

RETO 50: Inteligencia Artificial Ética

- Iñaki Retegi-

¿Qué es ético y qué no?

La ética tiene su origen en un término creado por Aristóteles. Él definió como:

“Disciplina filosófica que estudia el bien y el mal y sus relaciones con la moral y el comportamiento humano.”

Hoy en día sin alejarse demasiado su sentido original, lo definimos como: *“Conjunto de costumbres y normas que dirigen o valoran el comportamiento humano en una comunidad.”*

Si lo enfocamos a cualquier actividad o campo a explorar y a explotar como puede ser en el caso de la IA, podemos entender que debe ser compatible con las definiciones que anteriormente he mencionado. Al tratar estos dos conceptos en conjunto: “Ética” + “AI” veo que **tenemos un problema por no conocer sus dimensiones**. La ética no es un concepto matemático incuestionable, es una idea que dadas situaciones concretas es probable que no estemos de acuerdo con cual puede ser la opción más ética. Un ejemplo sin mencionar la IA puede ser la situación de “¿Si sólo podemos salvar a una persona y hay dos en riesgo, debemos salvar un bebé o un anciano?”

En los enlaces facilitados en este reto, se menciona un supuesto de un vehículo autónomo por AI, que en una situación en que tiene un frente un camión a punto de perder una carga, el vehículo inteligente debería girar al carril izquierdo colisionando con un vehículo, o si debería girar al carril derecho colisionando con un motorista, o si debería dejar caer la carga del camión encima para no perjudicar a ningún otro usuario. ¿Cual opción es la “ética”? Seguramente en la vida real por instinto de supervivencia giraríamos contra el motorista (no estoy seguro), y no sería muy ético. ¿O sí?

Un ejemplo que puede dar a pensar es un falso spot de un anuncio de un vehículo inteligente que por “ética” y sin dudarlo atropella a un niño. Dicho así poco tiene de ética, pero animo a ver el spot: <https://youtu.be/BCXBdKPgHxI> ¿Ya no parece tan poco ético? ¿O sí? Real desde luego no es, pero ante situaciones concretas extremas se podrían dar casos difícilmente comprensibles por la mayoría.

Se pueden reunir una serie de medidas a cumplir y legislar, para que para garantizar en la medida de lo posible el interés de las personas, con las sanciones correspondientes en el caso del no cumplimiento:

- Respetar la ética establecida
- Control de la AI por el ser humano
- Seguimiento y trazabilidad de las decisiones AI para entender, corregir y aprender
- Respetar la intimidad.
- Identificar el AI.
- Etc.

¿Son los datos justos? ¿Tienen sesgos? ¿Discriminan?

Los datos son justos, si los datos introducidos son justos. Entiendo que depende del ser humano que inserta los datos. Para ello (sin comprender a día de hoy su desarrollo) entiendo que hay que hacer un proyecto muy completo de datos, lo más completo posible, para el AI para que pueda ser aplicado lo más real y justo, y evitar lo máximo posible el riesgo de discriminación. Al redactar esta respuesta me ha recordado el video facilitado que muestra radiografías de casos médicos y es capaz de identificar tumores, y otras enfermedades. Si el sistema AI no tiene una “colección” completa de información para comparar, no podrá llegar a una conclusión real. Supongo que en parte se podría comparar con la educación: “Si no educamos con inclusión, diversidad y respeto, difícilmente será capaz un ser (antes de contrastar otros criterios) de no discriminar, y de ser justo”.

¿Cómo hacemos para que los resultados de la IA sean explicables y dejen de ser cajas negras?

Habría que hacer un sistema de trazabilidad de cada una de las decisiones. Alguna manera para que se pueda comprender por el ser humanos las decisiones de AI que no comprendemos, o los errores que puedan surgir. Un almacenamiento de: contexto + consulta de datos + decisión. (O algo por el estilo).

¿Cómo analizamos los parámetros/datos más influyentes? ¿Y su sensibilidad?

Los parámetros/datos más influyentes son los que más inquietan e influyen a las personas. A las personas las conocemos y comprendemos (en gran parte) y esos datos se pueden recopilar. Esos datos ya son tratados y medidos en otros campos como web, servicios públicos, prensa, etc.

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

No debería vulnerar la ética que se haya consensuado como tal en la Unión Europea. Y entre otras funcionalidades, deben garantizar una seguridad: Capacidad de detener un sistema AI cuando el ser humano lo decida o cuando no cumpla con los requisitos exigidos por la UO. Sea diseñado para que pueda ser comprensible su actividad, que no haya ninguna situación en el que el ser humano no pueda tener la actividad AI bajo su supervisión y control.

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

Creo que esta pregunta más que un dato puede ser una opinión, así que me animaré a dar mi opinión. Creo que la homologación limita la creatividad, ya que cuando una empresa trata de sacar un producto o servicio que pretende ser homologado, tratará de cumplir con estándares a veces sin alejarse, arriesgarse o innovar demasiado. Muchas veces parece que la homologación es un campo muy cómodo para grandes empresas y no tan cómodo para las pymes.

Creo que lo más correcto para dar más libertad a la creación sería: Sí homologar los productos y servicios que puedan poner en riesgo la vida de las personas, y para el resto de los casos marcar unas pautas éticas a respetar, y una legislación a cumplir, y a partir de ahí sin necesidad de homologar haya libertad para crear.

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

La AI es y será una tecnología creada por y para las personas, para la sociedad, por lo que la sociedad debe marcar cómo quiere que sea esa AI. Los valores y la necesidad lo marca la sociedad, y los saben identificar los profesionales como sociólogos, filósofos, psicólogos, ingenieros, etc. Como representantes de la sociedad ya están los expertos **“High-level expert group on artificial intelligence”** a nivel europeo (y no los gobiernos de cada país), que con mayor conocimiento y experiencia marcarán cómo debe ser la AI que tenemos cada vez más presente.