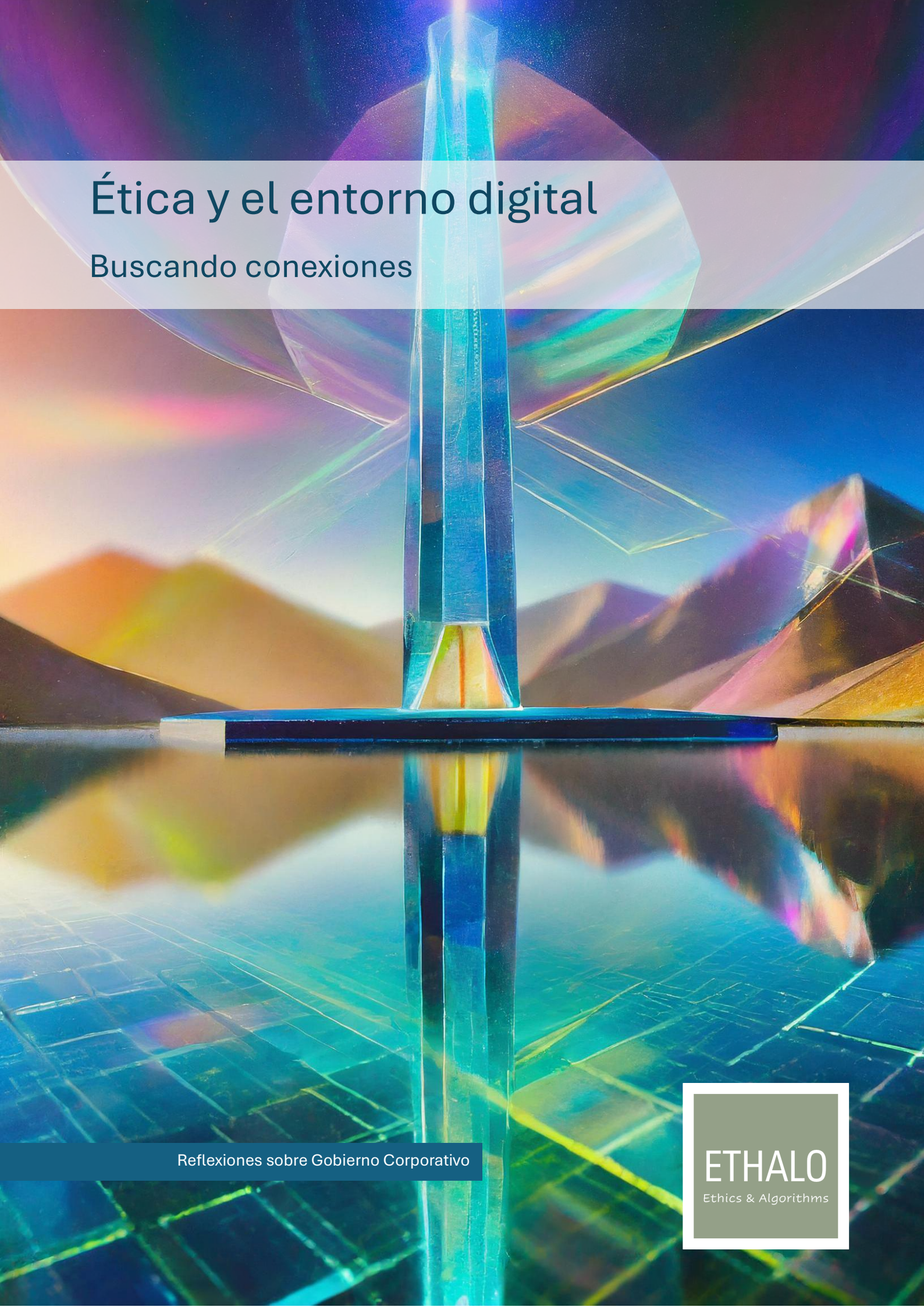




Ética y el entorno digital

Buscando conexiones

Reflexiones sobre Gobierno Corporativo

ETHALO
Ethics & Algorithms

Imágenes realizadas por el autor con IA (Firefly)

© Ethalo Advisory, SL
Avenida Isabel de Farnesio (1512), local J
28660 Boadilla del Monte
www.ethalo.es

Serie Reflexiones sobre Gobierno Corporativo
Primera edición: diciembre 2024

Se permite su distribución sin finalidades comerciales, bajo la condición de que el material se ofrezca de manera gratuita y se cite debidamente al autor como la fuente original. La distribución de este trabajo debe hacerse sin modificaciones y en su forma completa, salvo autorización expresa. Se anima a los lectores a compartir este documento, siempre que se respeten estas condiciones.

Se permite el uso de fragmentos o citas de este libro en cualquier medio, siempre y cuando se reconozca debidamente la autoría, indicando el nombre del autor y la fuente original del texto. Este permiso está limitado a usos que no tengan finalidad comercial y que busquen enriquecer otros trabajos a través de la referencia adecuada.

Este documento está protegido bajo las leyes de derechos de autor. Cualquier uso adicional de este trabajo debe ser autorizado por escrito por el autor.

Cita del presente documento:

Patxi Barceló. Ética y el entono digital. Buscando conexiones. Proyecto Ethalo (Ethics and algorithms) – Serie Reflexiones sobre gobierno corporativo. Diciembre 2024 (1ª ed).

Ética y el entorno digital

Buscando conexiones



Índice de contenidos

Introducción	5
¿Qué sabemos de la Inteligencia Artificial?.....	5
¿Qué queremos decir cuando hablamos de ética?	8
Posicionamientos éticos y su impacto práctico	12
La ética de la vulnerabilidad	13
Ética de los cuidados	15
La ética colectiva.....	18
Ética del empoderamiento	20
Ética de la responsabilidad	23
Ética de la gobernanza	25
Nuevo materialismo.....	27
¿Se puede introducir la ética en los sistemas de IA?	29
A modo de conclusiones	30

Introducción

Este cuaderno es un documento *in itinere*. Pretende generar reflexión y debate, manteniéndose abierto a mejoras futuras con los comentarios e ideas que vayan aportando todas las personas interesadas.

El objetivo es aportar valor a las personas que deben participar en reflexiones estratégicas en las organizaciones para que puedan reforzar su labor en el ámbito de la implantación de modelos, algoritmos y sistemas de inteligencia artificial, considerando especialmente sus aspectos reputacionales y de experiencia de cliente.

En este cuaderno se aborda la inteligencia artificial desde la perspectiva de su impacto en los procesos reales y las implicaciones que tienen para la buena gobernanza de las organizaciones. Usamos la palabra ética en el sentido originario, de las costumbres y hábitos que rigen la conducta de las personas o las organizaciones, generando un impacto en la experiencia de todos los afectados (clientes, empleados, proveedores, etc.). Por tanto, el foco está en el impacto y no en la valoración de la conducta en sí misma. Esta distinción, como veremos, es importante para facilitar llegar a conclusiones prácticas.

¿Qué sabemos de la inteligencia artificial?

Los coches de conducción totalmente autónoma ya son una realidad que, poco a poco, se hará habitual. En San Francisco y Phoenix están integrados en el día a día como vehículos de transporte de viajeros sin conductor a través de la empresa Waymo¹.

Esta empresa indica que tiene la misión de ser el conductor más confiable del mundo, haciendo que sea más seguro, más accesible y sostenible desplazarse, sin necesidad de que nadie esté en el asiento del conductor².

La pregunta inicial, sobre nuestra comprensión de la Inteligencia Artificial, no se refiere a las capacidades técnicas del vehículo autónomo, sino sobre nuestro conocimiento de las reglas de juego que utiliza: ¿Estaríamos dispuestos a coger un coche autónomo que sabemos que está programado para que, en caso de conflicto, priorice la vida de cualquier tercero externo frente a la vida de los propios ocupantes del vehículo?

También podemos hacernos la pregunta contraria ¿Y nos parecería un buen argumento de venta, y subiríamos al coche con mayor tranquilidad, si sabemos que dicho coche autónomo, por el contrario, está programado para proteger la vida de sus ocupantes en cualquier circunstancia, aunque ello pueda implicar la pérdida de otras vidas, sin importar cuantas ni en qué circunstancias?

Seguramente estemos tentados de pensar que la virtud está en un término medio y que depende de las circunstancias de cada caso. Esta es la cuestión ¿sabemos, cuando

¹ Empresa propiedad de Google y que da continuidad empresarial al proyecto de vehículo autónomo que Google lleva años desarrollando. En agosto de 2024 Waymo ya está operativo en San Francisco y Phoenix y pendiente de iniciar las operaciones en breve en Los Ángeles y Austin.

² En version original: We're on a mission to be the world's most trusted driver. Making it safer, more accessible, and more sustainable to get around — without the need for anyone in the driver's seat

estamos usando estas máquinas, realmente cuáles son las reglas de juego que se van a aplicar?

Cuando somos conscientes de que un amigo ha bebido, le invitamos a que no conduzca y vuelva por otros medios a su casa, incluso ofreciéndonos a acercarle nosotros. Somos conscientes del mayor riesgo y así actuamos. Cuando confiamos en la máquina ¿Sabemos si el software del coche autónomo ha sido debidamente actualizado? ¿sabemos que historial de fallos ha tenido el vehículo, al igual que normalmente sabemos la fiabilidad de nuestro amigo cuando nos ofrece llevarnos?

Veamos otra situación muy diferente. Los robots de asistencia quirúrgica son también un componente cada vez más habitual en los hospitales. En ocasiones, analizan las imágenes obtenidas del paciente, tanto con carácter previo a la operación como *in situ* para recomendar el mejor lugar para realizar una incisión, por ejemplo, para un injerto de ligamento en la rodilla.

La realidad, como veremos con los coches autónomos y debatiremos más adelante, es que estos sistemas apoyados por Inteligencia Artificial tienen un índice de éxito superior a la media de las operaciones tradicionales, por lo que, de forma general, logran una mejora en los resultados finales para la mayoría de los pacientes.

Ahora bien, analicemos esta situación desde una perspectiva diferente. Nos tienen que operar de ligamentos y nuestro doctor, con más de 30 años de experiencia en este tipo de operaciones, nos indica que va a usar un asistente quirúrgico para la operación, de forma que utilizará el sistema para determinar la mejor forma de proceder y el usará sus manos expertas para ejecutar la operación y para poder actuar ante cualquier imprevisto que requiera conocimientos adicionales.

Sin embargo, nos surge en esos momentos una duda y le preguntamos abiertamente. Sabemos que su experiencia es muy dilatada y tiene muy buena fama en estas operaciones, por lo que nos gustaría saber qué hará si no le convence lo que le sugiere la máquina. Es posible con que nos encontremos con una respuesta como está: “la máquina acierta en un 99,3% de los casos, por lo que su capacidad es muy superior a la mía, no tengo motivos para desconfiar de ella”.

