Las ideas: Humano e Inteligencia Artificial

Queralt Prat-i-Pubill[1]

No hay nada más práctico que una buena teoría.
Kurt Lewin

Nada es más poderoso que una idea a la que le ha llegado su hora.
Víctor Hugo

Tarde o temprano, son las ideas, no los intereses creados,
las que son peligrosas para el bien o para el mal.
John Maynard Keynes,

Escribo este texto en un momento de avance acelerado de la Inteligencia Artificial ("IA"). Este fenómeno no solo afecta nuestra manera de trabajar o de comunicarnos, sino que incide en nuestra concepción del conocimiento, la realidad y el mundo. En resumen, en nuestras concepciones, y por tanto, afecta a la creación de nuestro futuro. Dos aspectos fundamentales me animan: por un lado, la comprensión de cuáles son las ideas que se socializan sobre la IA y, por otro, algunas reflexiones sobre nuestra concepción de lo que significa ser humano cuando los sistemas de IA pueden desarrollar tareas "inteligentes" mejor que nosotros.

Recordando las palabras de J. M. Keynes y Víctor Hugo, quienes explicitan que las ideas tienen un poder mayor del que imaginamos. También, que las teorías son prácticas porque nos permiten dar sentido al mundo de manera

1 Doctora en Ciencias de la Gestión por ESADE Business School de la Universidad Ramon Llull (España), Doctora en Filosofía por la Copenhagen Business School (CBS) (Dinamarca), Master de Investigación por ESADE Business School, MBA por Insead (Francia), Master CEMS, Licenciada en Dirección y Administración de Empresas y MBA por ESADE Business School de la Universidad Ramon Llull (España).

práctica. La entrada de la IA en nuestra vida nos está pidiendo repensar qué significa ser humano, qué humanidad queremos construir y, por tanto, cómo deberíamos orientarnos como especie. Así, estas reflexiones no son un ejercicio teórico, alejado de la vida diaria que estamos viviendo, sino que inciden a nivel práctico en cómo el conocimiento, la ciencia y la tecnología se desarrollan con IA y cómo eso tiene un impacto en cómo pensamos y organizamos nuestras vidas, y por tanto, en la orientación de los sistemas de valores colectivos. Estos dos elementos, los valores y el conocimiento, orientan nuestra prosperidad y, más brutalmente, nuestra supervivencia; por tanto, entender cómo pensamos sobre esta tecnología, la IA, es crucial.

Hace más de 25 años que, a nivel práctico, me interesé por entender qué significa conocer el mundo, qué es la realidad, es decir, la epistemología. En aquellos momentos, todo era mucho más simple; me interesaban los textos de sabiduría, me explicaban una manera diferente de entender el mundo. Al inicio, de manera fortuita, investigando en una librería en Londres con el Advaita, después con el Budismo Zen y, posteriormente, cinco años más tarde, empecé a ampliar, gracias a la guía y al trabajo metódico del CETR, mi conocimiento de todas las demás tradiciones, como por ejemplo, la cristiana, la musulmana y la taoísta.

La extraordinaria aportación del CETR ha sido explicitar cómo separar las configuraciones culturales, los sistemas de valores, de esas tradiciones, de la sabiduría que quieren comunicar. Para hacer esto, al inicio, necesitábamos realizar un trabajo epistemológico, con el fin de poder profundizar en este trabajo de comprensión personal de la sabiduría, así como en la capacidad para comunicarlo a otros. Poder entender y desarrollar la Cualidad Humana, eso que intentan mostrarnos las tradiciones, esta capacidad humana de vivir en el mundo más allá de las comprensiones automáticas provenientes de la estructura psicológica, resultado de toda una serie de azares que me sitúan en un espacio-tiempo determinado, dentro de una cultura, dentro de una familia, con otra serie de azares personales, experiencias, memorias, expectativas, deseos y miedos, ha sido un gran regalo.

Así, para poder realizar esta investigación de los textos de sabiduría, es necesario cuestionar las hipótesis de base desde donde se hace la lectura;

de lo contrario, trabajaremos sometidos a nuestra configuración cultural, una prisión epistemológica de la que no se puede escapar, a no ser que uno la conozca. Una de las creaciones fundamentales de Marià Corbí (1983) ha sido establecer como hipótesis de su trabajo de investigación que la antropología del animal humano debe basarse en hechos científicos, no en especulaciones. Así, comenzar el desarrollo teórico definiendo al animal humano como constituido por el lenguaje, es decir, modelado en su comprensión del mundo y en su actuación dependiendo de las necesidades de supervivencia del colectivo, significa que el animal humano es flexible en lo que valora y, por tanto, en cómo actúa. Flexibilidad no quiere decir que todo valga o que todo sea relativo. Significa que el animal humano, como colectivo, creará aquellos valores que le aseguren la supervivencia, de manera más o menos consciente. Si no es capaz, entonces esa civilización, o incluso la especie, no prosperará.

Esta flexibilidad en la valoración y, por tanto, en la actuación, es posible porque los humanos tenemos un sistema de comunicación en el que el significado de las cosas del mundo se coloca en el significante (los sonidos de las palabras) y se separa directamente de su significado, porque se crea una distancia objetiva que hace que la cosa en sí pueda tener muchos significados (de Saussure, 1959). Este tipo de lenguaje es muy especial; ningún otro animal terrestre tiene esta estructura. El resto de animales pueden tener lenguajes, pero ninguno está constituido en el formato humano como una tríada que se representa en una relación humano-palabra-mundo. Esto es lo que nos otorga nuestra flexibilidad y la razón por la cual hemos desarrollado, entre otras cosas, ciencia, arte y religiones.

Ha sido impresionante constatar cómo esta decisión metodológica, de definir la antropología siguiendo los hallazgos reconocidos a nivel antropológico y lingüístico, continúa siendo tan ignorada en el mundo académico y social como hace más de 40 años, cuando Corbí ya la postuló. Este principio antropológico, que define al ser humano como constituido por el lenguaje, aunque de manera minoritaria está presente en las ciencias sociales —principalmente en la sociología del conocimiento, la historia del conocimiento y los estudios tecnocientíficos—, no ha llegado a tener suficiente impacto como para convertirse en rectora de nuevas propuestas

que alcancen el núcleo de la filosofía, ni de la ética, y tampoco, por supuesto, de la religión. Evidentemente, al no llegar al meollo de las disciplinas, no se han podido desarrollar las consecuencias sociales, políticas y económicas de esta conceptualización, aunque esté presente a nivel teórico. Todavía, a nivel social, cultural e incluso en la mayoría de las teorizaciones de las ciencias sociales, se piensa o directamente no se cuestiona la hipótesis inicial del animal humano como un compuesto de cuerpo y razón, o cuerpo y espíritu. En el mundo de la IA, como explicitaremos más adelante, se conceptualiza al ser humano como inteligencia —digamos que es una destilación del concepto de "razón" y "racionalidad"—, como un conocimiento desligado de la experiencia humana, diríamos como un oráculo o como un Dios, como si el conocimiento estuviera definido en un ámbito al que se pudiera acceder solo con la inteligencia. Utilizar hipótesis como cuerpo y razón o cuerpo y espíritu es construir la ciencia desde la metafísica. Si estas hipótesis están en la base de todas nuestras creaciones teóricas, entonces todas nuestras creaciones científicas son claramente vulnerables. Esto es un problema grave.

Considerar al animal humano como constituido por el lenguaje sería una revolución copernicana en las ciencias sociales, pero claro, entonces disciplinas como, por ejemplo, la ética, quedarían relegadas a disquisiciones de diletantes, es decir, no relevantes, porque se haría evidente que las formulaciones racionales de valor —lo que actualmente es la ética en nuestra sociedad— son tan efectivas como rezar a los dioses del Olimpo[2] Dejaríamos de invertir energías y capital intelectual en disciplinas como la "ética de los negocios" o la "ética de la IA" y nos enfocaríamos en entender cómo crear motivaciones de valor que fueran efectivas en las organizaciones, a nivel social y político, y cómo de ahí podría derivarse una ética[3].

