



INTELIGENCIA ARTIFICIAL

**EL FUTURO DE LA IA: CUESTIONES ÉTICAS
Y DIVERSIDAD**



Structuralia

Este documento es de uso único e intransferible para el alumno matriculado en el curso. Cualquier reproducción física o digital del documento sin permiso de los autores vulnera los derechos de propiedad intelectual de los mismos.

INDICE

INDICE.....	2
1. INTRODUCCIÓN.....	4
2. CUESTIONES ÉTICAS EN LA INTELIGENCIA ARTIFICIAL.....	4
2.1 Ejemplos relacionados con cuestiones éticas	6
2.2 Conclusiones.....	8
3. EL SIGNIFICADO DE DIVERSIDAD.....	9
4. EL FUTURO DE LA INTELIGENCIA ARTIFICIAL	10
5. REFERENCIAS	11

1. INTRODUCCIÓN

Cuando se trabaja con Inteligencia Artificial, ya sea en investigación o en empresas, debemos ser conscientes del alcance de nuestro sistema y del impacto que pueda tener en el potencial cliente o usuario final. Con el avance de la Inteligencia Artificial y la solidez que han ido ganando los sistemas inteligentes, nos encontramos con multitud de prototipos que se quieren implantar en el mundo real. Esto tiene una serie de consecuencias que debemos conocer para cumplir con una serie de medidas éticas, de inclusión y de diversidad.

En este tema se realizará una reflexión en cuanto a valores éticos y de diversidad que deberían tener todos los sistemas inteligentes, así como tenerlos presentes todos los ingenieros e investigadores encargados de diseñar estos sistemas y avanzar en el campo de la Inteligencia Artificial.

2. CUESTIONES ÉTICAS EN LA INTELIGENCIA ARTIFICIAL

Los sistemas inteligentes ya están implantados en nuestro día a día. Desde el robot roomba que limpia el suelo, el asistente (Alexa, Siri, Google) que nos ayuda y responde nuestras preguntas, hasta las recomendaciones y publicidad que recibimos diariamente en función de nuestros « me gusta » en las redes sociales. La particularidad de esos sistemas es que el impacto en el mundo real es limitado, pues afectan expresamente a la privacidad del usuario.

Sin embargo, cuando por ejemplo aplicamos inteligencia artificial para programar las noticias de un periódico, diseñar un vehículo autónomo, identificar a los mejores candidatos de una oferta de trabajo o monitorizar el entorno para actuar en consecuencia de peligro, hay que tener varios aspectos éticos en cuenta para que nuestros sistemas sean respetuosos con el usuario final y su actuación sea lo más objetiva posible.

A continuación, se enuncian una serie de pasos y sugerencias que habría que llevar a cabo para evaluar el componente ético de nuestro sistema :

- Los sistemas inteligentes han de someterse a una fase de entrenamiento en la que se estudiarán los distintos casos a los que podría enfrentarse dicho sistema. Esto incluye estudiar tanto casos típicos como atípicos. Habrá que obtener la mayor cantidad de datos estructurados posible para que nuestro sistema pueda basar sus decisiones en abanico amplio de posibilidades.

- Los sistemas inteligentes que llegan al mundo real siempre se someten previamente a diferentes pruebas en entornos controlados y simulados. Esto se realiza después de la fase de entrenamiento. El objetivo es predecir ante qué situaciones se podría encontrar el sistema inteligente en el mundo real y actuar en consecuencia. El sistema debería comportarse de forma educada y respetuosa con el usuario siempre que exista algún tipo de interacción. Igualmente, la presencia de los sistemas inteligentes se debería realizar de forma no invasiva. Si se detectan situaciones nuevas, habría que incorporarlas al proceso de entrenamiento.
- A pesar de que el entorno simulado puede parecerse mucho al que el sistema se encontrará en el mundo real, no será suficiente realizar exclusivamente esta batería de pruebas. El sistema inteligente ha de llegar a un entorno real 100% y comenzar a recoger datos del propio despliegue del sistema y de sus interacciones con el mundo real. En esta fase, aunque se realicen las pruebas en el mundo real, se dispondrá de un conjunto de ingenieros que analizarán cada uno de los movimientos y se asegurarán de que el sistema no causa ningún daño material ni humano.
- El siguiente paso será convivir 24/7 en el mundo real con el fin de detectar nuevos casos de uso y errores no identificados en las simulaciones previas. Habíamos hablado de interactuar en el mundo real de forma controlada. Una vez superada esa fase, es esencial que el sistema se despliegue durante una semana completa para observar y estudiar su evolución: interacción, aprendizaje, casos atípicos, etc.
- Habrá que tener en cuenta que las propias acciones de los sistemas inteligentes pueden dar lugar a situaciones que afectan a marcos legales. Hay que estudiar los sistemas inteligentes desde el punto de vista legal, el cuál dependerá del país en el que se quiera desplegar. Generalmente, la Unión Europea tiene una normativa mucho más estricta que la norteamericana para permitir que este tipo de pruebas se realicen en sus territorios.
- Accidentes en conducción, privacidad del usuario, manipulación, identificación errónea de sujetos, generar información no válida o falsa etc., son algunos de los problemas que podemos detectar de forma temprana en los sistemas inteligentes. Abordaremos algunos de estos casos en la siguiente sección.

