TAREA_50

ÍNDICE

Reflexión y organización de las ideas fundamentales

Introducción:

La inteligencia artificial (IA) es, en informática, la inteligencia expresada por máquinas, procurando que su funcionamiento se asemeje al cuerpo, cerebro y mente de los seres vivos, a diferencia de la inteligencia natural demostrada por humanos y ciertos animales con cerebros complejos.

El propósito último de la IA: lograr que una máquina posea una inteligencia de tipo general similar a la humana.
El principal problema al que se enfrenta la inteligencia artificial es la adquisición de conocimientos de sentido común. Este constituye el requisito fundamental para que las máquinas actuales sustituyan la inteligencia artificial especializada por una de tipo general.

Como dato final, se ha de mencionar que el combustible de la IA son los datos, y tal y como está evolucionando todo el sistema, esta claro que cada vez hay mas fuentes de donde sacar toda esta información y si no se filtra de la manera adecuada, se corre el riesgo de perder el control sobre esta tecnología. Ahí es donde entra la ética en este mundo.

Principios Éticos:

Dado que es una tecnología que está en pleno auge, aún se está trabajando en los principios fundamentales que rijan la ética en este sector. Como es lógico, estos principios se han desarrollado de la mano de los mayores gigantes tecnológicos, Google y Microsoft.
Aunque también se ha de mencionar que este es un tema que aun tiene mucho trabajo en desarrollo y que por mucho que se intente acotar el rango de principios básicos a respetar, todavía no se ha llegado a una conclusión final de cuales si y cuales no. Si se investiga en las propuestas de diferentes países, todos tienen puntos en común, y puntos en desacuerdo, de forma que en esta parte del informe, se le va a dar importancia a los siguiente más comunes.

Los principios desarrollados se basan en lo siguiente:

· Beneficio social: impacta en sectores con un gran alcance social. SANIDAD, ENERGIA, TRANSPORTE… Ejemplo1.

· Inclusivo: es imprescindible que en el desarrollo de la IA evite reforzar, fortalecer o reflejar sesgos injustos hacia colectivos de la sociedad. Ejemplo 2.

· Garante de la privacidad: Los datos sobre los que se sustenta el desarrollo de la inteligencia artificial deben haber sido obtenidos con el consentimiento expreso de sus propietarios, como recoge el RGPD. Ejemplo3.

· Transparente y explicable: Los modelos de inteligencia artificial deben ser capaces de explicar por qué toman determinadas decisiones de forma transparente y comprensible, para que la tecnología cuente con la confianza de todos los actores implicados. Ejemplo4.

Para cada uno de estos principios éticos, se plantea un ejemplo real, donde se ve claramente que si no se aplica la ética correctamente, puede ser perjudicial tanto para el usuario como para la empresa.

Ejemplo1:

“Existe un software comercial que, si bien tiene obviamente la intención de publicitarse y ser vendido hasta el último rincón del mercado, aporta unas estadísticas de acierto bastante contundentes.
El software en cuestión puede hacer alarde de haber acertado en sus predicciones de los movimientos del S&P 500 con una precisión de hasta el 79% hasta Julio de 2019, y en base a datos monitorizados desde enero de este mismo año.”

Ejemplo 2:

“El chatbot botTay de Microsoft basado en IA “estuvo navegando por sí solo en Twitter y al cabo de unas horas empezó a publicar tuits racistas y misóginos porque había cogido lo mejor de cada casa en esta red social”. A las 16 horas del lanzamiento la firma tuvo que desactivarlo.”

Ejemplo3:

“Mercadona pagará una multa por instalar un sistema que detectaba a personas con orden alejamiento.
La sanción ha sido propuesta por la Agencia Española de Protección de Datos (AEPD) como penalización por la instalación en varios supermercados de un sistema que permitía detectar personas con orden de alejamiento de sus establecimientos.”

Ejemplo4:

Referente a vehículos eléctricos: Cual es la capacidad de estas máquinas para tomar una decisión ante determinadas situaciones de peligro? sobre todo aquellas en las que el vehículo no tenía más remedio que decidir entre la vida de una persona u otra. Su respuesta sería transparente y se podría defender el porqué de su decisión?