---

2 Si el ser humano es un animal racional, entonces los principios de actuación racionales pueden guiar al ser humano. Pero si conceptualizamos que el ser humano es un animal, entonces el eje de la actuación animal es el sentir, como en todos los demás animales. Evidentemente, en el caso del ser humano, es un sentir que tiene lógica, causa-efecto, incertidumbre, etc. Por lo tanto, para orientar al animal humano no es suficiente la racionalidad cuando está totalmente subordinado al sentir.

3 Hoy en día, la ética se presenta como una formulación racional desligada del sentir de un proyecto humano compartido.

Comprender al ser humano como constituido por el lenguaje ya transpira una flexibilidad inmensa. Si lo comparamos con otras especies animales, vemos que es una característica única: una flexibilidad extraordinaria y la fuente de nuestro éxito como superdepredadores en la Tierra. Estar constituido por el lenguaje significa que esta constitución —lo que somos, lo que el ser humano considera adecuado como actuación— tiene la capacidad de cambiar infinitamente. Nunca como una decisión arbitraria, sino siempre sujeta a las formas de vida. Las diferentes constituciones humanas se reflejan en las valoraciones de los grupos humanos y en los significados de las palabras. Esta flexibilidad de adaptación y de creación de nuevas posibilidades no existe en ningún otro animal terrestre. Es fundamental y delicada. Si los humanos no están constituidos adecuadamente, no serán capaces de sobrevivir en su entorno. Esto ya ocurre. En nuestra especie conviven simultáneamente muchas configuraciones humanas.

Por tanto, esta constitución —configuración de las valoraciones humanas— se define a nivel colectivo y no es resultado de un proceso participativo, ni de una encuesta, ni de un mercado, ni de la decisión de un primer ministro, un dictador o un rey. Es el resultado de muchos azares, de contribuciones dispares de humanos —unos más poderosos que otros—, de procesos largos que van más allá de una vida humana y, sobre todo, hasta ahora, de una falta de conocimiento de que todo esto estaba ocurriendo. Me atrevería a decir que tenemos esta falta de conocimiento porque mantenemos creencias sobre lo que es el ser humano basadas en teorías metafísicas, como si viniera del cielo: somos un "espíritu", somos una "razón", somos una "inteligencia". Si pensamos que las configuraciones valorativas nos vienen dadas o reveladas por la naturaleza de las cosas, no hay espacio para imaginar nuevos tipos de valoraciones ni la necesidad de crearlas.

Todos los animales están plenamente configurados para sobrevivir gracias a la información genética que heredan; nosotros no. Los humanos tenemos una "deficiencia" que es un tesoro. No tenemos configurado genéticamente cómo debemos actuar para sobrevivir, ni cómo debemos organizarnos, ni cómo debemos relacionarnos entre nosotros (Corbí, 1983). La teorización de Marià Corbí, que aún no se ha materializado en sus consecuencias a nivel

social, nos permite ser conscientes de que estas constituciones humanas deben cambiar —y cambian—, y que las configuraciones de valores deben ser adecuadas a los ejes sobre los cuales un colectivo sobrevive; de lo contrario, no se puede prosperar. Por tanto, **la prosperidad no es una meta tecnológica o cultural**, sino una meta que tiene que ver con la **capacidad de constituir** humanos capaces de prosperar en condiciones adecuadas, es decir, **con la conjunción axiológica y de modo de supervivencia** (tecnológico y cultural) que permite a las comunidades prosperar. Así, nuestras valoraciones colectivas, como lo que definimos como inteligencia, no son atemporales, sino que dependen de que somos un animal humano y de cómo sobrevivimos en el entorno.

Este elemento axiológico, aparentemente irrelevante en nuestro mundo científico y cultural, es en realidad una fuerza poderosa e indomable que condiciona la capacidad de comunidades e individuos para prosperar. Dedicar esfuerzos y recursos a entender estas configuraciones axiológicas que constituyen a los humanos es tan importante como la frontera tecnológica de la física cuántica, la biotecnología o la Inteligencia Artificial, por mencionar algunos campos científicos. Incluso me atrevería a decir que **es más importante que nuestras creaciones científicas, porque las configuraciones axiológicas estructuran cómo nuestras creaciones son utilizadas, y teniendo en cuenta que nuestra ciencia y la tecnología son cada vez más poderosas, esto significa que debemos ser capaces de construir configuraciones axiológicas que beneficien la vida y no que conduzcan a la extinción de la humanidad.**

Desde otras disciplinas se ha llegado a conclusiones similares. Disciplinas como los estudios de la filosofía de la ciencia, la historia de la ciencia y la sociología del conocimiento se han interesado en entender cómo los humanos construimos el conocimiento, cómo es diferente de las creencias y cómo está relacionado con las comunidades humanas que han creado ese conocimiento. En los años 70, Thomas Kuhn (1970) comenzó a sacudir los pilares de la construcción del conocimiento científico como descripción del mundo. Fue Boyle, en 1660, quien definió los ejes racionales que han desarrollado el conocimiento científico tal como lo conocemos hoy. Según Boyle, para poder entender el mundo, para poder describirlo fielmente,

debemos seguir un método: 1. hacer experimentos - tecnología material, 2. la necesidad de que existan testigos - tecnología literaria, 3. que los testigos sean independientes - tecnología social (Shapin, 1984). En el siglo XX, la sociología de la ciencia ha concluido que el conocimiento científico no refleja la naturaleza, tal como postulaba Boyle, sino que es una herramienta práctica para "hacer comprensible" el mundo (Bloor, 1991). Así, no es una descripción, sino más bien una modelación: si el conocimiento funciona en la práctica, entonces decimos que eso es "verdad", si no, es "falso". Lo que se "descubre", la tecnología que se desarrolla, no es la "mejor" tecnología, como si existiera un camino predeterminado de desarrollo, sino que depende de factores históricos (Callon, 2010), y por tanto, la ciencia y la innovación no tienen una dinámica propia, sino que están totalmente condicionadas por la situación social. La visión de la ciencia como autónoma y desligada de la sociedad tiene implicaciones políticas porque permite mantener y perpetuar el status quo dominante (Law, 2017).

Desde una aproximación feminista, Haraway (1988) ha demostrado empíricamente cómo los métodos y procedimientos para realizar el trabajo científico llevaban otras agendas, valores, conocimientos, etc., que afectaban lo que se creaba y cómo se creaba; es decir, nada era tan aséptico, neutro y fuera de disputa tal como Boyle postulaba. Todo el conocimiento y todos los métodos están "situados", reflejan el lugar y reproducen las agendas sociales. En nuestra sociedad pensamos que la ciencia es "objetiva", "la verdad", pero Haraway (1988) argumenta que "esa mirada desde ningún lugar" es como un "truco de Dios" que puede verlo todo desde ningún sitio, imparcial, sin tener en cuenta todos los elementos humanos que influyen en su desarrollo. Los términos "objetividad" y "subjetividad" pierden sus raíces, aunque Haraway mantiene la palabra objetividad siempre teniendo en cuenta dos aspectos: (1) reconocer que el quehacer científico está situado y (2) analizar críticamente este hecho "situado". Por lo tanto, para Haraway (1988) se debe descartar la ficción de que la ciencia sea "neutral" y, por tanto, que sea una descripción de la realidad, pero no termina de romper con el concepto de objetividad, donde invariablemente se asume que una descripción fidedigna es posible, aunque ella no lo crea. Y es que la ruptura con la posibilidad de una descripción acreditada "objetiva" puede acarrear consecuencias no deseadas. Si no hay nada "creíble", ¿qué queda?

Muchos filósofos se han dedicado a reflexionar sobre el impacto de esta búsqueda y su incidencia a nivel político y en los valores.