Valoremos lo que subyace a esta afirmación. Aunque nuestro médico tenga la certeza de que la máquina no está haciendo la recomendación correcta, se encuentra ante una doble alternativa totalmente asimétrica. Si hace caso a la máquina, pero esta se ha equivocado, simplemente ha seguido el protocolo y la máquina ha cometido uno de sus posibles errores (muy bajos, pero reales). Asunto terminado. La alternativa es poco alentadora. Aunque esté seguro de que la máquina se equivoca, no tiene la certeza absoluta. Si no hace caso a la máquina y la operación sale mal, todas las miradas se dirigirán hacia él y la pregunta será evidente ¿por qué no hizo caso al sistema, que no se equivoca casi nunca? Es decir, la pregunta que subyace es ¿Qué incentivo tiene el doctor para llevar la contraria a la máquina, por muy seguro que esté del error?

Esta misma pregunta nos la podemos hacer para interrogarnos sobre los incentivos que puede tener un analista de crédito para autorizar una operación que el sistema considera

de riesgo elevado³ o un juez para dejar en libertad provisional a una persona cuando existe un sistema que está asignando a dicho individuo un riesgo de reincidencia alto⁴.

Analicemos una tercera situación real. Usted tiene un seguro de hogar, con la misma compañía hace ya varios años. La verdad es que no tiene especial queja, sólo tuvo un incidente, que se resolvió de forma eficiente, y el precio del seguro no ha subido significativamente a lo largo de los años (aunque nunca ha bajado, a pesar de no tener accidentes graves).

El tema de los seguros sale en una conversación y, tras darle varias vueltas, llega a la conclusión de que un compañero suyo tiene una vivienda muy parecida y sin embargo está pagando una cantidad muy inferior por el seguro. Incluso su interlocutor ha tenido más incidentes y está en la misma compañía.

Es entonces cuando descubre cómo funciona el algoritmo de su compañía de seguros para fijar los precios de las pólizas. Usa lo que en el argot se conoce como *price walking*. Este modelo trata de identificar la propensión que tiene el cliente a comparar precios de diversas aseguradoras e incluso cambiarse de seguro. De esta forma, a los clientes más fieles y que no tienen propensión al cambio, les mantiene los precios invariados o los sube conforme a parámetros habituales como el índice de precios al consumo. Sin embargo, en aquellos clientes en los que identifica una propensión al cambio de compañía, les ajusta los precios a la baja para evitar que acaben migrando a otra entidad.

Esta práctica ha sido habitual en el mercado⁵ y ha llevado a su prohibición expresa en el sector en varios países. Por desgracia, no es una práctica exclusiva de las compañías de seguros, sino que está extendida en otros proveedores de servicios (telefonía, televisión de pago, remuneración de cuentas de ahorro, etc.). De hecho, un análisis detallado de esta práctica permite concluir que esta perjudica en mayor medida a las personas más

³ La mayor parte de las entidades financieras españolas ya tienen sistemas de decisión automatizados que resuelven la mayor parte de las operaciones de petición de préstamos personales o tarjetas de personas físicas. Estos sistemas tienen una buena capacidad predictiva media, incluso con información limitada a unas pocas variables. Muchas entidades tienen totalmente automatizado el proceso, de forma que si el sistema arroja una puntuación baja (alto riesgo), no existe una revisión manual de la operación y se deniega automáticamente (aunque pudieran existir otras variables que permitirían reevaluar el riesgo de la operación).

⁴ Este sistema existe en Estados Unidos (COMPASS) y ha sido muy criticado por el sesgo implícito que asigna a las personas de color. En España existe un algoritmo semejante que se aplica en el modelo penitenciario de Cataluña, que tiene las competencias transferidas en la materia. El nombre del sistema es RisCanvi. Para un detalle divulgativo sobre su funcionamiento véase https://www.elconfidencial.com/tecnologia/2024-04-24/riscanvi-algoritmo-cataluna-prisiones-presos-inteligencia-artificial_3871170/. Adicionalmente, existe un informe de auditoría técnica sobre el sistema realizado por Claudia Herron, Xavier Clotet, Pol Colomer (2024). Informe Tiresias. Auditoria de l'algorisme RisCanvi. Dribia Data Research. (#788). Disponible en https://justicia.gencat.cat/ca/ambits/reinsercio_i_serveis_penitenciaris/serveis_penitenciaris/RisCanvi/

⁵ Véase el informe de la Financial Conduct Authority (FCA) de UK. General insurance pricing practices (Final Report, septiembre 2020).

desfavorecidas (mayores, población con bajo nivel de educación o con ingresos más ajustados)⁶.

Por eso, cuando preguntamos sobre el conocimiento de la Inteligencia Artificial realmente estamos preguntando, más allá del conocimiento técnico ¿sabemos las reglas del juego? Es aquí donde empiezan a aparecer conceptos como los valores y la ética.

No podemos negarlo, hablar de la ética de la Inteligencia Artificial está de moda. Sin embargo, tenemos que anticipar que algunas cuestiones nos van a obligar a salir de la caja y revisar algunos paradigmas. Al mismo tiempo, este paso nos abrirá puertas a nuevas formas de trabajar y oportunidades de negocio que pueden servir de catalizador en cualquier organización. En todo caso, iniciemos el camino...

¿Qué queremos decir cuando hablamos de ética?

Aunque la ética este ahora en boga, necesitamos comenzar clarificando que queremos decir cuando hablamos de este concepto. Veremos la utilidad de utilizar este término en su significado original, referido a los hábitos o costumbres. Esta forma de entender la ética nos ayudará a comprender que tenemos no solo la capacidad, sino también la responsabilidad de definir nuestro espacio ético específico (formas de actuar, hábitos) que queremos utilizar y hacer una reflexión explícita de esta elección y de sus consecuencias.

La ética habla de acciones y cuando hablamos de estas desde una perspectiva ética, tendemos a vincular de forma inmediata una valoración, positiva o negativa, conforme o no conforme a un ideal. Sin embargo, una revisión etimológica de la palabra ética nos debe llevar a reflexionar sobre este aspecto. La ética se refiere a los hábitos o costumbres, sin mayores consideraciones de partida, sin un a priori que pueda nublar nuestro entendimiento.

Este aspecto se hace todavía más confuso cuando usamos la palabra moral, asociada en mayor medida y de forma habitual a determinados valores, aunque en origen no existe esta conexión. Por ello es relevante empezar por el análisis de estos conceptos.

Seguimos para ello a Ortiz Millán (2016)⁷ que nos ilustra con claridad al respecto:

La palabra *ética* proviene del griego, y puede tener dos distintas etimologías, que son complementarias.

Una primera etimología nos dice que proviene de *hábito*, *costumbre*, *estar acostumbrado*, como cuando Aristóteles afirma en la *Ética* a Nicómaco: “Algunos creen que los hombres

⁶ Véase Citizens Advice (2018). Excessive prices for disengaged consumers. A super-complaint to the Competition and Markets Authority disponible en

<https://www.citizensadvice.org.uk/cymraeg/amdanom-ni/our-work/policy/policy-research-topics/consumer-policy-research/consumer-policy-research/excessive-prices-for-disengaged-consumers-a-super-complaint-to-the-competition-and-markets-authority/>

⁷ Gustavo Ortiz Millán (2016). Sobre la distinción entre ética y moral. Universidad Nacional Autónoma de México. (#E004). Disponible en https://www.scielo.org.mx/scielo.php?pid=S1405-02182016000200113&script=sci_abstract

llegan a ser buenos por naturaleza, otros por el hábito”⁸, en donde *hábito* se opone a *naturaleza*.

Según esta etimología, el término *ética* compartiría sus raíces con el término *etología*, costumbre, y razonamiento, estudio o ciencia. La etología estudia el comportamiento de los animales, sus hábitos y sus costumbres, en el medio ambiente o en el lugar que se encuentren.

La segunda posible etimología significa *lugar acostumbrado, morada, refugio, guarida*, como, por ejemplo, en Homero, en que se refiere al hábitat de los caballos, o en Heródoto, quien se refiere a la morada de los leones. Éste es el significado arcaico del término que luego también se usaría para referirse a la morada o a la habitación de los seres humanos, como en Heródoto, que lo usa como *lejos de su lugar acostumbrado*.

Muy probablemente la palabra “ética” hubiera pasado con similares significados al latín de no ser por Cicerón. Tal vez el problema de distinguir entre estos dos términos se lo debamos a él, porque se propuso “enriquecer” el idioma latino añadiendo la palabra “moral”. Cicerón comienza su tratado *Del destino* afirmando “ya que atañe a las costumbres, lo que los griegos llaman ἠθικός, mientras que nosotros solemos llamar a esa parte de la filosofía ‘el estudio de las costumbres’, pero conviene llamarla ‘moral’ para que se enriquezca la lengua latina”⁹. Cicerón, entonces, propone un neologismo, el término “moral”, para llamar así a la disciplina filosófica que estudia las costumbres y que los griegos llamaban “ética”.

Por tanto, etimológicamente hablando hay poca diferencia entre ética y moral; una y otra hacen referencia a una misma realidad o, al menos, una realidad próxima y con fronteras permeables.

Para concluir que, el uso de estos conceptos en el entorno actual no tiene una aceptación universal y, por tanto, depende del contexto y la distinción que se quiera estipular en cada ocasión.

Suele haber cierta controversia acerca de las definiciones y la distinción de los términos “ética” y “moral”. No hay nada en la etimología de las palabras “ética” y “moral”, ni en el empleo que diversos filósofos han hecho de estos términos a lo largo de la historia, que nos imponga un determinado significado para el uso de cada una de ellas. Se trata de una **distinción estipulativa**, es decir, depende de cómo se estipule que se van a usar los términos; por ello, no puede haber un único significado válido universalmente.

Como ejemplo, entre otros modos de estipular el significado de los términos, con frecuencia se afirma que el comportamiento ético, concebido como la afirmación de la conciencia autónoma y la autenticidad individual, debe tener un mayor peso que el ámbito de lo moral, entendido como la esfera de la observancia de reglas que nos son impuestas por la sociedad.

