a nuestros adversarios, abandonamos toda posibilidad de resolver pacíficamente nuestras diferencias; nuestros

adversarios se convierten en enemigos y lo que pretendes consecuentemente es hacer que desaparezcan, no hablar con ellos o llegar a acuerdos. Hobbes nos recuerda por su parte que cuando dos personas quieren algo que solo uno puede disfrutar, ambos se ven como enemigos y buscan la eliminación o el sojuzgamiento del otro. Ahora bien, aquello que debemos disfrutar no es algo que deba estar en unas solas manos, estamos hablando de un futuro próximo compartido, querámoslo o no.

Dejamos de momento de lado todas las manifestaciones y efectos psicológicos que la polarización está causando: el pensamiento dicotómico (pensar “en blanco y negro”, categorizar las experiencias y las personas de forma opuesta, extremista y excluyente) tiende a cierto catastrofismo (creer que siempre el peor resultado será el más probable), a cierta etiquetación (asignación acrítica de rasgos globales a uno mismo y a otros) y a una reafirmación del razonamiento emocional (creer que son los sentimientos quienes mejor interpretan la realidad). En el fondo, la polarización es una simplificación, una forma de pensamiento que a la postre hace que el sujeto no pueda explicar la complejidad, de ahí que su parquedad explicativa se traduzca en enemistad contra la misma realidad que le circunda.



Escuela de Atenas de Rafael_Estancias palacio Vaticano. Flickr

Es evidente que el “dataismo” que alimenta a las redes sociales, a la tecnología digital y a los medios de comunicación es proclive a la polarización. Ante la pregunta “¿posibilitan las redes sociales y la tecnología la democracia?”, la respuesta es claramente “no” si entendemos la democracia desde una sentido deliberativo. Como ya he subrayado, los problemas tecnológicos no tienen por qué tener soluciones tecnológicas, y nuestra propuesta va precisamente por ese camino: otras formas de relaciones humanas pueden propiciar una relación con las nuevas tecnologías también más humana. Frente a la forma de discurso habitual del espacio televisivo, del discurso de las redes sociales e internet, solo la deliberación parece poder cuestionar la dialéctica amigo/enemigo que subyace y que está alimentando la exclusión, la intolerancia y la exasperación. Ni qué decir tiene que la identificación entre valores (o ideas) y sujetos que algunas tradiciones culturales establecen, hace inviable nuestra propuesta. Si Dios u otros valores absolutos se consideran “la Verdad”, cuestionar determinadas ideas o valores se puede traducir como un cuestionamiento a las personas que los detentan. A continuación presentamos, de manera esquemática, nuestra idealista propuesta en la comunicación.

DEBATE	DELIBERACIÓN
Dialéctica amigo/enemigo	<i>Dialéctica ciudadano/conciudadano</i>
Busca las debilidades en la argumentación de la otra parte (fallos, contradicciones)	<i>Busca la fortaleza en la argumentación de la otra parte para (puntos de encuentro, aciertos)</i>
Subraya las diferencias respecto a nuestra postura	<i>Subraya las similitudes respecto a nuestra postura</i>
Se busca demostrar que el otro está equivocado	<i>Se busca lo que hay de cierto en los planteamientos del otro</i>
Busco que mi punto de vista triunfe y se imponga	<i>Se buscan contenidos de verdad que faciliten la acción conjunta</i>
Mi posición es la verdadera (verdad cuasi absoluta)	<i>Nuestras posiciones establecerán los criterios de verdad (susceptibles siempre de ser modificados)</i>
Se tiende a soluciones originales	<i>Se busca las mejores soluciones</i>

Las ideas tienden a asociarse a creencias personales: atacar a las ideas se considera atacar a la persona	<i>Las ideas no se asocian a creencias personales, Las ideas pueden ser compartidas y no son estrictamente personales</i>
Potenciación de la competitividad y de la polémica	<i>Potenciación de la colaboración y de la capacidad para llegar a acuerdos</i>
Expresiones: “Yo tengo (la) razón”	<i>Expresiones: “Es preciso ver quiénes pueden fundamentar mejor sus razones”</i>