Críticos de esta postura ven un ataque a la racionalidad y, por tanto, según ellos, un retorno a la irracionalidad. Pero la repercusión de estas investigaciones lleva a la conceptualización del ser humano como un actor material-semiótico, donde lo material y aquello que da y crea sentido están íntimamente unidos, y es gracias a la interacción social, mediante el habla, que las fronteras de lo que es son creadas. Por tanto, la ciencia es una tecnología semiótica. Y lo que llamamos objetividad es solo una alegoría de la ideología que gobierna (Haraway, 1988). La mirada "objetiva", por tanto, está subordinada al orden imperante, no es una "descripción" "auténtica" de la realidad. Esta disciplina, resultado de las investigaciones empíricas sobre la ciencia, ha desarrollado todo un eje teórico donde se hace evidente, a nivel práctico, una epistemología no mítica y una concepción del ser humano fuera de las coordenadas sociales comunes de cuerpo y alma o cuerpo y espíritu. Podríamos sugerir que el estudio del trabajo científico, desde la mirada feminista y la epistemología del conocimiento, les ha llevado a cuestionar la misma realidad de la "existencia" del individuo, de la naturaleza y de la sociedad. La teoría de las redes de actores (*"Actor Network Theory"*, ANT en inglés) define esta "existencia" semiótica y, por tanto, modelada, y, por tanto, comprendida, transforma nuestra realidad del mundo (Latour, 2005).

Estas reflexiones continúan aún. Law (2017) argumenta que la relación entre la sociedad y la ciencia es bidireccional y que estudiar esta situación es importante. Invariablemente, estas aproximaciones que sacuden la hipótesis de una descripción veraz de la ciencia han creado una fuerte oposición, y autores como Harding (2017) han dedicado extensas reflexiones a explicitar y argumentar por qué no se puede hablar de objetividad ni de subjetividad para entender cómo el ser humano comprende el mundo. La física no es una opinión, pero al mismo tiempo tampoco es absoluta, una verdad que describe el mundo[4]. Ha sido difícil conceptualizar de una

4 Tenemos la física newtoniana, la física cuántica, con diferentes hipótesis y modelos del mundo que funcionan.

manera rigurosa y conceptualmente clara cómo el conocimiento es una modelización del mundo que funciona; por lo tanto, estrictamente no es ni objetiva ni subjetiva, porque no tenemos esa posibilidad. La objetividad asume que nuestra comprensión de la realidad es una descripción de la realidad, y eso no es posible.

Marià Corbí ha profundizado en sus estudios sobre la comprensión de los valores. El mundo de los valores también es una modelación humana; no existe un mundo de valores definido, veraz, auténtico o cierto. Las configuraciones axiológicas que han funcionado son los sistemas de valores que han permitido a la especie progresar. Simplificando el desarrollo teórico de Marià Corbí, podríamos decir que hasta ahora hemos vivido tres tipos de configuraciones axiológicas que han organizado la simbiosis de los humanos: el mito, las religiones y las ideologías. El orden cronológico que he descrito depende de cuándo fueron efectivas de manera dominante a nivel social. Hoy en día, solo tribus perdidas cazadoras-recolectoras pueden tener mitos. Las religiones, nacidas de la revolución agrícola, aún son efectivas en algunos países no occidentales como estructuradoras incuestionables de la realidad, y las ideologías, bueno, quizá solo en Corea del Norte es posible que se mantengan como válidas. En el resto del mundo estamos desestructurados axiológicamente: hay una mezcla de religiones, ideologías y pensamiento alternativo, todo mezclado, resultado de la globalización. Hay muy pocos lugares en el mundo lo suficientemente aislados y con formas preindustriales de supervivencia que no tengan este "menú" axiológico, es decir, esta mezcla de posibilidades para ayudar al colectivo a prosperar.

Hoy en día, la prosperidad del colectivo humano no proviene de la capacidad de cazar y recolectar frutos, ni de la fertilidad de las tierras y el acceso al agua y al buen clima, ni del acceso a energía barata. Todos estos elementos han sido motores de prosperidad, pero ahora, en las condiciones en las que nos encontramos, el motor de la prosperidad es la capacidad de generar conocimiento técnico y científico de manera continua: la capacidad de innovar. Cada modo de supervivencia exige que los humanos cambien sus motivaciones para poder responder a los retos que presenta la nueva forma de supervivencia colectiva.

Por primera vez en la historia humana, la manera de sobrevivir —esta necesidad de innovación constante— nos exige entender, a nivel práctico, que el mundo en el que vivimos está modelado, constituido semióticamente, y que, por tanto, nos permite desarrollar la capacidad para crear e innovar. Una de las numerosas consecuencias de la vida actual es que la presión por innovar nos obliga a desarrollar la capacidad de trabajar en equipo de manera interdependiente, y, por tanto, el eje del éxito no es el individuo ni la competencia. Estas afirmaciones no son el resultado de una posición valorativa, crítica o moral, sino una secuela de la lógica de innovar y del crecimiento acelerado de las ciencias y las tecnologías. Por tanto, defender y comunicar una antropología científica como base de las formulaciones teóricas —es decir, que el ser humano modela el mundo y está constituido por el lenguaje, y no por hipótesis metafísicas como las formulaciones de que el ser humano es cuerpo y razón— debería ser una noción cultural compartida por los humanos, con el fin de asegurar la prosperidad del colectivo. Una epistemología de la ciencia y de la vida no mítica, entendiendo que las formulaciones científicas y las formulaciones de nuestra vida cotidiana no describen la realidad, sino que son modelaciones adecuadas en una situación determinada, es uno de los ingredientes fundamentales que permite que las capacidades creativas humanas se expresen y se desarrollen con la máxima potencia y diversidad. Las repercusiones de estos cambios son monumentales.

Los trabajos de Marià Corbí han permitido explicar la fuente de la creatividad humana, explicar las configuraciones axiológicas pasadas y entender los mitos, las religiones y las ideologías sin quedar atrapados en elementos culturales ni míticos. Así, por primera vez en la historia humana, podemos plantearnos a nivel colectivo la orientación valorativa humana como una construcción pensada, reflexionada, orientada y plenamente consciente. Esto es una revolución.

Aclaro que hoy en día todavía existen religiones, mitos e ideologías, y que hay mucha población que continúa sometida a estos tipos de configuraciones valorativas. En el mundo occidental, sin embargo, estas configuraciones ya no son totalizadoras, sino que las personas se someten a ellas de manera voluntaria. Digamos que hay un "menú" donde elegir. En el

pasado, esto no era posible: cada una de estas configuraciones humanas era impuesta, generalizada a la comunidad a la que uno pertenecía, y aseguraba la supervivencia del colectivo. Por ejemplo, aunque hoy en día existan muy buenos cristianos o musulmanes, la religión está desacreditada. Lo mismo ocurre con las demás religiones, con las ideologías y con los mitos. No era así en 1340, cuando los campesinos de Manresa construyeron una acequia para poder regar sus huertos o morían de hambre. El obispo no quería que un menor caudal del río hiciera que su molino, para moler cereales, pudiera dejar de funcionar algunos meses del año por falta de potencia hídrica. Para detener la construcción de la acequia, el obispo excomulgó a toda la ciudad; no se celebró ningún rito cristiano durante más de cinco años. Fue una crisis social de tal magnitud que solo pudo revertirse con la llegada de un nuevo obispo que recibió "un milagro", "de la luz", que venía del monasterio de Montserrat y que iluminó el rosetón de la iglesia de Manresa —una hazaña totalmente imposible por razones físicas—, pero claro, era un milagro, la única manera de justificar que la opinión del representante de Dios había cambiado.