Esta introducción debe hacernos conscientes de que cuando hablamos de ética en el contexto de la empresa (y en general en cualquier otro contexto), habitualmente damos por supuesto una estructura de valores y una filosofía concreta que da soporte a un

⁸ ‘γίνεσθαι δ’ ἀγαθὸν ὡς οἶονται οἱ μὲν φύσει οἱ δ’ ἔθει οἱ δὲ διδασχῇ’ (1179b21),

⁹ ‘quia pertinet ad mores, quos ἠθῆ Graeci vocant, nos eam partem philosophiae de moribus appellare solemus. Sed decet augentem linguam Latinam nominare moralem’ (Cic. Fat. 1, 1).

determinado posicionamiento. Sin embargo, aunque esto pudiera ser cierto, rara vez se explicita. Por tanto, **se estaría asumiendo** que existen unos valores compartidos y claros para todas las partes interesadas, es decir, **que la ética es algo que está ahí y lo único que hacemos es aplicarlo, cuando más bien la ética es la descripción de los comportamientos y las razones de estos.**

Incluso algunas declaraciones vagas y poco específicas pueden llevar a sospechar de la utilización de la ética como un lavado de cara o una forma de eludir la regulación en lugar de un examen de prácticas éticas¹⁰.

Esta situación requiere preguntarse si realmente existe ese conjunto común e indubitado de valores. La realidad nos demuestra que esto no suele ser así. Precisamente, en los momentos de conflicto y tensión es cuando se evidencia que el significado de estos no es igual para todos y que, adicionalmente, el orden de prioridad entre unos valores y otros puede ser muy diferente entre las partes.

Además, la prioridad de los valores está íntimamente vinculada también a una polarización entre la preminencia de lo individual y lo colectivo. Los propios términos de ética y moral, como hemos señalado más arriba, se asocian en no pocas ocasiones a la visión individual frente a las normas generales impuestas. Lo que nos interesa a estos efectos no es tanto esta distinción, sino señalar cómo uno de los ejes esenciales en la valoración de las decisiones está en el equilibrio entre el interés individual y el colectivo, aspecto que afecta de forma sustancial al entendimiento de los valores en cualquier contexto¹¹.

Uno de los riesgos de no explicitar con claridad los valores y prioridades es el traslado de la cuestión a cada persona en cada acción concreta. No existe una referencia, un espejo en el que mirarse, sino más bien buscar la aprobación del entorno, que va cambiando de forma paulatina y silenciosa según los intereses del momento. A posteriori, la institución valorará el resultado obtenido según las circunstancias, de forma que se refuerzan los incentivos (a través de los objetivos) al mismo tiempo que se exacerban las culpas.

Lo expresa perfectamente Víctor Lapuente, en otro contexto¹², cuando nos recuerda que está “modernidad” nos saca de la cueva y arrincona la culpa judeocristiana, haciéndonos los únicos responsables de nuestro destino:

La responsabilidad individual se ha alargado como una sombra siniestra sobre toda nuestra trayectoria vital. Perdemos el trabajo porque hemos tecleado una palabra de más en un tuit (y hemos ofendido a alguien) o un número de menos en la hoja de Excel (y nos ha superado

¹⁰ VVAA (2021). Inteligencia artificial, ética y sociedad. Una mirada y discusión a través de la literatura especializada y de opiniones expertas. Observatori d'Ètica en Intel·ligència Artificial de Catalunya (OEIAC). (#792). Disponible en https://vvvtest.udg.edu/ca/Portals/57/OContent_Docs/Informe_OEIAC_2021_cast-4.pdf (pag 27)

¹¹ Más adelante se tendrá ocasión de profundizar en este aspecto, pero ya podemos vislumbrar que la defensa de la privacidad es una prioridad en la mentalidad europea y así se refleja en las leyes sobre IA, en tanto que en la legislación China sobre IA existe una predominancia del bien común sobre los derechos individuales. Esta diferencia lleva a un tratamiento totalmente asimétrico en ciertos desarrollos de la IA.

¹² Víctor Lapuente. Por qué vivimos enfadados. (El País de 13 de agosto de 2024). En <https://elpais.com/opinion/2024-08-13/por-que-vivimos-enfadados.html>

en rendimiento el compañero de al lado). Hoy nos podemos ver en el banquillo de un juicio penal por fallos que antaño eran meras chapuzas, como empalmar mal unos cables de alta tensión. Que tengas que pagar una indemnización millonaria por servir café demasiado caliente ya no es una anécdota propia de los yanquis. Se ha judicializado la cotidianidad.

Alguien podría preguntarse incluso si, en este caso, estaríamos hablando de una ética de los resultados. La valoración depende del resultado final (otra forma del fin que justifica los medios).

Un individuo es autónomo (en el plano social) en la medida en que, “enfrentado a determinadas presiones y normas, las someta a una evaluación consciente y crítica, formándose intenciones y alcanzando decisiones prácticas, como resultado de su reflexión independiente y racional” (Lukes, 1975, p. 69¹³). La autonomía individual así entendida comprende la autonomía ética, es decir, la capacidad que tenemos de darnos a nosotros mismos los valores y principios con los que decidimos guiar nuestras vidas.

Lo mismo corresponde a cualquier institución, que define intenciones y reglas prácticas que generan lo que llamamos su forma de ser. No nos corresponde valorar su bondad, sino su éxito para lograr adhesión y confianza, dos características cuyo equilibrio parece esencial en el largo plazo.

A partir de aquí, podemos tratar de esbozar unas **primeras aproximaciones prácticas para la gobernanza**:

- Existe una tendencia en las organizaciones a expresar su misión, visión y valores. Estos enunciados suelen ser una enumeración de palabras evocativas de alto nivel, que rara vez se aterrizan en ejemplos concretos.
- Si queremos que la empresa tenga pautas de acción claras, es necesario definir esos comportamientos y prioridades a nivel operativo. Es decir, hablar de la ética como del comportamiento habitual de la empresa y del comportamiento que se esperan en situaciones concretas.
- Por tanto, la ética empresarial debe ser una manifestación explícita de comportamientos y prioridades, pública y que sirve de guía efectiva ante conflictos.

Podemos intentar mantener nuestra organización en la vaguedad, pero si algo bueno nos puede traer la inteligencia artificial es precisamente fomentar esa mayor transparencia y concreción¹⁴, no tanto de los entresijos del algoritmo (útiles para una auditoría de verificación), sino más bien de las decisiones clave que se han tomado en su construcción e implementación y que tienen impacto real en la salida del sistema y, por tanto, en el usuario y el entorno. Como decíamos antes, explicitar las reglas de juego.

¹³ Steven Lukes. El individualismo. Ediciones Península (1975)

¹⁴ Como ya señalamos en un artículo anterior, no es que el análisis ético sea necesario para la IA, sino para cualquier actividad. La IA nos da la oportunidad de poner este debate encima de la mesa, al representar un agente externo aparentemente impersonal. Esto permite reducir la sensibilidad de la discusión, al alejarla del comportamiento humano. Ver Patxi Barceló (2023). El gobierno de la innovación y el paradigma digital (Cómo innovar desde el consejo - capítulo 7). ISBN 978-84-09551-69-9

El entorno digital hace más necesario que nunca esa reflexión explícita, donde es conveniente detenerse y tomar decisiones claras en aspectos que, hasta ahora, no eran objeto de especial atención. La máquina representa, cada vez más, a la propia organización. En la medida en que hemos decidido delegarle un papel central en determinados procesos, sus acciones serán atribuidas, sin duda, a la voluntad de su dueño. ¿Alguien duda de que el algoritmo de *Google* proporciona sus resultados según las intenciones de la organización que lo gestiona? ¿Alguien puede asumir que los videos que muestra *Tik-tok* a nuestros adolescentes son un resultado aleatorio sin más, sin intencionalidad ninguna? Podemos traer de nuevo a colación los diversos ejemplos que pusimos al principio de este texto. Hagámoslo en genérico ¿algún sistema se implanta sin una intencionalidad?

La cuestión esencial es asegurar que la organización es consciente de esa intencionalidad. No siempre es evidente e, incluso en ocasiones, puede ser diferente a la deseada, por motivos muy diversos. La responsabilidad de cualquier consejo es asegurar que existe el debate y la indagación suficiente para que estas cuestiones tengan la importancia debida en los procesos estratégicos. Eso nos lleva a la siguiente pregunta.

Posicionamientos éticos y su impacto práctico

Como anticipa el Observatorio de Ética en Inteligencia Artificial de Cataluña (2021)¹⁵:

Este campo en la actualidad está dominado por la proliferación de “códigos de ética” de la IA que buscan guiar el diseño y despliegue de sistemas de IA, si bien éstos son limitados en aspectos clave ya que carecen de un marco universalmente acordado y no son vinculantes.

(...) Aunque existe un consenso cada vez mayor sobre cuestiones fundamentales como la responsabilidad, la justicia social, la privacidad y el consentimiento, sabemos que éstos pueden estar sujetos a diferentes significados en diferentes contextos.

Nuestro punto de partida es que **el objetivo de este trabajo no es determinar un posicionamiento ético dominante, sino que cada una de estas cosmovisiones tiene un valor intrínseco que debemos conocer y aprovechar para construir el de nuestra organización**. Vamos, para ello, a recorrer diferentes enfoques específicos y los aprendizajes que pueden surgir de cada uno de ellos.

En términos prácticos, buscamos apoyos con los que responder preguntas esenciales: ¿Cómo es la experiencia del cliente al usar un sistema de IA? ¿Cuál es la opinión pública que se genera ante determinados eventos provocados por la IA? Más allá de los aspectos legales, e incluso morales, de ciertos eventos ¿Cómo afectan a las organizaciones cuándo salen a la luz determinados resultados provocados por sistemas automatizados? ¿Tenemos evaluado el riesgo reputacional asociado?

¹⁵ VVAA (2021). Inteligencia artificial, ética y sociedad. Una mirada y discusión a través de la literatura especializada y de opiniones expertas. Observatori d'Ètica en Intel·ligència Artificial de Catalunya (OEIAC). (#792). Disponible en https://vvvtest.udg.edu/ca/Portals/57/OContent_Docs/Informe_OEIAC_2021_cast-4.pdf

La ética de la vulnerabilidad

Planteamiento formal

Aprovecharemos el trabajo de Antonio Luis Terrones (2023)¹⁶ para desgranar este posicionamiento, que parte del concepto de vulnerabilidad como un elemento esencial de la humanidad.

Se entiende por vulnerabilidad no como algo consustancial a la capacidad humana de sufrir (enfoque ontológico), sino desde una perspectiva relacional, donde se enfatiza la dimensión social y contextual de la vulnerabilidad.

La perspectiva relacional señala la existencia de diversos grados y circunstancias que condicionan la experiencia de la vulnerabilidad. Ser vulnerable significa ser susceptible de experimentar daños en los propios intereses cuando otros actores o actividades representan una amenaza. Si bien todo ser humano es vulnerable, el grado de dicha situación depende de múltiples circunstancias, en gran medida externas. Esta visión lleva a la necesidad del reconocimiento y cuidado de la vulnerabilidad de determinados grupos humanos.