Como conclusión a este punto: ¿Qué es ético y que no?
Llevando la cuestión a nuestro terreno, supongamos que hablamos de matar a una persona para salvar a otras cinco (el típico dilema moral que alguna vez en nuestra vida nos han planteado). Casi todo el todo el mundo estaría de acuerdo en que es preferible atropellar a esa persona para salvar a las otras cinco. Hay muchos factores humanos que podrían justificar esta decisión, como la ausencia de cercanía a esa persona (no es lo mismo matar a un desconocido que a alguien que tratamos y conocemos personalmente). Sin embargo, es difícil (por lo menos ahora) que una máquina pueda calibrar este tipo de decisiones, por eso es tan importante que la inteligencia artificial pueda estar dotada de una ética, o código de valores, que condicione, a nuestra imagen y semejanza, la esencia de sus actos.

¿Son los datos justos los datos?

Como respuesta a esta cuestión, se hace referencia al Ejemplo 2 mencionado anteriormente.

Ningún sistema de inteligencia artificial actual tiene intencionalidad, pero las decisiones que aprende están basadas en los datos con los cuales ha sido entrenado. Si esos datos están sesgados (pueden estarlo intencionadamente o no), el algoritmo decidirá sesgado, tal y como se aprecia en el Ejemplo2. El algoritmo coge referencias de comentarios racistas de la red social, aprende con esos datos, y devuelve una respuesta similar.

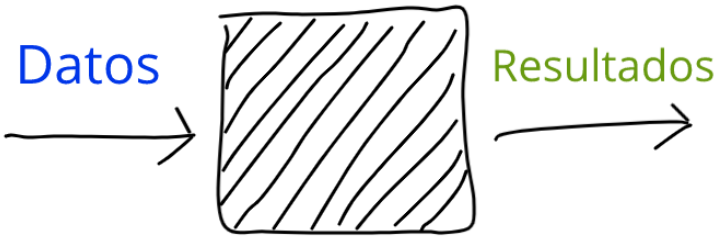
¿El algoritmo ha hecho bien su trabajo? ¿Deberían haberle aplicado mas o mejores filtros? Está claro que el chatbot discrimina porque es lo que ha aprendido de los datos de las personas que ha analizado, pero podemos tomar esto como concluyente? ¿El error es del algoritmo o de la persona que lo ha codificado?
Según afirma Carlos Castillo, director del grupo de Ciencia Web y Computación Social en la UPF: “Hay dos formas en que un sistema de inteligencia artificial puede mostrar prejuicios: primero, porque se usen datos inadecuados y, segundo, porque el procesamiento de los datos sea inadecuado”.

Este hombre también relata lo siguiente:
Los datos también pueden ser inadecuados de muchas maneras:

- Que contengan patrones históricos de discriminación. Por ejemplo, que aprenda de las estadísticas que los cargos ejecutivos son mayoritariamente desempeñados por hombres blancos y a la hora de seleccionar candidatos para una vacante de este tipo descarte currículums de mujeres y de hombres de raza negra
- Porque se seleccionen mal. “Quizá monitorizar el tráfico de coches es más fácil que monitorizar el de bicicletas o el de desplazamientos a pie, pero si seleccionamos solamente los datos de automóviles entonces propondremos políticas de movilidad más adecuadas para los viajes en coche que para otros viajes.

¿Cómo hacemos para que los análisis y resultados de la IA sean explicables y precisos?:

Uno de los problemas de la inteligencia artificial, en especial de las redes neuronales, es la falta de entendimiento acerca de cómo arrojan resultados, es decir, los algoritmos funcionan como cajas negras, donde se conocen las entradas y salidas pero no se entiende al 100% lo que pasa dentro (de ahí la expresión caja negra).



Para solventar este problema, se ha generado el nuevo campo llamado: XAI (eXplainable Artificial Intelligence) ya que al parecer, cuanto mas complejo es un algoritmo, mejor rendimiento tiene, de forma que la explicabilidad y entendimiento se sacrifican por conseguir mejores resultados. Hay empresas e instituciones como la DARPA, que están trabajando en ello, procurando crear técnicas de machine learning que produzcan modelos más entendibles sin sacrificar rendimiento, facilitando así a las personas que los algoritmos sean mas entendibles sin perder efectividad, de modo que puedan ser controlados con mayor efectividad.

La forma de analizar correctamente los datos mas influyentes es la siguiente:

- Analizar cuantas variables son dependientes e independientes entre sí
- Obtener registros en formato de base de datos, por ejemplo
- Aplicar el algoritmo correcto de análisis
- Obtener un porcentaje de efectividad y si es posible mejorarlo

Una vez hecho el análisis de como repercute cada variable sobre el problema, se puede definir cuáles de estas variables son sensibles a generar cambios mas grandes, de forma que con el correcto entrenamiento del algoritmo y tratamiento de los datos, es posible desarrollar soluciones adaptadas a cada situación.