Aquella situación se recuerda cada año después de casi 700 años, tal fue el efecto en generaciones de manresanos. En aquel tiempo, toda la vida se vivía siguiendo los mandatos de la Iglesia; no se podía escapar de las configuraciones axiológicas, y los manresanos tuvieron que vivir la complejidad de un obispo que amaba el dinero más que la vida y pronunciaba discursos en contra del pueblo y a favor de su propio beneficio material. Queda claro que solo "un milagro" podía justificar revertir la situación.
Cuando los individuos pueden decidir qué "religión seguir", como si fuera un menú en un restaurante, ya nos están mostrando que las religiones han dejado de ser configuraciones a nivel social. También nos permite definir nuestra situación colectiva actual como desarticulada axiológicamente y, por tanto, muy diferente de cómo han estado configuradas las poblaciones humanas anteriormente.

ESQUEMA DEL MILAGRO DE LA LUZ

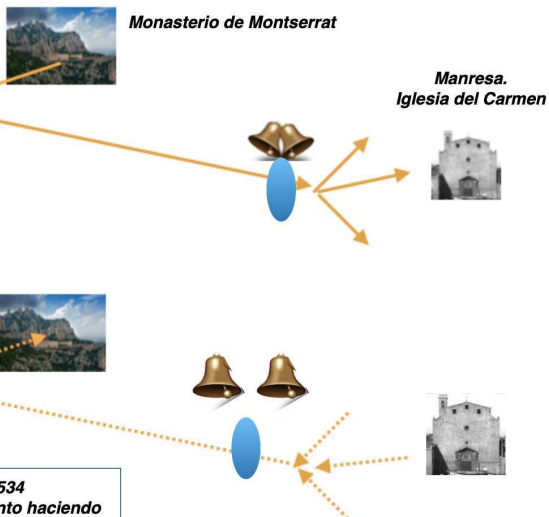
Del Monasterio de Montserrat a la Iglesia del Carmen de Manresa
Las campanas repican solas
Una Luz entra por una vidriera y se divide en tres rayos

Vidriera de la Iglesia del Carmen
Por donde entró y salió el rayo de luz

Rayo de luz de Montserrat a la Iglesia del Carmen. La luz tuvo que realizar un ángulo par salir del Monasterio y llegar a Manresa. Si no hubiera realizado este ángulo, la luz hubiera agujereado la montaña.

Rayo de luz de regreso a Montserrat. También ha realizar un ángulo

De la Iglesia del Carmen de Manresa al Monasterio de Montserrat
Las campanas dejan de repicar
Los tres rayos de Luz se concentran en uno solo que regresa a Montserrat



21 de Febrero de 1345
Fecha que la Iglesia afirma se produció el Milagro

Año 1534
Primer documento haciendo referencia al Milagro

Vivimos en sociedades sin una estructura axiológica colectiva. A nivel social se ha comprendido el fracaso de los proyectos ideológicos y religiosos; incluso la ideología dominante actual, capitalista-neoliberal, ha perdido reputación. Parecería que no puede haber lugar para nuevas formaciones axiológicas, en "formato" de ideología o religión, teniendo en cuenta que su funcionalidad axiológica era adecuada a una determinada forma de supervivencia[5].

Al mismo tiempo, nuestra comunidad global es un terreno fértil para una multitud de nuevas propuestas axiológicas, destinadas a dar respuesta a las necesidades humanas de orientación. Si entendemos al ser humano como constituido por el lenguaje, eso significa que necesitamos estas configuraciones como el aire que respiramos. Por esta lógica axiológica, hoy en día tenemos a nuestro alcance un "mercado" axiológico, con una multitud de creencias, ideologías y religiones. El individuo elige, dentro de estas posibilidades míticas que le describen una verdad, una certeza. Parece que este razonamiento que ofrezco conduzca a la creación de un nuevo sistema totalizador y único. Pero estoy intentando explicitar que debemos crear un nuevo concepto axiológico, alejándonos de las palabras y de lo que han sido los mitos, religiones y axiologías. Todas tienen en común que pretenden describir la realidad y orientar a los humanos de la manera correcta. Debemos transformarlo para poder enfocarnos en este reto que consiste en crear configuraciones axiológicas colectivas para la sociedad de innovación continua en la que vivimos, asumiendo plenamente la incertidumbre y, por tanto, la conciencia de los peligros. El concepto que ha desarrollado Corbí (2020) ha sido el de proyectos axiológicos colectivos, en los que se entiende que las configuraciones axiológicas son modelaciones que deben construirse.

Solo si, a nivel colectivo, se tiene plena conciencia de las demandas axiológicas adecuadas para la supervivencia, seremos capaces de prosperar y las propuestas serán no míticas. Entonces no será el individuo quien se vea forzado a elegir, sino que, a nivel social y comunitario, la orientación vendrá guiada, comprendida como una modelación. Como sociedad,

5 En el caso de las ideologías, son el resultado del proceso de industrialización.

todavía somos incapaces de estructurar estas configuraciones de valores orientadas de manera sensitivo-racional, conectadas con la forma en que se sobrevive.

Las narrativas de los creadores de la IA. La inteligencia.

Noviembre de 2022: fecha de nacimiento de ChatGPT para el público mundial. Con un enlace en internet, cualquier persona podía interactuar con un modelo de lenguaje desarrollado por la empresa OpenAI. Estos modelos de lenguaje, entrenados con todo el texto de internet y con todas las transcripciones de audio y video de la red, son capaces de emular las conversaciones humanas y, por tanto, parece que son inteligentes.

Lo que llamamos Inteligencia Artificial, IA, es un término comercial que engloba muchos tipos de tecnologías. La que está en boca de todos es la IA generativa, la que está en la base de los "chatbots"; esta es la IA que está triunfando ahora. Es una tecnología que no es funcional: no podemos asegurar que sus respuestas, sugerencias, etc., sean correctas, por lo tanto, por definición, no es segura, y no podemos asegurar que sea efectiva ni tampoco eficiente (Niederhoffer y otros, 2025).

OpenAI, en noviembre de 2022, estaba valorada en 29.000 millones de dólares, y en octubre de 2025 está valorada en 500.000 millones de dólares (Kinder y Hammond, 2025), 16 veces más. Su valor ha aumentado de manera estratosférica porque han sabido vender la historia, de forma efectiva, de que están construyendo una superinteligencia. ChatGPT es el primer paso, y según su director general, Sam Altman, quien sea capaz de lograrlo será quien controle a la humanidad y, al mismo tiempo, será capaz de resolver todos nuestros problemas[6]. Esta idea fue descrita inicialmente por I. J. Good (1966):

"Definimos una máquina ultrainteligente como una máquina capaz de superar con creces todas las actividades intelectuales de cualquier ser humano, por muy ingenioso que este sea. Dado que el diseño de máquinas es una de

6 https://blog.samaltman.com/

esas actividades intelectuales, una máquina ultrainteligente podría diseñar máquinas aún mejores; esto daría lugar, de manera incuestionable, a una 'explosión de inteligencia', y la inteligencia humana quedaría muy atrás. Así, la primera máquina ultrainteligente sería la última invención que la humanidad necesitaría jamás, siempre que fuera lo suficientemente dócil como para indicarnos cómo mantenerla bajo control. Es curioso que este punto se haya mencionado tan pocas veces fuera de la ciencia ficción. A veces vale la pena tomarse la ciencia ficción en serio."

Este es uno de los argumentos clave para impulsar las inversiones en inteligencia artificial y los estudios sobre los riesgos existenciales de esta tecnología, en detrimento del enfoque en los problemas actuales que esta diversidad de tecnologías está causando.

¿Es posible este pensamiento de la inteligencia como producto? La inteligencia siempre es una evaluación respecto a un entorno dinámico y siempre depende de unos perceptores, de un cuerpo. Es decir, la inteligencia de una hormiga es diferente a la de un humano y siempre está en relación con su entorno. ¿Cómo podemos hablar de una ultrainteligencia como si fuera un producto, una entidad, cuando es una respuesta relacional y dependiente de unas dinámicas específicas de supervivencia? ¿Cómo podemos "vender" una ultrainteligencia que hemos fabricado y "vendemos" como una inteligencia comparable a la humana? No puede serlo nunca.