La IA es una fuente potencial de generación de vulnerabilidad significativa, de forma que los algoritmos se convierten en una parte más del engranaje institucional y de las estructuras donde se asientan los sesgos.

El propósito de los algoritmos consiste en asumir el trabajo que desempeñan los humanos en la toma de decisiones para aliviar su esfuerzo, agilizar los procesos y abaratar los costes. Una intención que responde a una percepción errónea de las decisiones algorítmicas sobre la objetividad. La apelación a los algoritmos tiene su origen en un pensamiento históricamente cuestionado que establece un vínculo entre la objetividad y la imparcialidad, y sitúa a los humanos en desventaja respecto a las máquinas

Lorraine Daston y Peter Galison (1992) llamaron a este pensamiento “objetividad mecánica”, fundamentado en la creencia de que la tecnología es superior a la subjetividad humana porque elementos como los valores, los intereses o los sesgos no intervienen de forma decisiva. En términos de Evgeny Morozov, podemos hablar de un solucionismo falaz.

Cuando la sabiduría práctica, traducida en deliberación, buena elección y juicio sosegado, es desplazada por la quimérica objetividad y falsa neutralidad promulgada por el solucionismo tecnológico para defender la expansión de las decisiones algorítmicas, el ser humano está olvidando un aspecto esencial de su vida, la vulnerabilidad. Es preciso proponer estrategias humanas enraizadas en la sabiduría práctica que pongan en valor las narrativas particulares de los contextos.

Aristóteles (2005) empleó el término *phronēsis* para significar la excelencia de la inteligencia práctica. Esta virtud dianoética sobre la que el estagirita reflexiona invita a pensar la praxis de manera adaptativa a una realidad que es contingente. (...) Implica cultivar una mirada sensible a las narrativas particulares.

¹⁶ Terrones Rodríguez, A. L. (2023). Inteligencia artificial fiable y vulnerabilidad: una mirada ética sobre los sesgos algorítmicos. En *Vulnerabilidad digital: Desafíos y amenazas de la sociedad hiperconectada*. Dykinson. (#790).

Como remarca Terrones en sus conclusiones, es imprescindible conceder mayor importancia al análisis narrativo de los contextos, con la finalidad de desocultar aquellos aspectos que merecen especial atención para la sensibilización.

Aplicación en casos concretos

Los sistemas basados en inteligencia artificial cada vez integran un mayor volumen de datos y son capaces de particularizar cada vez más las soluciones de cada caso, alcanzando índices de “acierto” realmente elevados. Sin embargo, este proceso es fundamentalmente estadístico, la correlación se confunde en ocasiones con la causalidad o simplemente se ignora y el grado de acierto objetivo tiende a dejar de lado los errores por ser poco significativos en número.

Desde una perspectiva estadística, podemos visualizar los errores, en uno u otro sentido, como una campana de Gauss, donde los casos extremos, las colas, son pequeñas. Ese es el objetivo de una buena labor estadística, lograr que la tasa de aciertos sea del 99,9...%

Sin embargo, veremos que estos errores, dependiendo del contexto y el tipo de error, pueden requerir una reflexión profunda sobre su impacto real.

Analicemos primero el caso de un algoritmo de un buscador de contenidos en internet. Que logre que el 95% de los usuarios localicen lo que les interesa es un buen resultado. Los usuarios, si en el primer intento no logran encontrar lo que buscaban, están dispuestos a volver a intentarlo (el coste es bajo). Es más, eso no quiere decir que el usuario haya encontrado la mejor respuesta, sino que ha encontrado una respuesta que considera válida o suficiente. Es uno de esos casos donde el buscador puede decir, no me importa la causalidad (encontré lo que el necesitaba) sino la simple correlación (se fue satisfecho con lo que le ofrecí, aunque no sea el óptimo).

Pasemos ahora a un algoritmo de decisión de crédito bancario. Aunque los algoritmos son cada vez más sofisticados, para tener validez estadística tienen que manejar grupos suficientemente homogéneos de casos para lograr esa validez. Imaginemos que logramos un algoritmo que es capaz de evaluar de 1 a 100 la probabilidad de impago de mis clientes y, como banco, dado que normalmente cobro un interés por los préstamos al consumo del 7%, decido legítimamente no aceptar clientes con un riesgo de impago superior al 8%.

Leamos este resultado de forma práctica. Esto quiere decir que rechazaré a todos los clientes que tengan un scoring, por ejemplo, de impago del 10%. En ese grupo es posible que haya 20.000 clientes, de los que estoy seguro de que un 10% van a fallar, pero realmente no sé cuáles van a impagar. Por tanto, desde el punto de vista práctico, significa que rechazaré a los 20.000 clientes, de los cuales 2.000 hubieran impagado y 18.000 hubieran pagado correctamente su deuda. Aquí el impacto es muy diferente y tendré que evaluar si me compensa dejar que el algoritmo decida o añadir otros mecanismos manuales que permitan una mejor discriminación ¿realmente podrán unos humanos lograr mejor resultado? ¿cuál es el coste de esos recursos adicionales?

Pasemos a un último ejemplo, como puede ser el uso de cámaras de reconocimiento facial para identificar delincuentes que han cometido un delito (no hablamos de predicción de un delito, sino identificación de personas que efectivamente han cometido el mismo).

En este caso, ¿cuál es el impacto de un error de identificación? Ese error, como en los casos anteriores, puede ser en dos sentidos: o bien no identificar a un delincuente y dejarlo libre o bien identificar a alguien como delincuente a pesar de que sea inocente. ¿Son simétricos ambos errores? Evidentemente que no.

La identificación, detención y procesamiento de personas erróneamente por estos medios ya se ha producido en diversas ocasiones, con graves consecuencias para los afectados. Ousmane Bah fue arrestado por ser identificado como el autor de varios robos en tiendas de Apple; tuvo que demostrar que el sistema se equivocaba, hasta el punto de que uno de los robos tuvo lugar en Boston el mismo día en que asistía a su fiesta de graduación en Manhattan. En el caso de Porcha Woodruff, también arrestada ante la identificación realizada por un modelo de reconocimiento facial, su arresto se produjo frente a su hogar cuando la señora estaba preparando a sus hijas para ir a la escuela, estaba embarazada de ocho meses y fue esposada y trasladada al centro de detenciones de la ciudad, donde permaneció detenida durante 11 horas, sufriendo contracciones y dolores agudos en la espalda¹⁷.

Conclusión base

Los sistemas basados en IA tienen un enfoque estadístico, lo que implica que incluso en los extremos de la distribución incluirá situaciones que, por construcción, no sean bien tratados. Estos errores suelen afectar, además, a situaciones más vulnerables, pues suelen ser casos que se salen de la “normalidad”. Dependiendo del impacto de la decisión, será necesario establecer protocolos estrictos de supervisión que eviten errores de alto impacto. Lo más relevante es que este impacto debe ser medido como el impacto en el sujeto, no en la organización.

Seguramente es conveniente repetir aquí la frase de Terrones (2023), ahora a la luz de estos ejemplos: es imprescindible conceder mayor importancia al análisis narrativo de los contextos, con la finalidad de desocultar aquellos aspectos que merecen especial atención para la sensibilización

Ética de los cuidados

Planteamiento formal

Este posicionamiento parte de un contexto específico, como es el uso de robots dirigidos al cuidado de personas enfermas o con discapacidad. No obstante, su valor se puede extrapolar con facilidad, por lo que incluimos aquí su desarrollo, sobre la base del texto de Nuria Vallés-Peris (2021)¹⁸.

¹⁷ Casos obtenidos de Luis Pérez Orozco (2024). Daños ocasionados por sistemas de inteligencia artificial: desarrollo vs derechos (Inteligencia artificial para un futuro sostenible: Desafíos jurídicos y éticos). Casa Leiria. (#800). Disponible en https://www.academia.edu/122738097/Da%C3%B1os_ocasionados_por_sistemas_de_inteligencia_artificial_desarrollo_vs_derechos?email_work_card=view-paper

¹⁸ Vallés-Peris, N. (2021). Repensar la robótica y la inteligencia artificial desde la ética de los cuidados. Teknokultura. Revista de Cultura Digital y Movimientos Sociales, 18(2), 137-146. <http://dx.doi.org/10.5209/TEKN.73983> (#791)

El punto de partida está en considerar que un artefacto (un sistema inteligente) no es un mero objeto con una serie de funcionalidades, sino que encierra toda una red de relaciones en la que las consideraciones técnicas, científicas, políticas, económicas o éticas están íntimamente vinculadas en un todo.

Estos artefactos construyen una agencia en función de una comprensión específica del éxito que se mueve entre dos polos: la fidelidad a la forma en que los humanos actúan, en un extremo, y lo que se considera el ideal de inteligencia y racionalidad, en el otro.

Además, las condiciones de posibilidad de utilización de estos artefactos pueden ser diferentes de las que se pretendía originalmente y de las que estaban contempladas en su diseño.

Estas tecnologías posibilitan determinadas relaciones de los humanos con el mundo que de otro modo no habrían sido posibles. Sin embargo, las tecnologías no son intermediarias neutrales, sino mediadoras activas que contribuyen a la formación de las percepciones y la interpretación que los humanos hacemos de la realidad: son moral materializada, ofrecen propuestas materiales sobre cómo actuar (Verbeek, 2008).

Esta idea supone que el debate ético no recae en grandes temas filosóficos, sino en aspectos vinculados con el día a día. De este modo, en el ámbito de la robótica el interés se fijaría en identificar las prácticas, miedos y esperanzas que configuran las relaciones entre los artefactos y las personas y en la extensa red de interconexiones existentes. Por tanto, identificar qué imaginarios y preocupaciones prevalecen en el diseño tecnológico, qué relaciones de poder materializan los artefactos y a qué mundo contribuyen.

El objeto de discusión ya no es el robot, el artefacto, sino la red de relaciones en las que ese robot participa. Esta aproximación supera la conceptualización de las relaciones de los humanos con los robots desde un modelo binario, sin que debamos obviar que muchas de esas relaciones son desiguales y que es imprescindible escuchar el impacto, especialmente en los usuarios y, entre estos, con especial atención a los grupos minoritarios o más desprotegidos.

Desde esta visión podemos significar que todas las interacciones entre los sistemas de IA y las personas conllevan necesariamente un impacto en las personas, que puede darse de formas muy diversas y, por tanto, existe una dimensión ética. Este impacto debe ser entendido en la forma en que el sistema realiza el proceso o arroja el resultado -ambas acciones, no una más importante que la otra, como veremos a continuación- y afecta a las personas, el entorno y sus interrelaciones.