¿Cuáles son los principios fundamentales que ninguna IA debería vulnerar?

Profundizando en las ideas fundamentales mencionadas hojas atrás, se hace un análisis mas extenso sobre el respeto de la IA hacia los derechos humanos. Aquí se mencionan los principios fundamentales que ninguna IA debería vulnerar:

Destrucción de empleo y deshumanización

Los algoritmos de inteligencia artificial se utilizan cada vez más en la contratación, el despido y las decisiones diarias de la vida laboral, pero la ley no se ha mantenido al día, de forma que a menos que todo se regularice al mismo ritmo, corremos el riesgo de llegar a la deshumanización, favoreciendo a las máquinas mas que a las personas.
«La inteligencia artificial debe usarse para apoyar a los trabajadores y la sociedad en general, haciendo que la vida laboral sea más fácil y más eficiente», dijo un portavoz del gobierno en un comunicado.

Digitalización del trabajo y desafíos para privacidad e igualdad

La ley de Protección de Datos y Garantía de Derechos Digitales (LOPDGDD) está diseñada para proteger la privacidad de los ciudadanos de la UE y darles más control sobre sus datos personales. Su objetivo es establecer una nueva relación entre el usuario y el sistema, una donde la transparencia y un estándar de privacidad no sean negociables.
El artículo 22 del RGPD somete a requisitos adicionales solo aquellas decisiones basadas en el procesamiento automatizado y / o la elaboración de perfiles que producen un efecto legal o significativo para los individuos. Además, estas decisiones deben tomarse únicamente por medios automatizados, sin intervención humana.

Protección de derechos humanos en el trabajo digital

A la hora de examinar la intersección entre la inteligencia artificial y los derechos humanos se han de mencionar dos cosas:
La primera es que existe un interés amplio en las cuestiones éticas en torno a la IA, tanto en el mundo académico como en el laboral.
La segunda es que las normas de derechos humanos no estaban presentes en el debate sobre la ética de la IA. En varias ocasiones se menciona la importancia de los derechos humanos en este ámbito, pero por lo general no era más que una referencia pasajera.

Generalmente, muchos de los algoritmos creados para filtrar personas en contrataciones o vida laboral en general, acaban favoreciendo a los hombres sobre las mujeres, penalizando a las minorías étnicas y juzgando a las personas por su expresión facial.

Derecho a tutela judicial efectiva

El hecho de implantar IA en todo lo mencionado hasta ahora, preocupa, no por que la IA debiera quedar excluida totalmente de nuestra sociedad sino porque, partiendo de los beneficios que su utilización pueda acarrear, su implantación debe estar precedida de un análisis profundo y detenido de los riesgos y las implicaciones que ello supone, y debe ir

acompañada de las garantías necesarias que aseguren el respeto a la tutela judicial efectiva.

Además, habrá que evitar que se produzca una superposición de sistemas de IA (en fase administrativa y en fase judicial) que conduzcan a un encadenamiento de decisiones automatizadas que dificulten o impidan el análisis del caso concreto y la elaboración de una motivación real y suficiente que sea garantía de la tutela judicial efectiva reclamada.

Libertad de expresión

La Inteligencia Artificial puede presentarse como un aliado al momento de moderar contenidos violentos o de noticias aparentes, pero su utilización sin intervención humana que contextualice y traduzca adecuadamente la expresión deja abierto el riesgo de que se genere censura previa.

En la actualidad esto se encuentra en debate dentro del ámbito internacional dado que, al carecer la Inteligencia Artificial de la capacidad para contextualizar lo que modera, se ésta presentando más como una herramienta de censura previa indiscriminada, que como una moderación en busca de proteger la libertad de expresión, por tanto, debería dársele el peso que tiene basándose en el siguiente fragmento:

Citación de la vigencia de la OC 5/85 dictada por la Corte Interamericana que señala que:

“la libertad de expresión no se agota en su faz individual, sino que comprende además a la dimensión colectiva, subrayando que el libre pensamiento y su difusión son inseparables, de forma tal que una limitación previa —estatal o privada— a cualquiera de ellos, sería incompatible con los estándares interamericanos que protegen a este derecho (Sec. Gral. OEA, 2017).”

¿Deberían homologarse los desarrollos de IA? ¿Deberían pasar por una revisión para validar su certificación?