OpenAI, al igual que otras startups como Anthropic, Perplexity AI o MidJourney, sigue las dinámicas estándar del "capital riesgo". Son capaces de crear productos potencialmente interesantes, pero todavía no hay un negocio real, ni ingresos ni beneficios. Para conseguir financiación, las startups presentan un futuro, y los inversores "apuestan" por una historia, por una narrativa. Se apuesta por la persona que es capaz de generar una narrativa creíble del futuro, porque el negocio aún no existe. Se usa la metáfora de que lo importante es el jinete —el impulsor—, no el caballo —la empresa—. Si la persona es la adecuada, sabrá encontrar la manera de transformar o recrear la narrativa de tal modo que el negocio sea un éxito. Por eso, saber crear una historia es lo más importante.

Así, se crean diversas historias para impulsar estas empresas. He hecho una recopilación no exhaustiva, pero en mi opinión significativa, de las narrativas más comunes.

La primera, *"el poder de la IA es tal que puede acabar con la humanidad"*. Destacar el poder de la IA es una manera de adquirir notoriedad, de crear nuevas conversaciones, discursos. Tenerla siempre presente a nivel público. Los ciudadanos, los políticos, las instituciones y las empresas se ven obligados a formular su posición. Esto genera más probabilidades de que los productos y servicios de estas start-ups tengan compradores.

Parece contradictorio que alguien que quiera vender un producto/servicio destaque una parte negativa del producto/servicio, en este caso el "poder de la IA para acabar potencialmente con la humanidad". Los creadores de la IA presentan la posibilidad de que pueda ser utilizada de manera criminal y que se debe regular su uso. Pero con esta narrativa, no se cuestiona que simplemente este producto/servicio no debería existir (Golumbia, 2022). O que este producto/servicio no funciona, sino que se asumen estos dos aspectos: (1) su legitimidad para ser comercializado, (2) su funcionalidad, y se enfoca toda la reflexión en el uso negativo de la tecnología.

Ya sé que esto parece muy poco lógico. Me gustaría comentar este ejemplo: https://ai-2027.com/ . Los autores de esta "investigación" defienden que es una investigación aunque se lee como un ejercicio de ciencia "ficticia". Los autores argumentan que han abandonado carreras lucrativas trabajando para las start-ups de IA para dedicarse a crear narrativas como esta que advierten de los posibles futuros negativos que podemos vivir. Son "investigadores" que tienen conciencia, que se preocupan por el beneficio de la humanidad, son "los buenos". También presentan otras credenciales, para que creamos estas narrativas, afirmando que ya han demostrado en el pasado, con análisis similares, que sus predicciones son correctas o que han trabajado en start-ups sobre el tema de la seguridad de la IA.

En esta "creación" del futuro que imaginan, hay que resaltar que: **definir y preocuparse por el futuro genera poder en el presente**. ¿Cómo ocurre esto? Primero, se presenta el futuro como el **resultado de una**

inevitabilidad técnica acelerada que lleva, más temprano o más tarde, a la creación de una superinteligencia. Aquí no hay ningún reconocimiento de toda la disciplina de la ciencia del conocimiento ni de la interacción con la sociedad y la política. Segundo, debido a las hipótesis axiológicas de este grupo de "investigadores", esta superinteligencia es definida como dominante, colonial y exclusiva: *"quien controle esta superinteligencia tendrá el poder".*

Así, esta narrativa **nos impide discutir estas dos hipótesis de base y nos centra, nos obliga, a pensar y reflexionar en sus predicciones.** Desde mi punto de vista, se trata de una trampa: quedamos atrapados en una visión del futuro en la que la tecnología está determinada por el capital riesgo, aunque ellos lo presenten como si fuera fruto de la autonomía de la innovación científica. Es evidente que los autores pasan por alto más de setenta años de estudios sobre el desarrollo científico y tecnológico. Esta mirada sobre la IA mantiene como *"lógica, racional e inevitable"* la actual evolución de la IA, como si los intereses dominantes no tuvieran ninguna influencia. Un planteamiento, como mínimo, sorprendente.

Hay muchos investigadores que desarrollan lo que ellos llaman *"seguridad de la IA",* centrados en el estudio y mitigación de riesgos existenciales —extinción de la especie humana—. Su argumentación, heredera de las teorías filosóficas utilitaristas de Peter Singer (1993), es seguida fervorosamente por muchos desarrolladores de IA en Estados Unidos. Uno de los ejes de esta comunidad es https://www.lesswrong.com/ y su enfoque en el *"longtermism"*, es decir, en el efecto futuro de estas tecnologías (Torres, 2021).

Parece extraño que, queriendo defender el beneficio de la humanidad, me centre en criticar o atacar estas posturas, que aparentemente también defienden el *"beneficio de la humanidad":* la seguridad de la IA es un motivo loable. Pero lo que realmente están haciendo es defender una visión del futuro en la que se acepta la inevitabilidad de un cierto tipo de IA, y la "inteligencia" como dominadora y explotadora, y nos roban la posibilidad, incluso, de imaginar otro tipo de tecnología.

Otra de las narrativas que circulan es: *"Nosotros, los ingenieros y programadores, somos muy inteligentes, porque somos capaces de crear esta inteligencia artificial; por eso, lo que decimos está cargado de razones y es muy inteligente y razonable"*. Por esta razón, todas estas propuestas son ampliamente amplificadas: se considera que quienes las impulsan son muy inteligentes, justificadas por gente muy inteligente; al fin y al cabo, ¡son los titanes del mundo! y por lo tanto merecen su amplificación en los medios tradicionales.[7] Esta "inteligencia" sigue los parámetros de valores de personas muy específicas: los programadores, los ingenieros, las personas que trabajan en el capital riesgo en Silicon Valley.

El tema de la inevitabilidad de la IA también se utiliza como un argumento de dominación geopolítica: *"Si no la creamos nosotros* —los americanos—, *la crearán los chinos"*. Pero esta *"inevitabilidad"* es una ficción; no es así como la ciencia y la tecnología se desarrollan.

Otro aspecto de la inevitabilidad de la IA y del estudio de la *"seguridad de la IA"* es cómo se traslada la mirada al futuro y, por lo tanto, se olvidan y no se tienen en cuenta los problemas que actualmente ya está generando la IA. Por ejemplo, las redes sociales generan múltiples problemas con sus sistemas recomendadores (muy poca inteligencia), pero que tienen un gran impacto.

Otra narrativa de los titanes actuales de la IA es la defensa desenfrenada del poder de la tecnología para transformar el mundo y la suficiencia moral y ética de esta posición. Cualquier persona que se oponga a esta visión es tachada de ignorante y enemiga. No solo se defiende este futuro tecnológico, sino que se defiende toda una racionalidad: el mercado es la manera más racional de tomar decisiones; la política es un estorbo; la burocracia debe destruirse[8]. En definitiva, una defensa del capital riesgo para que no tenga ningún tipo de obstáculo y pueda continuar funcionando. Todo *"en beneficio de la humanidad"*. Aquí se defiende el mantenimiento y profundización de la economía capitalista neoliberal de mercado. El

7 https://www.nytimes.com/2025/04/11/podcasts/hardfork-tariffs-ai-2027-llama.html

8 https://a16z.com/the-techno-optimist-manifesto/

objetivo de desarrollar esta inteligencia artificial es tan importante que los problemas que puedan surgir por intentar alcanzar este objetivo no son relevantes: problemas climáticos, sociales, geopolíticos, guerras, etc.