Aplicación en casos concretos

Como hemos indicado, este posicionamiento nos alerta de como la interacción con sistemas digitales puede alterar el ecosistema de relaciones. Usaremos aquí casos muy sencillos pero que ayudan a evidenciar la cuestión.

Cada vez están más extendidos los sistemas de cola en los cuales se debe introducir algún tipo de dato de identificación personal para que el sistema te asigne un número en la cola, de forma que no existe una cola física y, además, los números asignados no son correlativos y suelen tener una codificación variada. Se usa en bancos, hospitales, centros de atención ciudadana, etc.

El resultado de estos sistemas no es solo que las reglas de juego no son transparentes, como ya hemos tenido ocasión de observar en otros ejemplos. La cuestión es que estos sistemas alteran el modelo de relación con un impacto significativo, y normalmente con mayor impacto en las personas o situaciones más vulnerables.

Pensemos en esos pequeños impactos. Desde el inicio, desconozco el tiempo estimado de espera, pues no visualizo la cola ni su tamaño. Tampoco puedo preguntar, pues las personas que están atendiendo tampoco disponen de esa información. Veo que los turnos que empiezan por la letra “T” van más rápido, pero desconozco si es porque realmente son casos más sencillos o porque se les está dando prioridad. Desconozco quien va delante o detrás, no le puedo pedir que me reserve el turno mientras salgo un momento para entregarle un documento a mi hija que está fuera. Tengo una urgencia y no puedo explicar a nadie el motivo de mi urgencia y si me permite adelantar el turno. Es decir, el modelo de relación con el resto de los usuarios y con el prestador de servicios se ve alterado. No entramos a valorar si el resultado global es bueno o malo, lo que nos preguntamos es si se han tenido en cuenta estas consideraciones al implantar el sistema, o se ha buscado la mera eficiencia económica.

Cuando los supermercados han implantado las cajas de auto servicio, queda la sensación de que elegir la opción de la cola atendida te sitúa en desventaja, pues el número de personas atendiendo ha bajado más que proporcionalmente al número de usuarios que demandan ese servicio y el tiempo de espera se incrementa. ¿Realmente se busca mejorar el servicio o sólo la eficiencia?

Los servicios de atención telefónica automatizada también apuntan a la eficiencia, pero la posibilidad de salir del sistema para ser atendido por un operador puede llegar a no estar disponible o la capacidad de dicho servicio atendido estar tan infradotada que supone un coste muy superior para el usuario, que incluso acaba abandonando.

Conclusión base

En general, los entornos digitales se sitúan como una barrera invisible e infranqueable entre el usuario y la organización o propietario del sistema. De esta manera se logra hacer más eficientes los procesos, tanto en tiempo como en coste. Sin embargo, estos procesos generan, no hay duda, sentimientos en los usuarios. En ocasiones estos son positivos, de eficiencia. En otras ocasiones, son negativos. Lo relevante es que ambos casos se hacen invisibles para la organización, pero tienen un claro impacto.

Si las empresas no establecen mecanismos sencillos, accesibles y atendidos por humanos para resolver estas incidencias, las percepciones negativas quedarán ocultas, pero su impacto no, dado que el usuario comentará con terceros su insatisfacción y esta se verá agravada por la imposibilidad de ser resuelta satisfactoriamente.

Pensemos que, normalmente, los procesos de queja o resolución de incidencias ya suponen de por sí una pesada carga para el usuario, pues aunque sea evidente que la empresa ha cometido un error, normalmente será necesario rellenar un formulario, acceder a determinado servicio de atención, esperar a la resolución de la incidencia para, finalmente, recibir simplemente la devolución del dinero o el cambio del servicio, sin siquiera un detalle o pequeña compensación adicional por las molestias ocasionadas.

En el ámbito digital esta circunstancia se ve especialmente agravada por la complejidad de lograr el contacto adecuado, exponer la situación y lograr su resolución, pues los sistemas rara vez contemplan una vía sencilla e inmediata para ello y el usuario se ve obligado a recopilar y documentar datos posiblemente redundantes e innecesarios.

La ética colectiva

Planteamiento formal

El colectivismo es un patrón cultural que se da sobre todo en Asia Oriental, América Latina y África (Talhelm, 2019)¹⁹. De forma simplificada, podemos afirmar que esta corriente pone el énfasis en la prioridad del grupo o colectivo frente a los derechos individuales cuando surge un conflicto de interés.

En África, la filosofía Ubuntu refleja ese espíritu. Se basa en el principio de que la identidad de una persona se forma y define a través de sus relaciones con los demás, resumido en la famosa frase "Umuntu ngumuntu ngabantu", que significa "una persona es una persona a través de otras personas". Esta idea destaca que los seres humanos solo pueden alcanzar su humanidad plena mediante la conexión con los demás y la prosperidad de una persona no tiene sentido si sus semejantes no prosperan²⁰.

Seguramente el ejemplo más relevante en términos de gobernanza política actual es el caso de China. La política China prioriza la idea del mayor bien para el mayor número de personas en lugar de un imperativo moral para proteger los derechos individuales, que predomina en la cultura occidental (Ortega, 2020)²¹.

Esta visión parece estar muy arraigada en el confucianismo. Para Confucio, la virtud individual no tiene sentido si no contribuye al bienestar común. La virtud es algo que debe manifestarse a través del comportamiento dentro del grupo, ya sea la familia, la sociedad o la nación. Esto se alinea con los principios del colectivismo, que promueven la importancia de trabajar en conjunto para lograr objetivos comunes y la estabilidad social. En el confucianismo, el estado es visto como una familia extendida, y el gobernante es el "padre" de la nación. Es una noción de estado paternalista, donde la unidad y el bienestar colectivo se priorizan sobre los intereses individuales.

¹⁹ Talhelm, T. (2019, 29 de octubre). Por qué su comprensión del colectivismo es probablemente errónea. Asociación para la Ciencia Psicológica.

<https://www.psychologicalscience.org/observer/why-your-understanding-of-collectivism-is-probably-wrong>

²⁰ Aunque la legislación en materia de IA es muy escasa todavía en África, si se observa esta característica diferencial en algunos casos. Por ejemplo, en Ruanda, la política nacional en materia de IA enfatiza la necesidad de que los avances de la IA deben ser beneficiosos y equitativos para el conjunto de la sociedad, mención que es difícil de encontrar de forma explícita en los documentos occidentales. Ver Dorcas Tsebee y Ridwan Oloyede (2024). State of AI Regulation in Africa: Trends and developments. Tech Hive Advisory Center for Law & Innovation. (#702). Disponible en <https://www.techhiveadvisory.africa/report/state-of-ai-regulation-in-africa-trends-and-developments>

²¹ Andrés Ortega Klein (2020). Geopolítica de la ética en Inteligencia Artificial. Real Instituto Elcano. (#797). Disponible en <https://media.realinstitutoelcano.org/wp-content/uploads/2021/10/dt1-2020-ortega-geopolitica-de-la-etica-en-inteligencia-artificial.pdf>

En el ámbito digital, una breve revisión de la legislación China en materia de Inteligencia Artificial²² refleja perfectamente esta filosofía y no deja indiferente, pues se puede describir bajo tres parámetros:

- Una regulación avanzada y que quiere estar en la vanguardia en la materia.
- Replica los valores también existentes en occidente y aprobados en organismos internacionales en declaraciones de amplio consenso²³.
- Con una salvaguarda omnipresente que excepciona esos principios cuando sea necesario por el bien común.

Aplicación en casos concretos

No hay duda de que uno de los mantras del big data y la inteligencia artificial es la hiperpersonalización de contenidos, productos y servicios. Dentro de la cultura occidental este proceso encaja con su cultura individualista. Sin embargo, surgen reflexiones relevantes en este sentido.

El caso más obvio nace del derecho a la privacidad e incluso al anonimato. Por un lado, el derecho al olvido, pero, por otro, la profusión generalizada de requerimiento de datos para optar a determinados productos o servicios.

Habitual es en los servicios gratuitos, donde se asume que la contraprestación es la cesión de información personal a cambio del uso, sin que vayamos ahora a entrar en la proporcionalidad de este intercambio. Este hecho se ha extendido a circunstancias que pueden considerarse más anómalas, como es el caso de determinadas ofertas comerciales, que actúan de gancho, pero que para poder conocer en detalle las condiciones es necesario rellenar un formulario con datos personales e incluso consentir todo tipo de actividad comercial. Igualmente, empresas de reconocido prestigio y que dicen promocionar la divulgación científica y el conocimiento, ponen documentos a disposición del usuario, pero requiriendo la previa identificación e incluso la cesión del correo profesional, no sujeto a las mismas restricciones publicitarias que el correo personal.

Más allá de esta casuística, la hiperpersonalización también impacta de forma significativa en otros ámbitos, incluso adulterando la esencia de algún negocio. En concreto, el mundo de los seguros nace del principio de la mutualización, es decir, compartir riesgos cuantificables estadísticamente, pero indeterminados en cuanto a su impacto individual.

Sin embargo, en el ramo del automóvil es hoy posible obtener una batería enorme de datos que permite asociar un riesgo individual más preciso. Pensemos que entre los sensores del propio automóvil y los que integramos en el teléfono u otros *wereables* es posible

²² Patxi Barceló. Regulación de la IA en China: leyes clave (traducción al español no oficial). Ethalo Advisory – Serie legislación y jurisprudencia. 1ª edición mayo 2024. (#749)

²³ Como referencia global usamos la Recomendación sobre la ética de la Inteligencia Artificial. Unesco, 2021. (#689). Disponible en https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa. Este documento fue firmado por China y otros 192 países y fue pionero en la materia. Solo Rusia y la República de Irán han realizado matizaciones a su adhesión, en la medida en que pueda entrar en conflicto con sus propias regulaciones. Estados Unidos e Israel no han participado, dado que no pertenecen a la UNESCO.

determinar parámetros como la velocidad habitual, los cambios bruscos de dirección, la intensidad de las frenadas, las situaciones de emergencia de colisión, cambios involuntarios de carril, las horas de conducción, etc. Pensemos el grado de conocimiento que se puede adquirir y la posibilidad de individualizar el seguro según estos resultados. Pasamos de un modelo de mutualización a un modelo de previsión. Hay alguna compañía de seguros que ya ha comercializado seguros de vida donde la prima se ve reducida al portar un *wearable* que permita conocer el estilo de vida del usuario.

Este planteamiento, que puede parecer fronterizo, se complica si esos datos se usan para denegar el acceso al seguro médico a determinados individuos o si, en el caso del seguro del hogar, se observa como la prima de seguro sube sustancialmente porque se han recopilado datos que ahora identifican que la vivienda está situada en una zona fácilmente inundable, aunque esta lleva construida más de 30 años.