Algunos prejuicios de nuestra ingenua relación con la tecnología

Los prejuicios a los que me referiré a continuación fundamentan en buena medida buena parte de los planteamientos tecnológicos actuales. Estos “juicios previos”, no fundamentados, no tienen el mismo carácter discriminatorio que se arroja contra determinadas personas, grupos sociales o actitudes; curiosamente, si pecan de algo es de un optimismo desahogado, de una fe incuestionable con respecto al supuesto “progreso” tecnológico, progreso que, por otra parte, no coincide en absoluto con el de la condición humana. La prueba de todo ello está en la innumerable multitud de seres humanos que todavía no viven, sino sobreviven. Somos capaces de crear “casas inteligentes” habitadas de individuos estúpidos, o capaces de crear robots sin darnos cuenta de que el comportamiento de muchas personas se ha automatizado. Si en un futuro próximo somos capaces de crear seres más inteligentes que nosotros, será la prueba definitiva de nuestra falta absoluta de inteligencia. Lamentablemente, al firmar el contrato con esta nueva infraestructura no leímos la letra pequeña: “La servidumbre será la *conditio sine qua non* de la tecnología”. Buena parte de la tecnología que supuestamente usamos nos usa a nosotros. ¿Puedo deshacerme de mi móvil o de mi ordenador? Sólo en teoría; dentro de no mucho, ni en sueños. La autonomía que creamos en la tecnología es directamente proporcional a la dependencia que generamos de ellas.

Prejuicios tecnológicos:

1. La tecnología es autónoma, e incluso independiente: el ser humano ya no puede impedir o revertir la evolución tecnológica.

Objeción: A pesar de que podemos corroborar cierto “progreso” tecnológico, de momento nada muestra ni demuestra que no se pueda desacelerar, revertir o detener. Me temo que

vemos la tecnología desde una visión tecnológica.

- Si algo puede hacerse debe hacerse.

Objeción: “Puedo, pero no quiero” (Ellul), los valores y las finalidades deben de primar sobre la praxis tecnológica.

- Síndrome Titanic: el mundo tecnológico nos permite prever y controlar el mundo que nos rodea.

Objeción: La imprevisibilidad en sistemas complejos (“Teoría del caos”) es negada por el control que tenemos en distintos ámbitos cotidianos y tecnológicos. Como la magia, la tecnología propone un universo determinado y omnicomprensivo (Lévi-Strauss). Es necesario aplicar en este ámbito también el principio de indeterminación.

- Toda acción humana cuenta con la reacción pertinente de signo contrario.

Objeción: Pueden existir efectos secundarios tan imprevisibles como irreversibles. Toda acción por principio es irreversible de forma total.

- Mito positivista: el “objeto” y el “dato” son objetivos.

Objeción: El “objeto” y el dato” son construcciones epistemológicas humanas (Quine). El nuevo dato ya no es una determinada información que recibimos (datum), sino una información que construimos.

- La tecnología es neutral.

Objeción: La tecnología no sólo depende de su uso que de ella se haga, sino que tiene un componente metafísico (Heidegger) y político (Winner) dado el enorme poder de transformación que conlleva. Las constantes proclamas sobre la neutralidad de la tecnología muestran su posición.

- La tecnología es una prolongación de nuestra humanidad.

Objeción: La tecnología también prolonga nuestra inhumanidad al establecer una “segunda naturaleza” tanto entre nosotros como en el ámbito natural (Ellul).

- Desmedida fe en el progreso y en el progreso tecnológico.

Objeción: ¿cuál es la meta de la humanidad? Solo si pudiéramos contestar a esta pregunta podríamos saber si nos acercamos o nos alejamos de la meta. El progreso tecnológico no garantiza en absoluto el progreso de la humanidad.

- Todos los problemas humanos cuentan con soluciones tecnológicas.