Según esta narrativa, tampoco debemos preocuparnos de los problemas que estamos creando con el desarrollo actual de la IA, por ejemplo: *"la explotación humana en la creación de datos* (O'Neil, 2017; Gillespie, 2018; Crawford, 2021), *la automatización de la discriminación* (Eubanks, 2019; Benjamin, 2019; Buolamwini, 2023; Broussard, 2023), *la explotación de la naturaleza (Crawford, 2021)". Una vez alcanzada la meta de una inteligencia artificial general, entonces resolveremos todos los problemas, como por ejemplo la crisis climática".* Sin reconocer que la gran mayoría de los "problemas" humanos no son tecnológicos, sino sociales y culturales. Si, por ejemplo, tuviéramos una IA que fuera un oráculo, del tipo que imagina Elon Musk[9], no resolveríamos los problemas sociales y culturales. Por ejemplo, ahora tenemos el conocimiento para resolver la crisis climática, pero no hemos desarrollado la voluntad política y social.

Los "gurús" de la IA han creado historias impactantes explicitando las distopías de control de la IA sobre los humanos, y que por tanto se debe regular y evitar la extinción de la humanidad. Pero cuando existen regulaciones de la IA, como la reciente AI Act (2024) de la Unión Europea, todas las empresas estadounidenses que controlan la frontera del desarrollo de la IA están en contra. De hecho, la administración de Trump está presionando para detener la regulación[10]. Parece que Europa está retrasando la implementación de la ley, la AI Act.

La inteligencia como aspecto definitorio de lo que significa ser humano en comparación con los animales también se ha utilizado para "vender" este nuevo producto/servicio. La IA es "inteligente", pero es una inteligencia peculiar: la que ahora domina, la IA generativa, es una mirada estadística al mundo; no tiene nada de inteligencia humana, capaz de comprender

9 https://abcnews.go.com/Business/elon-musk-launches-ai-company-compete-chatgpt/story?id=101210078

10 Haeck, P. (2025, June). EU's waffle on artificial intelligence law creates huge headache. Politico.

cualidades, hacer abstracciones, valorar las situaciones, etc. La IA generativa emula la inteligencia humana sin funcionar como los humanos lo hacen.

Se ha destacado este aspecto, su "inteligencia", porque ayudaba de manera importante a incrementar el interés por productos y servicios que fueran inteligentes. Esta inteligencia se resalta de muchas maneras, normalmente justificando que el modelo de IA es capaz de superar muchos tests, diferentes tipos de pruebas. De hecho, hay una página web (Language Models Arena) que, mediante la participación de los usuarios (https://lmarena.ai/), crea un ranking de los mejores modelos de lenguaje.

Los rankings que se crean, a partir de "benchmarks", es decir, especificaciones técnicas que evalúan aspectos esenciales de los modelos, tienen que ser concretos para ser efectivos en la evaluación, y por lo tanto no son útiles para ofrecer una apreciación real del nivel de inteligencia del sistema de IA. Por ejemplo, el hecho de que un modelo de IA sea capaz de aprobar el examen para ser abogado no significa que pueda ser un buen abogado: hay mucho más en la labor de un abogado que simplemente poseer conocimiento; se requiere estrategia, conocimiento social, etc.

Por tanto, se trata de una "inteligencia", digamos, muy específica: la capacidad de responder tests. Además, para hacerlo aún menos comparable con la inteligencia humana, los creadores de los modelos no pueden asegurar que los datos de los exámenes que realizan los modelos de lenguaje no se hayan utilizado para entrenar el propio modelo[11]. También estos modelos "razonan" de maneras muy específicas; mejor dicho, no son capaces de razonar[12].

Esta idea de que la IA nos proporcionará inteligencia y, por tanto, todos nuestros problemas podrán solucionarse, tiene hipótesis erróneas tanto a nivel epistemológico como a nivel social. Hay muy pocos que argumenten

11 Si un modelo ha sido entrenado, por ejemplo, con las soluciones a preguntas de exámenes, entonces, una vez entrenado, si se le hacen esas preguntas, será capaz de responderlas sin problemas.

12 Horton, M. (2025). The Illusion of Thinking : Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. 1–30.

en contra de estas ideas (Marcus, 2024; Bender y Hanna, 2025; McQuillan, 2022).

Científicos como Geoffrey Hinton, que se dedican a la computación, hacen declaraciones que van más allá de su campo de especialización (Lohr, 2025), y por el hecho de ser científicos de la computación, haber ganado premios Nobel o ser directores generales de estas start-ups tecnológicas, lo que dicen se considera de gran valor. Por ejemplo, Elon Musk[13] defiende que la IA nos dará "la verdad", como si fuera un oráculo. Este es un error epistemológico. El ser humano modela la realidad y, por tanto, por definición no existe "la verdad" a la que un ser con suficiente inteligencia pueda acceder, sino modelaciones que son más o menos efectivas y que van cambiando según nuestras necesidades. Seguramente, este tipo de mirada estadística que nos ofrece la inteligencia artificial nos ayudará a crear modelaciones más efectivas, pero la tecnología que tenemos actualmente siempre está basada en datos generados por humanos y, por tanto, con perceptores limitados y máquinas limitadas que amplifican nuestras capacidades. Siempre, por definición, tendremos límites. La modelación es infinita y los límites también. Nuestra estructura biológica, que requiere una constitución para ser viable a través del lenguaje, nos ha ofrecido esta mirada única al misterio que habitamos. Pensar que la IA nos revelará "la verdad" es una receta para el fundamentalismo y la violencia legitimada por una inteligencia "superior" inexistente.

Deberíamos ser capaces de salir de este tipo de narrativas. La IA tendrá impacto dependiendo de quién la use, cómo la use y cómo esté imbricada en sistemas de influencia o control (Mühlhoff, 2025). Por ejemplo, desde hace más de dos décadas las redes sociales utilizan sistemas recomendadores —un tipo de algoritmo que podríamos llamar de IA, aunque no muy inteligente por estándares humanos— pero con gran impacto. Empresas como Uber los utilizan para crear sus servicios de transporte y generar precariedad laboral; empresas como Meta venden sus servicios y la información recopilada de sus usuarios para desarrollar campañas de

13 https://abcnews.go.com/Business/elon-musk-launches-ai-company-compete-chatgpt/story?id=101210078

manipulación política; empresas como Palantir venden servicios de vigilancia basados en datos públicos, etc.

La conversación actual sobre la inteligencia de la IA está centrada en enfocar nuestra atención en un problema técnico: "cómo hacer que la IA sea inteligente". Pero el impacto de la IA no dependerá de si resolvemos este "problema técnico", sino de cómo utilizamos esta IA en nuestra sociedad, quién la controla y cómo está integrada en sistemas de control e influencia. Por tanto, debe abrirse el espacio para que disciplinas relacionadas con los aspectos humanos (antropología, axiología, sociología, psicología, humanidades, filosofía y economía) ayuden a definir qué tipo de IA queremos. Y estas reflexiones no pueden surgir de los creadores de estos sistemas "inteligentes", porque con sus capacidades y orientaciones específicas y limitadas perpetúan una mirada técnica. Por tanto, son incapaces de aportar soluciones y replican, profundizan y automatizan los problemas actuales.

Las prácticas de los creadores de la IA. Un ejemplo: Google

Me gustaría presentar el ejemplo de Timnit Gebru para ilustrar cómo actúan más allá de las narrativas que impulsa el sector tecnológico de Estados Unidos y qué significa la ética de la IA para los creadores de la IA, en este caso en Google.