Podemos hacer la misma reflexión respecto a la personalización de ofertas y descuentos. Hoy muchos centros comerciales y supermercados ofrecen descuentos individualizados al finalizar la compra. Nos podemos preguntar cómo están construidas esas asignaciones, si en base a ampliar la base de consumo, aunque sea innecesaria; a incrementar el margen ofreciendo aquellos productos que más beneficio dejan o que otros criterios se siguen. Al ser descuentos individualizados, nos podemos preguntar si se está siendo coherente con la aplicación de criterios a todos los clientes o, como en el caso de los seguros que se comentaba al principio de este documento, se ajustan los resultados en función del comportamiento del usuario, de forma que a los calificados como más consumistas se les hacen ofertas menos atractivas, pues su propensión al consumo ya es alta. ¿Cuáles son las reglas del juego? ¿Están declaradas?

Conclusión base

La promesa de la individualización de los productos y servicios parece un reclamo atractivo, pero tiene sentido si es una opción del cliente y las reglas son claras.

Por un lado, el colectivismo nos obliga a reflexionar sobre si el resultado de estas políticas genera un resultado neto positivo, neutro o negativo para el conjunto. De esta forma podremos abordar con tranquilidad la prueba de la publicidad, es decir, ¿qué pasaría si nuestro modelo de decisión se transparentara en la portada de la prensa? Y, aunque en términos generales pudiera pasar la prueba, ¿el caso más extremo podría poner en duda el proceso de decisión? Si fuéramos nosotros el cliente ¿aceptaríamos el resultado?

Ética del empoderamiento

Planteamiento formal

El término empoderamiento tiene acepciones muy diversas según las culturas y el contexto (Narayan, 2002)²⁴. A los efectos que aquí nos interesa resaltar, podemos resumir su esencia a través del conocido proverbio chino “Dale un pez a una persona y comerá hoy. Enséñale a pescar y comerá el resto de su vida”.

²⁴ Narayan, D. (2002). Empoderamiento y reducción de la pobreza. Bogotá: Alfaomega / Banco Mundial.

Bula Caraballo (2012)²⁵ sugiere que el empoderamiento puede comprenderse a partir del concepto de “causa adecuada” de Spinoza, asociado a lo que éste entiende por “libertad”.

Somos libres, en la medida en que somos la causa de lo que nos acontece, y padecemos, somos esclavos, en la medida en que las causas de nuestro acontecer yacen fuera de nosotros. Parafraseando la canción de Simon & Garfunkel, se trata de ser el martillo y no el clavo.

Un individuo libre es aquel que ejerce efectivamente sus derechos en una sociedad. Esta definición es independiente del contexto cultural: no se afirma que las mujeres sean libres si usan la burka o si no la usan (en Francia se les prohíbe, en Irán están obligadas a llevarla); es libre en la medida en que la decisión a este respecto provenga del sujeto empoderado.

Un aumento en la libertad implica un aumento en la variedad, en sentido cibernético, de un sistema cualquiera. La variedad es el número de estados posibles de un sistema determinado (Beer, 1993: 21)

Preguntar por el empoderamiento de una comunidad es preguntar: ¿qué tanto son éstas personas causa, y no efecto, de lo que les acontece?

Por tanto, uniendo ambos conceptos y haciendo referencia al entorno tecnológico, estamos ante la confluencia de tres aspectos clave:

- la capacidad de comprender el sistema que vamos a usar;
- la posibilidad de usarlo sin restricciones limitantes de la libertad que no sean consustanciales a su propia naturaleza;
- e, incluso, la posibilidad de salir del sistema y disponer de alternativas adecuadas.

Aplicación en casos concretos

Esta concepción ética nos obliga a revisar la propia visión y misión de las organizaciones. Una primera pregunta apunta a saber si estamos prestando un producto o servicio que facilita la vida al usuario, pero al mismo tiempo lo sitúa en una posición de dependencia y necesidad.

Esa dependencia puede tener otras derivadas, como explica Omar Hatamleh²⁶:

En la época en la que yo iba al colegio, si querías aprender algo ibas a la biblioteca y sacabas libros, había muchos dónde elegir. Luego pasamos a internet, buscamos las cosas en Google, el orden en el que aparecen los resultados depende de intereses comerciales, pero tienes bastantes opciones. Ahora vamos a preguntar directamente a sistemas de IA como ChatGPT y la respuesta que dé normalmente va a ser la que va a aceptar la gente. Habría que ver qué compañía está detrás y cuáles son los intereses porque puede que la respuesta esté influenciada por quien controla ese sistema.

Las capacidades crecientes de los sistemas basados en inteligencia artificial implican una percepción cada vez mayor de superioridad e incluso infalibilidad. A modo de nuevos

²⁵ Germán Ulises Bula Caraballo. Spinoza: empoderamiento y ética de la composición. Universitas Philosophica 58, año 29: 197-215 (Junio 2012). Disponible en <http://www.scielo.org.co/pdf/unph/v29n58/v29n58a09.pdf>

²⁶ Citado por Teresa Guerrero. El Mundo 10 de septiembre de 2024. En <https://www.elmundo.es/papel/lideres/2024/09/09/66dedcb3e85ecef2678b45ab.html>

dioses cuyos designios son imperativos, volviendo a épocas ancestrales, donde el desconocimiento de las leyes naturales implicaba un temor reverencial hacia la naturaleza o quienes tenían poder para dominarla.

En este apartado, los ejemplos pueden ser casi infinitos. La mayor parte de los sistemas están desarrollados pensando en hacer la vida más fácil al usuario (siendo precisamente el motivo de venta y la razón de satisfacción del cliente). No obstante, es precisamente cuando el usuario adquiere cierta consciencia de otras necesidades o intereses cuando observa perplejo como el sistema no permite alternativas ni existen otras vías de acceso.

En algunos casos, esta situación se desarrolla en términos casi monopolísticos (muchos servicios de Google o Meta pueden encuadrarse en esta categoría), lo que genera un refuerzo de la pérdida de libertad.

Aterricemos en un ejemplo real que pone de manifiesto la realidad subyacente a muchos de estos sistemas. Uno de los principales bancos españoles publicita un servicio que proporciona información sobre el precio objetivo de las acciones a diferentes plazos basado en un sistema de inteligencia artificial.

Su propia documentación²⁷ explicita:

El proceso de estimación de precios implica la selección, adquisición, y transformación de variables, seguido del entrenamiento y optimización de modelos de Machine Learning. Es un proceso robusto, donde los modelos se validan y mejoran constantemente. Esto permite una mayor precisión y eficacia en la obtención de precios objetivos, tratando de aportar más información a inversores para que puedan adelantarse a los mercados, y a comprar y vender según sus intereses.

La publicidad del modelo, junto con el lenguaje empleado, parecen apuntar a un potente sistema para tomar decisiones de inversión con gran fiabilidad. Ofrece pescados para comer, bajo un modelo de pesca sofisticado y que no es posible personalizar.

Sin embargo, una revisión más detallada de la información disponible en otras fuentes²⁸, nos lleva a afirmaciones como:

En el caso de Amazon hemos visto que una de las variables que más afecta a la compañía históricamente, de esas más de 2.000 que tenemos en cuenta, es el tráfico en el puerto de Shanghái. Pero a la vez también tiene impacto en Amazon lo que haga el *treasury* a 3 meses.

¿Han detectado algún sector en el que la predictibilidad sea mayor y alguno en el que los aciertos sean menores? En tres meses, por ejemplo, donde más exactitud tenemos es en tecnología, con un 67%. A un mes, hemos llegado a tener una ratio de acierto del 64% en el sector financiero; a seis meses, del 67% en consumo cíclico y a 12 meses, del 87% en

²⁷ Disponible en:

https://www.openbank.es/assets/static/pdf/Inversiones/informacion_broker_inteligencia_artificial.pdf?_gl=1*vb0wx5*_gcl_au*MTI3ODcxOTY1Mi4xNzE1MjY2MzAz*_ga*MTE5MTA2MTQ2Ny4xNzE1MjY2MzAz*_ga_F8PG17BNY5*MTcxODE4MTEwMS43OS4xLjE3MTgxODI5NzkuMC4wLjA.&_ga=2.188779112.298530743.1718010688-1191061467.1715266303 a 10 de septiembre de 2024 (#V001)

²⁸ Fuente: <https://www.eleconomista.es/mercados-cotizaciones/noticias/12697202/02/24/gonzalo-pradas-openbank-el-grado-de-acierto-de-la-ia-en-la-prediccion-de-precios-objetivos-ya-es-del-56-pero-no-creo-que-llegue-al-80.html>

tecnología también. Los sectores en los que la exactitud ha sido menor han sido los siguientes: en el plazo de doce meses en el de *healthcare*, con un 37%; a seis meses, en el de *utilities*, con un 30%; a tres meses, en el de energía, con un 48% y a un mes en el de real estate, con un 50%.

Nos reservamos el derecho de admisión en cuanto a qué datos poner y preferimos no alimentar ruidos. Los bancos nacionales que están en el Stoxx Europe 600 no los incluiremos, y nos reservamos la discrecionalidad de quitar acciones que creamos que no tienen sentido porque haya polémicas o cosas que puedan confundir al inversor porque esta herramienta va de todo lo contrario, de ayudar a que se anticipen al mercado, de ayudar a invertir, de ayudar a que puedan tener estimaciones que para nosotros son serias.

Su lectura lleva a varias preguntas clave: ¿El sistema ayuda a mejorar las capacidades inversoras del usuario? ¿El usuario sabe que el proveedor puede decidir “que datos pone” sin más explicaciones? ¿El usuario conoce las limitaciones del sistema, hasta el punto de que determinados casos pueden tener un error superior al mero azar? ¿El usuario puede intervenir para observar las variables empleadas en cada caso o filtrar determinados aspectos?

Conclusión base

La tecnología no es neutra. Los procesos operativos condicionan la libertad de acción del usuario. La forma en que se pone a disposición la tecnología condiciona el resultado, generando usuarios dependientes o usuarios más libres y capacitados (empoderados).

Cada organización debe decidir el grado de transparencia del sistema y la forma de dicha información (generar conocimiento asequible) y el grado de libertad del usuario (generar libertad de alternativas). Estas decisiones adquieren mayor relevancia y responsabilidad en cuanto hablamos de procesos monopolísticos, no tanto por el alcance de ese monopolio (extensión o tamaño), sino por la inexistencia de alternativas razonables (lo tomas o lo dejas).