Objeción: Hay problemas humanos que pueden tener soluciones tecnológicas; y problemas tecnológicos que exigen soluciones sociales y culturales.

1. Toda acción humana cuenta con la reacción posible y previsible de signo contrario.

Objeción: Pueden existir efectos secundarios tan imprevisibles como irreversibles. Toda acción por principio es irreversible de forma total.

1. Prejuicio democrático: los avances tecnológicos revierten en el conjunto de la humanidad.

Objeción: El pasado y el presente de la tecnología cuestionan tal principio. La tecnología no es ni gestada ni controlada de manera democrática.

1. Prejuicio antropocéntrico: la referencia central de todas nuestras reflexiones debe ser el ser humano.

Objeción: El ser humano forma parte de un todo, de un ecosistema vital (Riechmann).

1. Prejuicio temporal agustiniano: solo existe el presente, preocuparnos por el futuro no tiene sentido.

Objeción: Tenemos una responsabilidad con respecto a las generaciones futuras (Jonas). No heredamos la tierra, la tomamos prestada de las generaciones venideras.

Sobre la Inteligencia Artificial: de la ética a los postulados

Lo primero que debo decir para no decepcionar excesivamente al lector es que no creo demasiado en la expresión “Inteligencia Artificial”: al igual que podemos hablar de distintos usos, grados y tipos de pensamiento y de razón en el ser humano, también podemos establecer distintas formas de inteligencia. Más allá de la teoría de las inteligencias múltiples de Gardner, la inteligencia puede ser definida desde múltiples perspectivas. ¿Es la inteligencia una facultad asociada al discernimiento y a la comprensión

(intellegentia)?, ¿una forma de conocimiento racional encargada de establecer los mejores medios para alcanzar determinados fines?, ¿un conjunto de funciones que nos permiten adaptarnos a distintas circunstancias?, ¿todo nuestro conocimiento está constituido exclusivamente por nuestro intelecto? Estas cuatro preguntas, asociadas al ámbito filosófico, psicológico, biológico y sociológico respectivamente, muestran la complejidad de tema que nos ocupa. Desde el punto de vista filosófico, la inteligencia exige (entre otras cosas) comprensión profunda y (toma de) conciencia, algo que hoy por hoy tan solo es una promesa asociada a una hipotética singularidad. Desde el punto de vista biológico, adaptativo, debemos decir que tal vez el ser humano es el animal menos inteligente del planeta, ya que si de adaptación hablamos, ninguna especie tiene la capacidad que sepamos de poner en serio peligro tanto a su propia especie como a todas las demás. No es casual que buena parte de los estudios que se realizan sobre IA, sobre todo en Asia, se dirijan al mundo de los insectos. Nuestra aproximación psicológica –a pesar de su simplismo, por entender las relaciones mentales solo desde la dicotomía de medios y fines– parece, hoy por hoy, la que más aproxima al ser humano con la IA. Por último, una valoración de la inteligencia más amplia, percibida desde un punto de vista social y cultural, nos hace caer en la cuenta de que la inteligencia humana no es independiente de las actitudes, hábitos, habilidades, instituciones, herramientas o experiencias que mantenemos. Por ello mismo, cuando hablamos de IA, deberíamos tal vez referirnos a (*Integrates Abilities*), “Aptitudes Integradas”, o como sugiere Floridi, “Agencia Instrumental”.