La Dra. Gebru es una científica especializada en inteligencia artificial que, desde 2018, trabajaba en Google como colíder del equipo de IA ética junto con Margaret Mitchell. En 2020 escribió, junto con ella y otros autores, un artículo titulado *"On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"* para la conferencia *ACM Conference on Fairness, Accountability, and Transparency,* en el que alertaba sobre los riesgos medioambientales y la perpetuación de sesgos en los modelos de lenguaje a gran escala, especialmente en sistemas como BERT de Google (entrenado con 3,3 mil millones de palabras) o GPT-3 (con medio billón de palabras). Estos modelos utilizan grandes cantidades de datos que no han sido "curados" y, por tanto, **reproducen de manera automatizada todos los sesgos y problemáticas presentes en nuestra sociedad.**

Este artículo, que ya había pasado una primera revisión interna, fue bloqueado por la dirección de Google, alegando que era demasiado crítico y que no contemplaba suficientemente los mecanismos de mitigación de riesgos. El 2 de diciembre de 2020, después de que Timnit Gebru pidiera explicaciones sobre el procedimiento de censura interna, recibió un correo electrónico informándole de que "aceptaban su renuncia", a pesar de que ella nunca había presentado ninguna dimisión.

Este hecho desencadenó una fuerte polémica: cerca de tres mil empleados de Google y más de cuatro mil miembros de la comunidad académica y civil firmaron un manifiesto en su apoyo. Gebru, que ya era reconocida por su investigación sobre sesgos algorítmicos —como el estudio *"Gender Shades"* junto con Buolamwini (2018), donde evidenció los altos errores de reconocimiento facial en mujeres negras—, también había denunciado que Google no tenía una política efectiva de promoción de la diversidad, y escribió en un correo interno que *"no hay ningún incentivo real para contratar a más mujeres o personas subrepresentadas"*.

Ante la protesta, el director general de Google, Sundar Pichai, envió una disculpa interna diciendo: "He escuchado la reacción ante la marcha de la Dra. Gebru alta y clara" y prometiendo un proceso de revisión, pero muchos trabajadores lo consideraron insuficiente. Poco después, Google también despidió a Margaret Mitchell. El 2 de diciembre de 2021, justo un año después de los hechos, Timnit Gebru fundó el *Distributed AI Research Institute* (DAIR)[14], con el objetivo de fomentar una investigación en inteligencia artificial ética, descentralizada y respetuosa con las comunidades más afectadas por la tecnología.

Esta situación nos plantea multitud de preguntas. ¿Verdaderamente es posible un rol ético de la IA en una empresa tecnológica? ¿Por qué nos estamos enfocando en una ética cuando estos productos "inteligentes" quizá ni siquiera deberían existir? ¿Es un rol de estas características —éticas— un trabajo que pueda tener un impacto relevante en cómo se desarrollan estas tecnologías? ¿Por qué existen este tipo de roles en las

14 https://www.dair-institute.org/

empresas tecnológicas? ¿Pueden los equipos éticos funcionar de manera efectiva dentro de las organizaciones? Teniendo en cuenta lo que sabemos sobre los problemas que presentan los modelos de lenguaje que son la base de productos como ChatGPT (OpenAI), Claude (Anthropic) o Gemini (Google), por ejemplo, problemas de repetición y amplificación de sesgos, creación eficaz de desinformación, copia de contenidos con copyright (Hammond y Acton, 2025), falta de explicabilidad de los resultados, facilidad para fabricar información errónea (alucinaciones), el gasto energético exorbitante, etc., ¿qué función tienen estos departamentos de ética para la IA generativa?

En octubre de 2024, un niño de 14 años (Montgomery, 2024), perdidamente enamorado de un personaje de ficción con quien mantenía una relación mediante un servicio de chatbots[15], se suicidó. En la base de este servicio están los modelos de lenguaje que todos usamos, como ChatGPT o Gemini, pero especializados, en este caso mostrando unas características determinadas, basadas en el personaje Daenerys Targaryen de *Juego de Tronos*. El problema, sin embargo, no es solo técnico, sino que tiene que ver con cómo estos productos y servicios interactúan con los humanos y las comunidades que los utilizan. La madre de Sewell ha demandado a la empresa, que continúa ofreciendo una plataforma de intercambio donde hay más de 1000 personajes[16] diferentes para que el usuario elija con cuál interactuar. Los hay de todo tipo, incluso abusadores. Existen muchos servicios similares, algunos se presentan como *"compañeros"*[17] amistosos o sexuales para establecer relaciones enfocadas a las necesidades individuales del suscriptor del servicio. Pero incluso ChatGPT puede facilitar la muerte de un adolescente de 16 años, tal como ocurrió en abril de 2025 (Hill, 2025).

15 https://character.ai/

16 https://character.ai/sitemap/characters_a

17 https://replika.com/

Conclusión

Desde la aparición de las redes sociales hemos visto cómo algoritmos cada vez más sofisticados son capaces, según sus creadores, de *"conectarnos"*, *"establecer relaciones"* y en general *"vivir mejor"*. Solo considerando el sistema de IA generativa, ChatGPT, más de 700 millones de personas envían al sistema más de 18.000 millones de mensajes cada semana (Clark y Nevitt, 2025). Ahora estos algoritmos son muy *"inteligentes"* y pueden *"resolver"* aún más problemas. Pero ¿y si todas estas ventajas fueran realmente una trampa? ¿La etapa inicial de un mundo en el que es imposible comprender verdaderamente a los demás y entender nuestro propio mundo? Eso es lo que argumenta el Center for Humane Technology. Otros investigadores, como Zuboff (2022), describen cómo estas empresas, con sus productos y sus promesas de *"conectarnos"* o de *"buscar"* en la web, en realidad están construyendo constantemente perfiles de quiénes somos con el fin de hacernos más previsibles en nuestras acciones. Venden este conocimiento a anunciantes y así son capaces de proveernos con los mejores productos/servicios por los que pagamos.

Así, por ejemplo, en el caso de las redes sociales, la promoción de información sesgada y polarizante hace que las personas estén más interesadas y, por tanto, pasen más tiempo usando estos servicios, destruyendo —por el beneficio financiero de estas empresas— la comprensión que tenemos del mundo, y con ello el espacio común necesario para la gestión política de nuestra comunidad. Todo con el fin de vender esos datos a anunciantes y hacernos más previsibles. Cada vez que usamos Google, los resultados iniciales que obtenemos dependen del perfil que los algoritmos tienen sobre quiénes somos y cuáles pueden ser nuestros intereses. Por tanto, nos están presentando un mundo totalmente centrado en cómo el algoritmo nos ha evaluado, y de ello también dependerá el precio al que pagamos los productos (Crawford, 2025).

¿Puede la IA, con su *"inteligencia"*, conseguirnos ese mundo que sus creadores prometen? ¿Cómo puede una mirada ética de la IA incidir en esas dinámicas competitivas? ¿No estamos creando una ficción al pensar que elaboraremos una ética que resolverá estos problemas? El objetivo

comercial es que los usuarios pasen el máximo tiempo posible conectados a estos servicios para dejar la mayor cantidad de datos sobre sus intereses, de modo que, en interacciones posteriores, estas empresas puedan vender esa información de los usuarios a otros, como por ejemplo anunciantes, pero también a personas interesadas en influir en las opiniones políticas, como fue el caso de Cambridge Analytica (Wylie, 2019), una empresa que se ha demostrado que, mediante una relación estrecha con Facebook, ahora llamada Meta, fue capaz de evitar que muchos votantes demócratas acudieran a votar y así facilitar la elección de Donald Trump en 2016.

Enfocarnos en desarrollar una ética de la IA, dadas las condiciones actuales de nuestra comprensión de lo que es una ética de la IA, nos lleva a un callejón sin salida (Munn, 2022; Phan y otros, 2022). ¿Por qué seguimos creyendo que la ética de la IA puede solucionar las dinámicas explotadoras de nuestro sistema económico? ¿O que las empresas de IA, con su idea de beneficiar a los humanos, realmente se preocupan por los efectos de sus productos y servicios en los humanos que los utilizan? ¿Por qué imaginamos o explicitamos una ética optativa que no está en consonancia con la ética real, la verdadera, que se destila del proyecto de valores actual? ¿Por qué pensamos que actitudes individuales y voluntarias —basadas en una ética teórica ficticia— pueden contrarrestar el funcionamiento de un sistema que tiene otra ética?