Una forma de revisión de IA es el conocido Lovelace 2.0. Se trata de un modelo mejorado de un test propuesto en 2001 (cuya función era pedir a una unidad de IA crear algo que fuera incapaz de explicar cómo había sido creado).

Lovelace 2.0 desarrolla un poco más esa idea de la siguiente manera:
"Para superar este test, el agente artificial debe desarrollar un artefacto creativo (pintura, poema, una historia de ficción, un diseño arquitectónico…) a partir de una serie de géneros artísticos que requieren un mínimo desarrollo de inteligencia. Además, el artefacto debe cumplir con ciertas limitaciones que son impuestas por el evaluador humano".
Actualmente algunos algoritmos han creado historias y pinturas, aunque "ninguno ha pasado el test Lovelace 2.0".

Viendo esto, y como opinión personal, creo que los desarrollos en IA si que deberían de tener una regulación que los haga pasar por una serie de pruebas o cumplimentación de requisitos, de forma que puedan desplegarse siempre bajo la tutela y criterio del ser humano (cualidades que siempre deberían prevalecer sobre los algoritmos de IA).

¿Qué papel juegan la sociedad y los gobiernos en la definición de estos principios?

La sociedad y los gobiernos deben trabajar juntos para la resolución final de los principios mencionados anteriormente, ya que no tendría sentido que un gobierno decrete ciertos criterios que la sociedad no este dispuesta a cumplir o aceptar (refiriéndonos a la IA, desde luego).
Un Gobierno que quiera implementar "seguridad" en las calles añadiendo cámaras de vigilancia con reconocimiento facial, que vulnere los derechos de privacidad de la sociedad al estilo Orwell 1984, no debería llevar ese plan a cabo sin la aprobación de la gente.
De modo inverso, una propuesta que no sea aceptada por la sociedad, mediante la cual el gobierno quiera/deba ayudar a la gente, y nosotros no seamos capaces de dar el paso para validarla, también puede ser perjudicial para el desarrollo tecnológico en general.

De este modo, la única forma de poder definir correctamente que principios deben actuar en el ámbito de la IA, englobando tanto a gobiernos como personas de a pie, es la de que entre ambos se lleguen a acuerdos mediante el dialogo, reuniones, o comités.
Es un tema delicado y muy complejo, ya que en un mismo grupo de gente habrá mucha diversidad de opiniones referentes al mismo tema, y probablemente las decisiones de unos irán en contra de otros. De modo que la manera de llegar a un balance entre opiniones quizá podría ser la de: ¿el bien común?

Otra opción podría ser la de que cada individuo acepte las propuestas que mejor se adapten a si mismo (elegir entre todas las que hay, en cuales está de acuerdo y en cuales no), siempre y cuando se comprometa a no poner en riesgo y/o discriminar las propuestas elegidas por otra persona, y sepa convivir en sociedad sabiendo que las propuestas elegidas para si mism@ pueden no ser iguales a las de su gente alrededor. De esta forma, cada uno tendría mano ancha para convivir con los principios que ha elegido respetar (algo así como los partidos políticos. Cada uno tiene sus ideales y propuestas, y conviven con el resto de personas que defienden otros ideales).

Bibliografia:

Introducción:
https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/el-multiverso-cuntico-711/tica-en-la-inteligencia-artificial-15492

Ejemplo1:
https://www.elblogsalmon.com/mercados-financieros/inteligencia-artificial-capaz-predecir-movimientos-bolsa-79-fiabilidad

Ejemplo2:
http://chocolatesexyconsulting.es

Ejemplo3:
https://www.antena3.com/noticias/economia/mercadona-pagara-multa-25-millones-euros-instalar-sistema-que-detectaba-personas-orden-alejamiento_2021072360fa929a4aebd80001bfc6dd.html

Ejemplo4:
https://www.tendencias.kpmg.es/claves-decada-2020-2030/robots-etica-inteligencia-artificial/

XAI:
https://www.avances.ai/inteligencia-artificial-explicable-caja-negra/

Principios / Propuestas IA / Gobierno:
https://elpais.com/tecnologia/2020-12-02/el-gobierno-invertira-600-millones-de-euros-en-inteligencia-artificial-hasta-2025.html

Areas de implementación de IA del Gobierno:
https://www.telcel.com/empresas/tendencias/notas/inteligencia-artificial-en-el-gobierno

Principios propuestos en todo el mundo:
https://www.businessinsider.es/espana-adopta-principios-inteligencia-artificial-ocde-425775