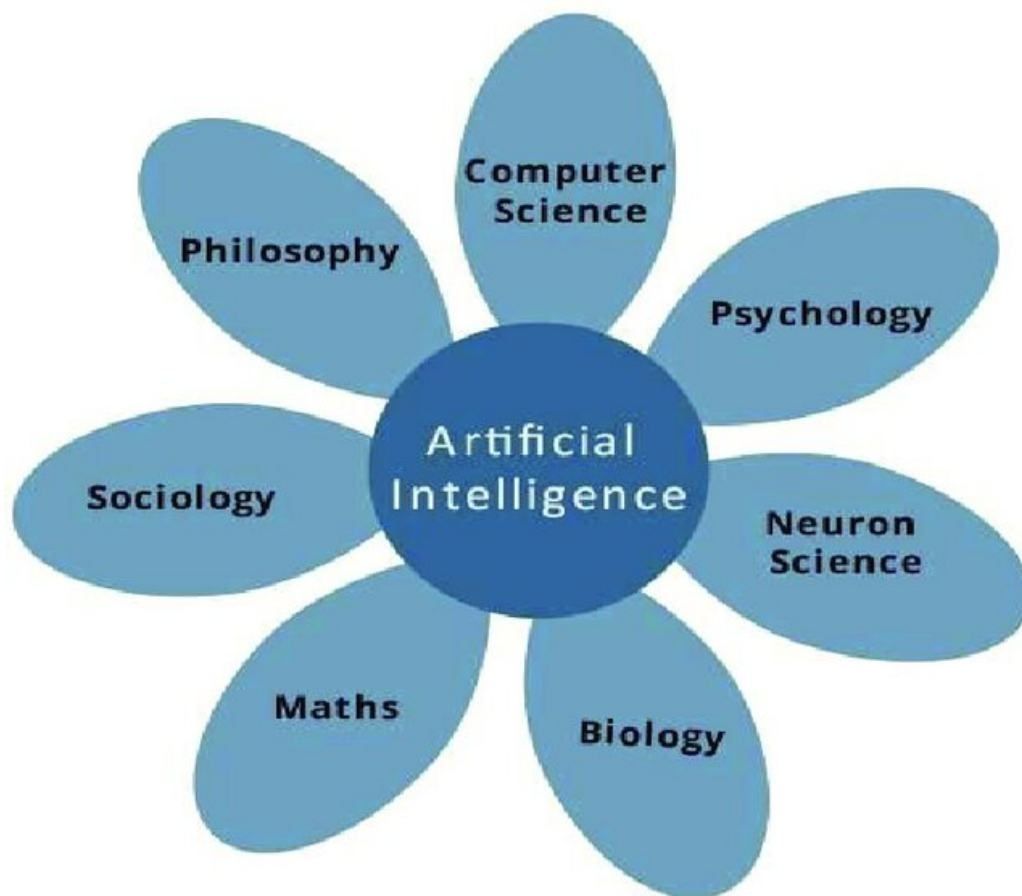


Figure 1. Artificial intelligence and its application in various disciplines. Available via license: [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

La inteligencia puede ser también imaginativa, ingeniosa, sentiente, afectiva, intuitiva e incluso ética. La polivalencia de la inteligencia es tal vez su característica y virtud fundamental. Como bien precisa Gould, la inteligencia no es algo único e innato asociado a la letra **g**, es más bien un conjunto de capacidades en ocasiones muy diversas. En cualquier caso, parece que inevitablemente nos referimos siempre a la inteligencia humana, lo que hace que caigamos en cierta circularidad, pues tendemos a rechazar todas aquellas capacidades o aptitudes que no están asociadas a nosotros. En cualquier caso, la IA se crea y potencia –en principio– para desarrollar supuestamente todas y cada una de las aptitudes humanas. Tal vez, una forma de aproximarnos a la inteligencia es mediante sus antónimos es decir, no desde un planteamiento sintáctico, sino desde un punto de vista semántico. Algunos de estos antónimos son: la estupidez (o estulticia), la torpeza, la cerrazón, la