2.1 Ejemplos relacionados con cuestiones éticas

En esta sección se presentan diferentes sistemas inteligentes que se ven afectados por diversas cuestiones éticas debido al uso final que tiene la plataforma.

- **Sistema de conducción autónomo.** Cuando hablamos de coches autónomos, una de las primeras preguntas que nos surgen está relacionada con su fiabilidad. ¿Es seguro el coche autónomo? La respuesta es sí, ya que ha sido entrenado previamente con multitud de formas de conducción y ha generado la suya propia basada en todo ese histórico de muestras. Para que un sistema de conducción autónomo llegue a nuestras manos como consumidores ha tenido que pasar una serie de procesos legislativos y de certificaciones de calidad y seguridad vial. Como ingenieros, técnicos o investigadores, la parte que más nos debería preocupar desde el punto de vista ético es esta última. Debido al sistema de visión en tiempo real que posee el coche autónomo, es necesario que se haya entrenado en diferentes entornos (urbano, rural, autovía etc.), con diversos climas y orografías. Lo más complejo siempre es conducir en urbano, pues la interacción con humanos es mayor y las calles no siempre son amplias. Habrá que prestar atención a todos los tipos de seres vivos que podemos encontrarnos: adultos, niños, animales etc. Es posible que no siempre se nos ocurra incluir a todos en nuestra batería de pruebas. Lo mismo sucede con los comandos de voz en el coche, hay que preparar el sistema para reconocer a cualquier usuario. Por supuesto el sistema autónomo deberá contar con un botón fácilmente identificable que detenga el modo autónomo al pulsarlo. Para resolver los dilemas éticos en cuanto a accidentes de tráfico, especialmente aquellos que sean inevitables, el Massachusetts Institute of Technology (MIT) inició este año una encuesta global [1] con gente de todo el mundo para conocer las preferencias de la población y así poder diseñar un modo de actuación u otro ante dichas situaciones de peligro inminente. En [2] puedes ver la encuesta y participar en la misma dando tu opinión.
- **Robot ayudando a personas.** En el terreno de los robots sociales es habitual encontrarse sistemas basados en robots que ayudan a niños, a personas con problemas de movilidad o neurodiversidad o a personas mayores. Cuando se diseña un robot de estas características es importante que el material sea agradable al tacto del humano (generalmente se utiliza plástico) y que no sea muy pesado, para poder interactuar con facilidad. Por otro lado, es muy importante que haya una interacción humano-máquina por medio del habla o bien a través de una pantalla táctil. Si el robot se utiliza para ayudar a personas mayores o a niños es más sencillo implementar un sistema que base la

comunicación en una pantalla táctil, pues la interacción por voz puede dar lugar a situaciones difíciles de predecir o controlar; además de que a estos usuarios se les entiende con mayor dificultad. En lo referente a dilemas éticos, habrá que analizar el movimiento del propio sistema en el entorno (no se le puede hacer daño al usuario) y el sistema inteligente deberá controlar muy bien los imprevistos, como puede ser que el niño se mueva o que el anciano no reaccione en un tiempo corto esperado. También hay que prestar atención a los datos que guardamos de los usuarios cuando se comunican con el sistema (privacidad), pues según dónde se desarrolle esa interacción, por ejemplo, en espacios abiertos, pueden no habernos dado permisos para almacenarlos. Otra variante de estos robots son los robots de servicios, que intentan resolver una tarea que se les ha ordenado. Es más común que sean estos robots los que se encuentren en espacios abiertos como ferias, congresos, universidades, aeropuertos o centros comerciales. En estos casos también será imprescindible la diversidad del dataset en cuanto a distintos tipos de personas que nuestro sistema sea capaz de reconocer, no precisamente en cuanto edad sino en cuanto a raza o idioma. Habrá que tener en cuenta también las formas que tiene cada cultura para iniciar una conversación, finalizarla y el nivel de cordialidad esperado. Según la región en la que se desarrolle el sistema, la forma de relacionarse será diferente.