Ética de la responsabilidad

Planteamiento formal

Su origen está en el principio de responsabilidad de Hans Jonas, que invierte la prelación de términos en el ámbito de la ética. De una ética antropocéntrica a una ética global, donde el resto del ecosistema es igual de importante; de una ética de los actos a una ética de las posibilidades; de una ética de las consecuencias de los actos a una ética de los actos necesarios para lograr consecuencias.

En palabras del propio Jonas²⁹:

El poder y el saber eran en otro tiempo tan limitados que, en lo referente al futuro, casi todo había que confiarlo al destino y a la permanencia del orden natural; por ello toda la atención se dirigía al recto hacer lo actual y concreto. Y donde mejor asegurado estaba el recto hacer

²⁹ Hans Jonas (1979). El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica. Herder. (#803). Disponible en <https://etica.uazuay.edu.ec/sites/etica.uazuay.edu.ec/files/public/uazuay-etica-principio-de-la-responsabilidad-hans-jonas.pdf>

era en el recto ser. La ética, en consecuencia, tenía que ver principalmente con la «virtud», que representa el mejor ser posible del hombre, y atendía poco a lo que no era próximo.

En la era de la civilización técnica, que ha llegado a ser «omnipotente» de modo negativo, el primer deber del comportamiento humano colectivo es el futuro de los hombres. En él está manifiestamente contenido el futuro de la naturaleza como condición sine qua non; pero además, independientemente de ello, el futuro de la naturaleza es de suyo una responsabilidad metafísica, una vez que el hombre no sólo se ha convertido en un peligro para sí mismo, sino también para toda la biosfera.

Para ello, es imprescindible una análisis del futuro, con mirada prudente, identificando escenarios y decidiendo, activamente, acciones que eviten los escenarios no deseados. No vale el inmovilismo. En definitiva, una gestión proactiva de riesgos.

Terrones (2018)³⁰ nos ayuda a esbozar esta visión:

El poder de la tecnología es inconmensurable e impredecible, las consecuencias de la acción humana pueden implicar un alto porcentaje de riesgo.

La lógica de producción tecnocientífica ha conducido al hombre a un cierto nihilismo, que desemboca en inmovilismo, como si no fuera posible cuestionarse nada que tuviera que ver con el progreso tecnológico.

Sin embargo, se trata de partir desde cierto velo de ignorancia para llegar a un saber propio de la ética de una dimensión nueva, una ética vinculada con el saber predictivo para poder acompañar y guiar la vigilancia a que le debe ser sometido el poder de la tecnociencia.

El poder cognitivo recae sobre la imaginación y sobre el sentimiento, porque nos anticipamos para conocer y a la misma vez nos conmovemos, algo fundamental para asumir responsabilidad.

En definitiva, la ética de la responsabilidad apela directamente a la función más estratégica del consejo. La dirección y la pervivencia. La gestión de riesgos. La obligación de actuar (la culpa por omisión) en un nuevo entorno donde las organizaciones pasan a tener una capacidad no sólo inconmensurable, sino también inmediata. Por tanto, es responsabilidad de cada organización prever los escenarios posibles, aunque puedan ser remotos y asumir como regla de acción no sólo la sostenibilidad de la empresa, sino la pervivencia del entorno amplio. Además, dicha responsabilidad debe perdurar en el tiempo. Un cambio de las condiciones del entorno no exime de la responsabilidad.

Aplicación en casos concretos

Existen numerosos ejemplos en otras industrias que pueden servirnos de analogía. La industria del tabaco, la talidomida o el uso de glifosatos en los pesticidas son casos reales donde la ética de la responsabilidad es aplicable. ¿Se hicieron las preguntas adecuadas sobre el futuro? ¿Se midieron las consecuencias?

Adicciones, trastornos de la personalidad, pérdida de capacidades cognitivas, incitación a la violencia, facilitación de la comisión de delitos, violaciones de la privacidad, armas de

³⁰ Terrones Rodríguez, A. L. (2018). Inteligencia artificial y ética de la responsabilidad. Cuestiones de Filosofía, 4(22), 141–170. <https://doi.org/10.19053/01235095.v4.n22.2018.8311> (#775)

ingeniería social y otros tantos atributos que pueden llegar a asociarse a determinado sistema.

La detención en París del propietario de Telegram, bajo la acusación de 12 delitos³¹ relacionados con transacciones ilícitas, pornografía infantil, fraude y negativa a comunicar información a las autoridades, es un claro exponente, más allá de otras cuestiones geoestratégicas, de la realidad de estas situaciones.

Conclusión base

La inteligencia artificial, que es un artefacto fundamentalmente de base estadística, obliga a usar esa misma base en la gestión de las organizaciones. Frente a un modelo normativo monocromo, de blanco sobre negro (“lo que no está prohibido, está permitido”), nos enfrentamos a un entorno mucho más complejo, donde los actos, tanto los realizados como los no realizados (omisión) pueden provocar fuertes seísmos futuros.

Las organizaciones tienen que integrar la gestión de riesgos como un elemento central de sus procesos de decisión. La gestión de riesgos identifica posibilidades, las mide en términos probabilísticos, identifica medidas de mitigación y las aplica, para situar el nivel de riesgo asumido por la organización dentro del nivel de tolerancia aceptado. Y cómo estamos en un entorno incierto, establece cortafuegos y otras medidas que tratan de reducir también el efecto de lo muy improbable. Siendo conscientes además de que cada suceso muy improbable, en sí mismo, no tiene relevancia, pero el conjunto de sucesos improbables es tan grande, que la probabilidad de que alguno suceda es significativa.

Ética de la gobernanza

Planteamiento formal

Tiene su origen en un aspecto común a diversas culturas, aunque en muchas ocasiones criticado como paternalista. En Atenas era la aristocracia, en Europa la ilustración y en Asia el confucianismo al poner el énfasis en la responsabilidad del gobernante para con el pueblo.

Quizás sea precisamente la forma de expresarlo en la versión de la filosofía oriental la que más luz nos puede dar. El foco no está en que el gobernante debe tener las capacidades para dirigir al pueblo, que en la masa demuestra su ignorancia. Más bien el énfasis debe estar precisamente en la responsabilidad del gobernante para, desde la humildad, ser capaz de buscar lo más adecuado. Y desde esa humildad, la participación de todas las partes es una pieza esencial (y no sólo de las mayorías que puedan garantizar su mantenimiento en el poder, que es el talón de Aquiles de las democracias occidentales).

Precisamente este enfoque permite visualizar la humildad desde el otro lado del espejo. En realidad, la humildad no es más que el reconocimiento de la complejidad y, por tanto, la imposibilidad de que un único punto de vista o unos pocos puedan contener los elementos suficientes para tomar decisiones. Por tanto, la participación no es una delegación de

³¹ Fuente: <https://elpais.com/internacional/2024-08-26/un-juez-prolonga-la-detencion-en-francia-de-pavel-durov-fundador-de-telegram.html>

responsabilidad ni una forma de justificación, es un proceso que permite interiorizar la complejidad para poder tomar decisiones desde una perspectiva amplia.

Como señala Terrones (2020)³², citando a Moratalla y Gracia:

Una ética de este tipo nace de la posibilidad de reconocer perspectivas, pues el ámbito de la ética es contingente, y como señala Moratalla: “La deliberación es el método de análisis ético una vez que reconocemos la contingencia de los asuntos humanos y comprendemos que nuestro ámbito de acción y decisión es el de la incertidumbre y la complejidad. (2017:41)

La actitud deliberativa implica la necesidad de alejarse de determinados “vicios” propios del modo de pensar lineal deductivo, donde todo se reduce a aspectos cuantificables y dilemas. Según Gracia, si queremos emprender la vía deliberativa, es necesario huir de los dilemas, pues “más que dilemas, hay problemas, es decir casos con múltiples cursos de acción posibles que será preciso tener en cuenta a la hora de tomar una decisión razonable o prudente”. (2016:13)

Aplicación en casos concretos

Los sistemas basados en inteligencia artificial pueden ser capaces de autorregularse y modificar su conducta con nuevos datos; el proceso de desarrollo del sistema implica múltiples partes con perspectivas diferentes; su construcción requiere decidir determinados parámetros y asunciones que impactan en el resultado final; el sistema en ocasiones requiere actualizaciones o parches que modifican su comportamiento; su aplicación puede extenderse en ámbitos no previamente considerados, y, a su vez, puede depender de otros sistemas automáticos o manuales, los datos de construcción pueden tener errores, vicios o incluso cambios de criterio no considerados y otras muchas vicisitudes pueden ocurrir en el proceso.

Por tanto, la complejidad de estos sistemas es un hecho, no desde el punto de vista técnico, sino de proceso. La realidad es que la asignación de responsabilidad en este entorno puede ser una tarea imposible, motivo por el que existen amplias voces que reclaman la aplicación de principios de responsabilidad objetiva, de responsabilidad por riesgo o de inversión de carga de la prueba. Y más allá de las consecuencias legales y económicas, no pueden olvidarse las consecuencias reputacionales.

Por ello, más allá de la necesidad de participación de todas las partes interesadas, lo cual exige un cambio en los procedimientos internos de la mayor parte de las organizaciones, es necesario también evaluar el papel de la dirección en la decisión de implantación de estos sistemas.

Conclusión base

La complejidad y la participación no implica delegación, como se especificaba más arriba en el planteamiento formal de la cuestión. El consejo debe adoptar un papel activo de indagación. Como hemos visto a lo largo de los numerosos ejemplos expuestos, el foco no

³² Citados por Antonio Luis Terrones Rodríguez (2020). Inteligencia artificial, responsabilidad y compromiso cívico y democrático. Revista CTS, nº 44, vol. 15, junio de 2020 (253-276). (#794). Disponible en <https://ojs.revistacts.net/index.php/CTS/article/view/166/157>

está en la complejidad del algoritmo, sino en su implantación efectiva y su impacto en los usuarios. Es decir, en este punto, podemos entrelazar el concepto de ética con la experiencia de usuario. Como se indicaba más arriba, la tecnología impacta en el modelo de relación y en la configuración de la propia persona, siendo este impacto el que debe ser evaluado a la luz de las intenciones de la organización, la sensibilidad del entorno y la diligencia debida con el presente y futuro de la sociedad y las personas.

Nuevo materialismo

Planteamiento formal

El nuevo materialismo rompe con el dualismo ontológico (materia y espíritu) y el antropocentrismo. Parte del difuminado de las fronteras entre el hombre y la máquina y del hombre y el entorno. Son las interacciones entre los diferentes entes (en sentido amplio, más allá del concepto de seres vivos, denominados fenómenos) lo que constituye a estas. No puede entenderse un ente fuera de sus interacciones. El hombre no es un ente principal, incluso el propio concepto de hombre no es fijo e inmutable, sino que depende de su contexto. El hombre es parte del mundo, no es algo que vive en el mundo, con una realidad independiente.³³

Esta conceptualización del mundo tiene una especial trascendencia en el ámbito de las nuevas tecnologías, pues entiende que, en ese concepto de hombre variable, la introducción de la tecnología cambia la propia naturaleza del hombre. Da entrada a los conceptos de *cyborg* o de la inteligencia artificial como entes que también forman parte del mundo e interactúan con él, transformándolo.