cortedad, la desavenencia, la brutalidad, la ignorancia, la ineficacia, la insensatez, el instinto, la incapacidad o la ineptitud. Dos últimas cuestiones sobre esta cuestión: la inteligencia no es inmune a un esquema espacio temporal, es decir, a una vinculación histórica y cultural: la hipervaloración que de la tecnología hacemos hoy en día frente al conocimiento de otros ámbitos (Arendt) o nuestra utilización desmedida de la razón instrumental (Escuela de Frankfurt) hacen que nuestra visión de la inteligencia esté asentada bajo unas coordenadas muy específicas. Por poner algún ejemplo de ello, puede sorprendernos que para un pensador del siglo XVII como Gracián, la inteligencia sea una afortunada combinación de juicio, ingenio, inventiva, agudeza, prudencia, discreción, talento, destreza e interés en el aprendizaje; características todas ellas sobre todo individuales pero de aplicación social. En segundo lugar, es necesario hablar también de una intelligenza en torno a la IA, es decir, de auténticos lobbies, grupos de poder en determinados casos pero también de interés, no centralizados, dispersos pero muy influyentes, con poderosas motivaciones económicas y políticas. Solo así se explica la insistente vinculación e influencia que quiere ejercer el mundo de la IA en todos los ámbitos de nuestras vidas. La IA vende: todo lo que tiene que ver con la IA se sobrevalora en virtud de las posibilidades económicas y de poder que genera y que puede generar. Es indudable que esto plantea un problema en absoluto menor: cuando el capital o la motivación económica es quien dirige las investigaciones, no solo se querrán obtener determinados resultados a corto plazo, sino que dichos resultados, por pequeños que sean, deberán sobrevalorarse para seguir obteniendo valor. Para que esto suceda, es importante establecer una adecuada campaña de marketing en la que la publicidad de su inevitabilidad, nuestra absoluta dependencia e incluso el miedo jueguen su papel.

Propondré en primer lugar una definición de ética para comprobar hasta qué punto podríamos hablar de una “ética de la IA”; en segundo lugar, intentaré ver algunas de las posibles relaciones entre la dimensión ética y los avances de esta tecnología.

La ética es aquella disposición culturalmente adquirida, desde la cual los seres humanos pensamos y decidimos comportamientos y acciones individuales entre aquello que nos mejora y desarrolla y lo que nos perjudica y envilece. La ética presupone libertad y exige intersubjetividad y reciprocidad.

Una disposición exige ciertas capacidades y aptitudes. En ocasiones puede significar talento, no obstante, las condiciones que nos capacitan son adquiridas dentro de una determinada cultura. ¿Esto significa caer en un relativismo cultural? No necesariamente: cada momento histórico presenta ciertos universales culturales y valores asumidos en la

mayoría de los casos de manera no consciente. Sabemos que es más difícil establecer unos universales culturales y éticos a través de la historia de la humanidad que seleccionar algunos preceptos éticos actuales asumidos por la gran mayoría de culturas presentes. Los Derechos Humanos (DDHH), a pesar de su carácter idealista, utópico y manifiestamente occidental, siguen siendo la referencia inmediata. En cualquier caso, urge una nueva elaboración de los DDHH con un protagonismo bien distinto.

Otro problema: La definición se refiere al pensamiento y a la capacidad decisoria a la hora de establecer aquello que nos mejora y nos desarrolla evitando lo que nos perjudica y envilece. Al parecer la IAFuerte actual no sólo establece procesos de pensamientos próximos al que puede realizar el humano, sino que supera con mucho en algunos procesamientos de la información las capacidades humanas. De momento, las formas en las que el ser humano procesa y establece el pensamiento no pueden ser imitadas (ni de lejos) por la IAFuerte. No son pocos los que desde el campo de la neurociencia señalan que, sabiendo que poseemos 10^{10} neuronas y que cada una de ellas tiene una media de 7500 conexiones, es poco probable que se pueda generar una réplica material de dichos procesos: también los microchips tienen una limitación molecular evidente. Es posible que los avances de la IAGenerativa (por ejemplo, en el análisis molecular de los más de 200 millones de proteínas conocidas o en el reconocimiento facial) nos hagan creer ingenuamente que su aplicación no tiene límites. Respecto al problema de la decisión y de la deliberación humana, parece que ello nos remite al problema de la libertad. Hasta hace poco, los problemas que una IADébil podía resolver estaban dentro de una determinada programación, no obstante, la estocástica de las nuevas y complejas programaciones permite cierta aleatoriedad no determinista, cuantificable tan solo de forma probabilística. En cualquier caso, hay que reconocer también que las decisiones humanas pueden ser cuantificables desde el punto de vista sociológico por la estadística. La computación cuántica (todavía en pañales) parece prometer una estocástica exponencial.