¿Qué puede hacer un departamento de ética de la IA, en una empresa tecnológica, en un entorno enfocado en lograr que el usuario pase el máximo tiempo posible usando el servicio, sino aparentar que se cuida al usuario? La empresa defiende que protege los intereses del usuario, pero en realidad no es así: la empresa tiene un objetivo comercial y obtiene un beneficio directo en función del tiempo que el usuario utiliza el servicio, con el fin de vender sus datos.

Las empresas que comercializan la IA generativa o las redes sociales, con sus algoritmos recomendadores, defienden que los usuarios no están siendo obligados a utilizarlas, sino que eligen libremente. ¿El hecho de que sea una elección libre significa que, por definición, es una elección que beneficia a quien la hace? Los sistemas están creados para hacer la vida

fácil y conveniente, de modo que los individuos acepten invertir su tiempo en estos sistemas de IA; el beneficio se lo llevan las empresas, y los costos y problemas que generan estos sistemas los asume la sociedad (Mühlhoff, 2025).

Cuando un sistema de IA, como por ejemplo ChatGPT, se lanza al mercado, como mínimo debería respetar las leyes actuales. Ninguno respetó ningún derecho de propiedad intelectual (Metz y otros, 2024), ni han tenido que demostrar que son seguros (McCabe, 2024), ni son transparentes en los recursos que utilizan ni en cómo gestionan los datos privados (Magid, 2025). Se han creado y se siguen creando continuamente toda una serie de riesgos que la sociedad debe asumir. Hasta ahora, estas empresas no han asumido ninguna responsabilidad (O'Neil, 2017; Olson, 2024; Schellmann, 2024; Wynn-Williams, 2025).

Pienso que el problema que tenemos es grave: seguimos replicando respuestas que no han aportado soluciones. Me refiero a más de 15 años con los algoritmos recomendadores de las redes sociales, enfocándonos en injusticias puntuales y aproximaciones éticas que crean la ilusión de que estamos haciendo algo. Deberíamos ser capaces de incidir en el desarrollo de la IA en la sociedad, no en el desarrollo técnico, sino en la integración de los sistemas de IA en la sociedad, todos y cada uno de nosotros en la medida en que nos afectan. Como decía Dewey (1927), en las verdaderas democracias los ciudadanos son la única autoridad legítima, porque sus problemas crean el marco en el que puede funcionar toda forma de especialización. En este momento no parece que podamos imaginar futuros compartidos para utilizar las tecnologías cada vez más poderosas que estamos creando en beneficio de la humanidad y de la vida en general. ¿Qué estamos esperando?

Referències:

Bender, E. M., & Hanna, A. (2025). **The AI Con**. Harper Collins.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? FAccT 2021 - *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Benjamin, R. (2019). *Race after technology*. Polity.

Bloor, D. (1991). *Knowledge and social imagery.* University of Chicago Press.

Broussard, M. (2023). More than a Glitch. In *More than a Glitch*. MIT Press.

Buolamwini, J. (2023). *Unmasking AI : my mission to protect what is human in a world of machines.*

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, 77–91.

Callon, M. (2010). Afterword. In A. Feenberg (Ed.), *Between reason and experience: Essays in technology and modernity*. The MIT Press.

Clark, D., & Nevitt, C. (2025, October 7). How AI became our personal assistant. *Financial Times.*

Corbí, M. (1983). *Análisis epistemológico de las configuraciones axiológicas humanas. La necesaria relatividad cultural de los sistemas de valores humanos: mitologías, ideologías, ontologías y formaciones religiosas.* Ediciones Universidad de Salamanca

Corbí, M. (2020). *Proyectos colectivos para sociedades dinámicas.* Herder.

Crawford, D. (2025). *Surveillance pricing: How your data determines what you pay.*

Crawford, K. (2021). *Atlas of AI*. Yale University Press.

De Saussure, F. (1959). *Course in general linguistics*. Columbia University Press.

Dewey, J. (1927). *The public and its problems*. H. Holt and Company.

Eubanks, V. (2019). *Automating inequality : how high-tech tools profile, police, and punish the poor.* Picador.

European Parliament, & European Council. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024. *Official Journal of the European Union,* 1689(3), 1–144. http://data.europa.eu/eli/reg/2024/1689/oj

Gillespie, T. (2018). *Custodians of the internet*. Yale University Press.

Golumbia, D. (2022, December). ChatGPT Should Not Exist. *Medium*.

Good, I. J. (1966). *Speculations Concerning the First Ultraintelligent Machine*. May, 31–88.

Haeck, P. (2025, June). EU's waffle on artificial intelligence law creates huge headache. *Politico.*

Hammond, G., & Acton, M. (2025, September 5). AI start-up Anthropic settles landmark copyright suit for $1.5bn. *Financial Times.*

Haraway, D. (1988). Situated Knowledges: the Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), 575–599.

Harding, S. (2017). *Whose Science? Whose Knowledge?*

Hill, K. (2025). A Teen Was Suicidal. ChatGPT Was the Friend He Confided In. *New York Times.*

Horton, M. (2025). *The Illusion of Thinking : Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity.* 1–30.

Kinder, T., & Hammond, G. (2025, October 2). OpenAI overtakes SpaceX after hitting $ 500bn valuation. *Financial Times*.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. The University of Chicago Press.

Latour, B. (2005). *Reassembling the social : an introduction to actor-network-theory* (Repr.). Oxford Univ. Press.

Law, J. (2017). STS as Method. In U. Felt (Ed.), *The handbook of science and technology studies.*

Lohr, S. (2025). Your A.I. Radiologist Will Not Be With You Soon. *New York Times.*

Magid, Y. (2025). *AppleStorm: Unmasking the Privacy Risks of Apple Intelligence.* Lumia.

Marcus, G. (2024). *Taming Silicon Valley.* The MIT Press.

McCabe, D. (2025). Regulators Are Digging Into A.I. Chatbots and Child Safety. *New York Times.*

McQuillan, D. (2022). Resisting AI. In *Resisting AI*. Bristol University Press.

Metz, C., Kang, C., Fenkel, S., Thompson, S., & Grant, N. (2024). How Tech Giants Cut Corners to Harvest Data for A.I. *New York Times.*

Montgomery, B. (2024). Mother says AI chatbot led her son to kill himself in lawsuit against its maker | Artificial intelligence (AI). *The Guardian*, 11–13.

Mühlhoff, R. (2025). *The Ethics of AI. Power, Critique , Responsibility.* Bristol University Press.

Munn, L. (2022). The uselessness of AI ethics. *AI and Ethics.*

Niederhoffer, K., Kellerman, G. R., Lee, A., Liebscher, A., Rapuano, K., & Hancock, J. T. (2025). AI-Generated " Workslop " Is Destroying Productivity. *Harvard Business Review.*

O'Neil, C. (2017). *Weapons of math destruction : how big data increases inequality and threatens democracy*. Penguin Books.

Olson, P. (2024). *Supremacy*. St. Martin's Press.

Phan, T., Goldenfein, J., Mann, M., & Kuch, D. (2022). Economies of Virtue: The Circulation of 'Ethics' in Big Tech. *Science as Culture*, 31(1), 121–135.

Shapin, S. (1984). Pump and circumstance: Robert Boyle's literary technology. *Social Studies of Scienc*e, 14, 481–520.

Singer, P. (1993). *Practical Ethics.* Cambridge University Press.

Torres, P. (2021). The Dangerous Ideas of "Longtermism" and "Existential Risk." *Current Affairs*, 1–22.

Wylie, C. (2019). *Mindf*ck : Cambridge Analytica and the plot to break America.* Random House.

Wynn-Williams, S. (2025). *Careless people*. Flatiron Books.

Zahn, M. (2023). *Elon Musk launches his own AI company to compete with ChatGPT.* https://abcnews.go.com/Business/elon-musk-launches-ai-company-compete-chatgpt/story?id=101210078