- **Sistema recomendador de noticias.** Los sistemas recomendadores fueron una de las primeras aplicaciones de la Inteligencia Artificial orientada al marketing y al negocio. El origen surge de Amazon y su habilidad para desarrollar ese primer sistema que comenzaba a analizar las relaciones entre la compra de unos productos y otros así como el momento y lugar en el que se realizaba esa compra. Esto aplicado al periodismo se ha utilizado en los últimos años para enviar al email noticias de interés, newsletters con contenido personalizado, o incluso para personalizar la portada de un periódico en su versión online. Aparentemente, en cuanto a cuestiones éticas sólo aparecería de nuevo la privacidad del usuario referido a sus gustos personales, ideología y característica similares. Sin embargo, en estos dos últimos años han aflorado los casos de ‘fake-news’ o noticias generadas con contenido falso que inundan las redes, pues su objetivo es viralizarse en tiempo record para dejar una huella en la mente del lector. A pesar de que a las horas se eliminan o se lance una rectificación, el alcance que tiene el contenido falso es tan directo y sencillo de entender que el usuario no se preocupará por analizar su veracidad. Por este motivo, hay que evaluar el nivel de manipulación de este tipo de

sistemas, tanto los inteligentes como los que no lo son. Se puede realizar manipulación ideológica tan solo con la personalización de la portada del periódico, pues ya se está decidiendo qué contenidos se encuentran a un *click* y a cuáles costará más acceder. Recientemente, distintos países han experimentado estos fenómenos en época de elecciones, como Estados Unidos [3], Reino Unido y México [4]. Ahora más que nunca cobra muchísimo poder el disponer de datos, mediciones, patrones y del acceso directo a grupos indecisos de población, por ejemplo, a través de las redes sociales.

- Chatbots para resolver incidencias. La interacción con humanos a través del lenguaje es uno de los grandes retos de la Inteligencia Artificial. En los últimos años se han dado grandes pasos, como hemos visto en temas anteriores, pero a pesar de ello, seguimos cometiendo errores a la hora de entrenar los sistemas inteligentes. El origen del problema reside en la dificultad de encontrar un corpus amplio que sea representativo y que le enseñe al bot cómo se debe relacionar, así como el grado de cordialidad. Muchos ingenieros e investigadores optaron por utilizar las redes sociales como fuente de generación de ejemplos. Es de índole popular que en las redes sociales la cordialidad es nula y el respeto a veces escasea. Debido a esto, en 2016 un experimento de Microsoft tuvo que ser retirado a las pocas horas de su lanzamiento [5]. El sistema en cuestión aprendía a relacionarse con los humanos a través de tweets de 140 caracteres, tal y como se hace en la red social Twitter. Como se retroalimentaba de los tweets que la gente publicaba, en pocas horas se convirtió en un ser racista, sin escrúpulos y con valores éticos cuestionables. Por este motivo, durante la fase de aprendizaje de los sistemas inteligentes hay que saber evaluar la calidad de la información así como filtrar el posible ruido y adecuar los ejemplos al contexto en el que queremos que el bot se desenvuelva.

2.2 Conclusiones

Una vez analizados estos cuatro casos de ejemplo, enumeramos una serie de conclusiones para que os sirvan de guía a la hora de analizar vuestros sistemas:

- Debemos conocer cómo funcionan los algoritmos por dentro. Qué elementos o datos concretos los llevan a tomar las decisiones.
- Los seres humanos tenemos sesgos, somos imperfectos [6]. Los sesgos se propagan de forma consciente o inconsciente en los datasets con los que se entrenan los sistemas inteligentes. Hay que preguntarse si mi sistema, a pesar de funcionar correctamente, contiene algún tipo de sesgo.

- Cuando elaboramos los datasets para proporcionar conocimiento a nuestro sistema inteligente debemos preguntarnos si hemos tenido en cuenta ejemplos de diferentes clases, no sólo de nuestro entorno cercano.
- Ideologías, expresiones, información técnica validada, luz natural/artificial, casos atípicos etc. todos estos aspectos deben ser evaluados en mayor o menor medida dependiendo del caso de uso del sistema inteligente y del usuario final y su geolocalización.