Respecto al problema de la consciencia, la IASingular (la singularidad de la IA) también está lejos de aparecer. En el terreno de la filosofía y de la psicología, ni siquiera disponemos de una definición precisa de consciencia humana. Si por consciencia entendemos “la capacidad de ser y reconocer la realidad circundante y de relacionarse con ella”, parece evidente que eso ha sido más que logrado. No obstante, si la consciencia se entiende como el reconocimiento que el sujeto (o un ser) tiene de sí mismo, de sus reflexiones y de sus actos (algo evidentemente importantísimo desde el punto de vista moral), la distancia de momento entre nosotros y la IA es enorme. Es interesante establecer

por contraposición la incapacidad manifiesta de otro ámbito: el del inconsciente. Parece improbable que una IAFuerte, Singular o Generativa pueda tener distintos niveles de procesos e informaciones ocultos para sí misma. Como no podía ser de otra forma, el problema de la consciencia remite de manera inmediata al problema de la conciencia (conscientia lat.), ya que tener un reconocimiento de las reflexiones y actos propios significa entenderlos y asociarlos a lo correcto (al bien) y a lo incorrecto (al mal) en cierta medida o grado. En el ser humano (o al menos en su gran mayoría) el espacio que existe entre la consciencia intelectual y la conciencia moral es mínimo, por no decir inexistente.

Si bien es cierto que la IAGeneral parece establecer en determinados ámbitos una conciencia procedimental, analítica, empírica, epistemológica y metódica, otros procesos mentales y morales como la voluntad, la responsabilidad, el remordimiento, la nostalgia, la melancolía, la vergüenza, el escrúpulo, el recato, el sentido común, la prudencia, la culpabilidad, la intencionalidad, la mentira, la contrariedad, o la conciencia transcendental, metafísica, infeliz, social, crítica, utópica, revolucionaria o histórica parecen estar muy lejos de su competencia.

¿Qué significa “individual” en nuestra definición? Significa que no puede existir una forma serial o en serie a la hora de afrontar un problema determinado. Cada sujeto debería confeccionar una perspectiva autónoma a la hora de establecer una posible solución ante una decisión o conflicto. La autonomía exige fundamentalmente llegar a ciertos acuerdos entre determinados sujetos. A pesar de que cada uno tenga una perspectiva individual, es necesario para una sociedad establecer acuerdos, pactos, convenios, compromisos, tratados y convenciones. Para ser precisos, y siguiendo la distinción de Rousseau y Kant, los sujetos estamos “sujetos”, compartimos un campo común vecino a nuestro propio ámbito autonómico. El individuo propiamente dicho no existe, su afirmación da por sentado la independencia absoluta de los sujetos.

Es evidente que nuestra definición de ética concierne exclusivamente al ser humano. En el caso de que la IA tuviera buena parte de las características y facultades humanas, deberíamos evidentemente actualizar dicha definición. Lo paradójico más significativa al respecto es que creo que, mucho antes de la prometida llegada de la singularidad, entrará en escena una nueva especie en distinto grado de humanidad llamada cyborg. Un cyborg como Habisson mantiene que no usa la tecnología, sino que “es tecnología”. El proceso de alienación en este ámbito, el alejamiento de la autenticidad de lo que consideramos que es un ser humano con respecto a la tecnología, ya ha comenzado. Otro proceso más de alienación no causa sorpresa.