En este sentido, todos los componentes de la naturaleza, incluso a la luz de los avances de la física moderna, se convierten en fuerzas vivas. Como dice Franca Maccioni (2020)³⁴:

Dichos avances obligan a abandonar las consideraciones sustancialistas de la materia y a asumir, en cambio, que cuando nos enfrentamos a procesos de materialización, tratamos con fuerzas, energías e intensidades más que con sustancias, y con procesos complejos, incluso *randoms*, más que con estados simples o predecibles.

Todo es material, pero la materia es siempre algo más que “mera” masa determinada y estable: un exceso, fuerzas plurales, procesos relacionales complejos mantienen a la materia activa, creativa, productiva, impredecible, es decir, en constante proceso de materialización.

Considerar que la materia posee sus propios modos de auto-transformación, autoorganización y direccionalidad, es decir, ya no pensarla simplemente como pasiva o inerte, perturba, también, el sentido común que considera que los agentes son exclusivamente humanos.

Esta conceptualización del mundo nos permite llegar a una sostenibilidad integral. Más allá de la sostenibilidad de moda, entendida como preservar el entorno para seguir pudiendo

³³ Para más detalle, ver Puzio, A. The entangled human being – a new materialist approach to anthropology of technology. *AI Ethics* (2024). <https://doi.org/10.1007/s43681-024-00537-z> (#802)

³⁴ Maccioni, F. & Jorge, J. (2022). Nuevos Materialismos. Aproximaciones al materialismo vibrante de Jane Bennet. *Cuadernos Del Sur Letras*, (52), 167–176. Recuperado a partir de <https://revistas.uns.edu.ar/csl/article/view/3815> (#F001)

aprovecharlo en el futuro, se genera una cosmovisión integradora donde el ser humano es un ente más en una compleja interacción de todo tipo de entes (donde la inteligencia artificial puede ser otro ente, en la medida en que interactúa y transforma el entorno).

Si desarrollamos esta idea, podemos incluso imaginar al hombre como un microbio (en su sentido etimológico de vida diminuta) que forma parte de la naturaleza (entendida como el universo, tanto conocido como no conocido).

Desde esa perspectiva, el hombre habría evolucionado de ser una bacteria del universo, en un principio miedosa como un ratón, que poco a poco fue adquiriendo una autoconsciencia que le hizo sentirse fuerte dadas sus características intrínsecas, como un león en la sabana.

Con el tiempo, esa bacteria, ya evolucionada a virus por su mayor complejidad e interacción con el mundo, empezó a ser consciente de su capacidad de dominar el mundo, convirtiéndose en un parásito, en la medida en que comienza a aprovechar los recursos que le rodean (otros entes) para sus propios fines, incluso a costa del entorno.

Cuando la virulencia de esa actitud es exagerada, se llega a convertir en un tumor, una masa de células que crecen de forma descontrolada y destruyen el entorno más próximo. Incluso comienzan las mutaciones, generando otros entes (entre ellos, la inteligencia artificial) que refuerzan esa dominación descontrolada, hasta el punto de que dicha ausencia de control puede llegar a afectar al propio hombre, convertido ahora en dios creador de otros entes.

Visualicemos esta situación en términos prácticos. El propio hombre o sus creaciones – la inteligencia artificial- tienen la capacidad de autodestruirse, por ejemplo, generando una deflagración nuclear de alcance planetario. No obstante, esto no alteraría el cosmos en su conjunto, simplemente daría espacio a que la naturaleza desarrolle otros entes que ocuparían, con el tiempo, el espacio destruido. El hombre en su concepto más estricto de “vida diminuta” (bacteria) para el conjunto del universo, desaparecería. La naturaleza seguiría su curso y el resto de las interrelaciones entre entes se verían afectadas parcialmente, pero no desaparecerían.

Aplicación en casos concretos

El nuevo materialismo nos lleva directamente a reflexionar sobre el impacto de las nuevas tecnologías en el entorno. Estas tienen la capacidad de modelar lo que es el hombre y cambiar su propio concepto.

Esto no es nada nuevo. No hay duda de que la aparición a lo largo de la historia de la tecnología (la rueda, la imprenta, la máquina de vapor, ...) ha supuesto un cambio en nuestro propio autoconcepto. Como se señalaba antes, incluso pasar de sentirse una parte indefensa del entorno a tener la capacidad de dominarlo y someterlo.

No obstante, estos cambios han sido siempre lentos, pues requerían de una difusión de las tecnologías y de cambios en los hábitos de vida que implicaban largos periodos, acompañados de un debate social sobre su modo de integración.

Sin embargo, el nuevo ecosistema digital ha cambiado las reglas de juego. Hoy se da la confluencia de dos elementos clave: la capacidad de influir de forma directa en las personas y de hacerlo de forma masiva e inmediata.

Los procesos de priorización y viralización de contenidos, de retención de los usuarios, de administración de recompensas y otras técnicas de neuro impacto tienen la capacidad de conformar el mundo de creencias de las personas y de hacerlo de forma masiva e inmediata en el tiempo.

Como ejemplo ilustrativo, es interesante ver la recopilación de patrones adictivos recopilados en un informe de la Agencia Española de Protección de Datos y que son una muestra de esa enorme capacidad de conformar el entorno³⁵. No hay duda de que hoy, los videos de Tik-tok y los contenidos de Instagram, por poner ejemplos muy integrados en la cultura occidental, están influyendo en la conceptualización del mundo y de la ética de las relaciones de varias generaciones, con mayor intensidad y direccionalidad de lo que pueden contrarrestar los procesos de formación tradicionales, incluyendo la propia familia.

Por tanto, es imprescindible que la gobernanza de la sostenibilidad beba de estas fuentes. La sostenibilidad entendida como la responsabilidad sobre el impacto que nuestras acciones como organización tienen, no solo en el entorno ecológico, sino en la conformación de la sociedad del futuro y en la mecánica de interacciones con el resto de los entes. Un reto que, al mismo tiempo, da un sentido mucho más trascendente a la actividad de la propia empresa y puede usarse de palanca de cohesión y regla de comportamiento.

Conclusión base

La capacidad digital permite a las organizaciones multiplicar de forma exponencial sus capacidades y poder de influencia. Cuando estamos ante una relación entre iguales, esto implica mayores riesgos y por tanto requiere mayores cautelas; pero cuando esta situación genera una asimetría de proporciones gigantescas entre la empresa y los usuarios, se hace evidente que la responsabilidad de la empresa por sus actos y el nivel de exigencia debe incrementarse de forma proporcional, y la elusión del debate y la gobernanza interna no es una opción, especialmente al nivel del consejo.

¿Se puede introducir la ética en los sistemas de IA?

La respuesta es negativa en términos absolutos. En la medida en que aceptamos la necesidad de participación de las partes afectadas (donde las personas son un elemento central) y que el contexto es relevante, tratar de establecer reglas unívocas de regulación atacaría estos principios.

Esto no es obstáculo para que los sistemas de IA deban tener protocolos de actuación claros y transparentes, definidos a priori, que sirvan de guía en su acción. Pero esa

³⁵ VVAA (2024). Patrones adictivos en el tratamiento de datos personales. Implicaciones para la protección de datos. Agencia Española de Protección de Datos (AEPD). (#798). Disponible en <https://www.aepd.es/guias/patrones-adictivos-en-tratamiento-de-datos-personales.pdf>

circunstancia, necesaria por sí misma, no exime de la existencia de otros protocolos para revisar las situaciones que lo requieran, gestionar las excepciones y asumir la responsabilidad por aquellas acciones inadecuadas.

Esta postura implica negar la autonomía plena de los sistemas de IA. Algo que hoy seguramente sea aceptado de forma mayoritaria, pero que la lógica del mercado llegará a ponerlo en duda. Es ahí donde se puede producir una bifurcación significativa en la historia de la humanidad, en la medida en que se pueda admitir esa autonomía objetiva, que implicaría enfrentar a dos piezas que sólo pueden sobrevivir conjuntamente en términos de igualdad individual.

Cuestión muy diferente es la necesidad de que exista una regulación expresa que asegure la debida diligencia de todos los actores y la posibilidad de reparar los daños. Desde esa perspectiva, debemos recordar que la obligación de indemnizar cuando ocurre un daño depende fundamentalmente de la ocurrencia de un hecho antijurídico, la ocurrencia del daño y la existencia de una relación de causalidad entre el hecho y el daño (Pérez Orozco, 2024³⁶). La existencia de un hecho antijurídico es crucial y la ausencia de regulación supone un impedimento efectivo a la reparación.

Por ello, tres aspectos parecen clave para un desarrollo sostenible de la IA:

- Una regulación suficientemente abierta para incluir la responsabilidad de forma objetiva, de forma que la inacción pueda ser considerada un desencadenante del hecho antijurídico.
- Una interpretación analógica razonable de la ley desde otros ámbitos, de forma que, ante la falta de regulación, pueda operar la vieja máxima del comportamiento esperado de un buen padre de familia.
- Un compromiso de las organizaciones por expresar en concreto su compromiso, a modo de autodeclaración responsable vinculante en las relaciones con terceros.

A modo de conclusiones

Existen múltiples visiones éticas. Cada entorno cultural, e incluso la dimensión tiempo, marcan la preponderancia de unas sobre otras en un tiempo y lugar concreto. Ni el mínimo común múltiplo ni la acumulación de todas las alternativas son soluciones adecuadas.

Un campo de juego homogéneo, propiciado por una legislación lo más consensuada posible, es un primer paso.

No obstante, es necesario ir más allá. Cada organización debe hacer explícito su posicionamiento ético, con suficiente detalle como para permitir a los usuarios tomarlo en consideración en su proceso de selección y para exigir responsabilidad en caso de incumplimiento.

³⁶ Luis Pérez Orozco (2024). Daños ocasionados por sistemas de inteligencia artificial: desarrollo vs derechos (Inteligencia artificial para un futuro sostenible: Desafíos jurídicos y éticos). Casa Leiria. (#N001)

Para ello, es imprescindible que las organizaciones reflexionen desde diferentes perspectivas y desarrollen su apetito de riesgo en el ámbito digital, así como que establezcan los procedimientos de gestión de riesgos que aseguran que se mantienen dentro de los límites auto impuestos.



Boosting a healthful data driven society



www.ethalo.es

 info@ethalo.es