3. EL SIGNIFICADO DE DIVERSIDAD

Como se ha planteado en la anterior sección, los datasets no son perfectos. Tenemos que controlar los sesgos y la diversidad e inclusión del propio sistema inteligente. En primer lugar definiremos ambos términos:

- La diversidad se refiere a todo aquello que marca una diferencia dentro de un grupo, bien sean características físicas, organizacionales, psicológicas o conductuales.
- La inclusión se refiere a la consideración de todos los colectivos a la hora de realizar cualquier acción o llevar acabo cualquier proyecto.



Ilustración 1: Ejemplo gráfico de diversidad en personas

A continuación, se muestra una lista de elementos a tener en cuenta para fomentar mayor diversidad e inclusión tanto dentro del sistema inteligente como previo a su desarrollo.

- Desde el momento en el que se empieza a diseñar el sistema inteligente hasta que se desarrolla, se evalúa y se despliega en el mundo real, se debe analizar la propia diversidad del sistema y el equipo de desarrollo.
- Detección de personas de diferente raza o de diferente altura en tareas de visión artificial. Tener en cuenta tanto luz artificial como luz natural.
- Reconocimiento del habla entrenado en hombres, niños o mujeres, pues utilizan diferentes frecuencias. Cuanto más mayor sea la persona más pausada será su forma de hablar, habrá que tenerlo en cuenta para saber cuándo ha terminado la frase el usuario.
- Adaptabilidad para diestros y zurdos, especialmente en dispositivos móviles donde se utilizan sensores como el acelerómetro y giroscopio para saber si coger la llamada, colgarla, girar la pantalla, etc.
- Diseño de accesorios unisex o para hombres y mujeres, teniendo en cuenta las dimensiones de cada uno. Se ha dado el caso con los smartwatches, pues la mayoría son muy grandes e incómodos para la muñeca de una mujer [6].
- Accesibilidad para personas mayores y discapacitados. Debemos pensar en cómo incluir a la población más anciana en la tecnología, es decir, hacerla accesible. Cuando seamos mayores nos gustará que alguien piense en nosotros a la hora de diseñar los sistemas inteligentes.
- Se debe hacer un esfuerzo por reducir la brecha tecnológica. Hasta el momento la tecnología es un sector altamente masculinizado con mayoría de raza blanca. Esto afecta a la forma en la que se desarrollan los proyectos, la alcanzabilidad y el impacto socioeconómico.

4. EL FUTURO DE LA INTELIGENCIA ARTIFICIAL

Finalmente, para dar por terminado este curso, se enuncian a continuación las líneas presentes y futuras en las que se está trabajando en la Inteligencia Artificial. En resumen, todo gira alrededor de la transparencia de los propios sistemas inteligentes.

- Marco regulatorio legal: se empieza a legislar cada vez con más ímpetu sobre las distintas aplicaciones que se generan con Inteligencia Artificial. Su implantación en el mundo real

depende de dicho marco legal. También se establecen directrices para la evaluación de dichos sistemas.

- Sistemas con alto componente transversal: Existen multitud de campos donde anteriormente no se había aplicado la Inteligencia Artificial. Ahora sectores como la lingüística, la medicina, el periodismo, el derecho, el arte etc. están trabajando mano a mano con técnicos especialistas en este campo.
- Explainable AI es el nombre que cobra la nueva línea de investigación que apuesta porque los sistemas inteligentes nos cuenten cómo han tomado sus decisiones de forma explícita.
- Fairness es el nombre que cobra la nueva línea de investigación que apuesta por evaluar la diversidad y el componente ético de los sistemas. Que el sistema inteligente sea “justo”.

5. REFERENCIAS

- [1] MIT Medialab (2018) <http://moralmachine.mit.edu/hl/es>
- [2] MIT Medialab. How should autonomous vehicles be programmed? (2018) <http://news.mit.edu/2018/how-autonomous-vehicles-programmed-1024>
- [3] A new study suggests fake news might have won Donald Trump the 2016 election (2018) https://www.washingtonpost.com/news/the-fix/wp/2018/04/03/a-new-study-suggests-fake-news-might-have-won-donald-trump-the-2016-election/?utm_term=.881a8ecc33a8
- [4] La inteligencia artificial definirá las próximas elecciones presidenciales (2018) https://www.huffingtonpost.com.mx/2017/11/18/la-inteligencia-artificial-definira-las-proximas-elecciones-presidenciales_a_23281732/
- [5] Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. (2016) <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>
- [5] Cathy O'Neil – La era de la fe ciega en los algoritmos ha de terminar. TED talks (2017). https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end?language=es
- [6] Why most all smartwatches suck for small wrists (2017) <https://www.imore.com/apple-gets-it-right-why-android-wear-sucks-for-women>