¿Es o será entonces la ética algo exclusivamente humano? ¿Réplicas biológicas tendrían ética? ¿Podrían ser considerados como humanos aquellos cíborgs que superaran las condiciones físicas y mentales del homo sapiens sapiens? Esta problemática nos conduce a su vez al problema de la “reciprocidad”. Solemos dar por supuesto que ningún sujeto vale más que otro en su dignidad y derecho, en este sentido, los imperativos categóricos kantianos parecen imponerse con rotundidad. No obstante, de nuevo, algunas cuestiones problemáticas vuelven a aparecer: ¿distintas IAGs mantendrán dicha reciprocidad?, ¿tendrán todas las mismas capacidades y valores?, ¿la ética imperante de todas ellas será la ética kantiana?, ¿la reciprocidad entre seres humanos será comparable a la que pueda existir entre el humano y la IAG o la que distintas IAGs puedan tener entre sí? Me temo que en la complejidad del asunto pueda significar una gran imprevisibilidad de problemas para los que posiblemente no estamos preparados. Para no dar nada por sentado, presento una serie de postulados en torno a la IAG que, como diría Descartes, son ciertos, claros y evidentes. Entendiendo por “postulado” aquel principio no demostrable desde la razón teórica que exige una actuación desde la razón práctica.

Postulado 1: La existencia de la IAG podría ser un hecho a corto o medio plazo.

Postulado 2: La existencia de la IAG podría conllevar más perjuicios que beneficios al ser humano.

Postulado 3: La existencia de la IAG podría en última instancia crear el peor de los escenarios posibles para el ser humano: su extinción.

Consecuencia práctica de los postulados:

Estos postulados exigen del ser humano una actuación conjunta de carácter preventivo.

Para terminar: una píldora de realismo político en el mundo de las IAs

- A. La política con carácter democrático tiende a ser cortoplacista; otro tipo de política no democrática tiende a la autocomplacencia. En ambos casos, el imperativo de Jonas, “obra de tal modo que los efectos de tu acción sean compatibles con la permanencia de una vida auténtica en la tierra”, tiene la fuerza de un “cuídate mucho”.
- B. El imperativo de Jonas parte de dos falsas creencias:
- C. La existencia de la familia humana como una entidad unitaria: las diferencias entre los seres humanos, lejos de reducirse, aumentan. Dentro de no mucho será cómico hablar de “la” humanidad.

- D. La creencia de que la “tierra” existe: existen zonas, países, comunidades, regiones y áreas de influencia. Para colmo, muchas de esas áreas ven como competidoras e incluso enemigas a otras áreas.
- E. Los principios de precaución constituyen una serie de sugerencias bienintencionadas en absoluto vinculantes.
- F. Responsabilidad: a pesar de que pueda probarse que existen alternativas más seguras, la inmediatez, el oportunismo y el coste económico determinan en buena medida las acciones.
- G. Respeto y prevención: Las actuaciones preventivas solo se llevarán a cabo frente a la posibilidad de un riesgo propio.
- H. Obligación de conocer e informar: El secreto forma parte esencial de un mundo constituido por un puzle cuyas piezas no pueden encajar debido a sus distintos tamaños económicos y formas de poder.
- I. Obligación de compartir el poder: El poder sólo se comparte si con dicho gesto se incrementa el poder. Es absurdo creer en la democratización del mundo cuando muchos países no creen ni siquiera en la democracia a nivel interno, y otros consideran de manera incuestionable la superioridad de su democracia. La lucha por el poder (y el conocimiento) es, lamentablemente, irrenunciable.

Para citar esta entrada

Moraleja, Alfonso: «Polaridad, prejuicios y postulados en relación a las tecnologías y a la IA». *Niaia. Revista interdisciplinar sobre formación y resolución de problemas morales*, nº4. 2025. ISSN 2990-0409 Acceso [en este enlace](#).

Creemos en el libre flujo de información. Republique nuestros artículos libremente, en impreso o digital, bajo licencia Creative Commons, citando la fuente



La Web de NIAIÁ y sus publicaciones (salvo aquellas en las que se especifique de otra manera) están bajo una [Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional](#) Etiquetas: [Ética aplicada](#)