[06] ¿CUÁLES SON LOS PRINCIPALES PROBLEMAS ÉTICOS RELACIONADOS CON LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL?

A lo largo de la ultima década se ha ido construyendo una falsa narrativa sobre las bondades de la inteligencia artificial que tiene tendencia a ignorar todos aquellos aspectos que los expertos y académicos están expresando. De hecho, los sistemas de IA presentan comportamientos no explicables, con una probabilidad garantizada de error que es significativa y no pequeña. Esto los hace esencialmente discutibles en determinadas situaciones críticas en las que los errores podrán poner en riesgo vidas humanas (diagnóstico médico, vehículos autónomos o el caso de las armas autónomas, en las que los errores se convertirán directamente en vidas humanas) y en las que puede ser difícil la rendición de cuentas.[1] Los principales problemas éticos relacionados con estos sistemas derivan de su **carácter oscuro y no explicable,**[2] de su **fiabilidad limitada**[3] **y probabilidad de error,** del hecho de **no poder resolver las ambigüedades** que aparecen en las situaciones reales, de sus **inevitables sesgos,** y de la **necesidad de post-supervisión** en el caso de aplicaciones críticas. Los sistemas de IA aprenden, captan y actúan. Pero **pueden ser imprevisibles** y no siempre lo hacen de la manera esperada.

Uno de los campos de aplicación que se nos presentan como más prometedores en el campo de los sistemas de IA es el de los sistemas autónomos. La **diferencia constructiva** entre un sistema automático y uno autónomo es que en el primer caso (puertas automáticas, lavadoras y lavaplatos), su comportamiento es previsible, mientras que los sistemas autónomos, que han sido construidos para poder actuar en base a "sus" decisiones (basadas en su aprendizaje y en la información que captan), pueden ser imprevisibles, además de no explicables y en casos, erróneos.

Pero los problemas éticos relacionados con los sistemas de Inteligencia Artificial provienen no tanto de sus potencialidades constructivas sino de **su uso, que es el que puede ser ético o no.** En este sentido podemos distinguir tres casos:

- Los sistemas basados en IA y autónomos desde un punto de vista constructivo que se usan bajo control humano. Sería el caso de los vehículos autónomos que deberían ser constantemente controlados por la persona que los conduce. En este caso, los problemas éticos asociados a los sesgos, baja fiabilidad y no explicabilidad, quedan resueltos por la existencia de una persona responsable que puede y debe asumir una posible rendición de cuentas.
- Los sistemas basados en IA que se usan con post-supervisión. Sería el caso de los sistemas de ayuda al diagnóstico médico en los que los expertos acaban tomando la decisión final. Como el caso anterior, su uso suscita pocos problemas éticos, aunque es importante recordar el sesgo de automatización,[4] que habría que tener siempre en cuenta.

- Los sistemas basados en IA que funcionan de manera autónoma, sin control ni intervención humana. En este caso, los problemas éticos asociados a los sesgos así como a la baja fiabilidad, comportamiento imprevisible y no explicabilidad, son muy relevantes.

En este contexto, algunos autores como Alan Winfield y Marina Jirotka proponen que los robots y los sistemas autónomos deban ir equipados con una "caja negra ética" que seria el equivalente de los grabadores de datos de vuelo de los aviones, y que registraría continuamente los datos y el estado interno de los sistemas de inteligencia artificial en aplicaciones críticas.[5] Esta caja negra ética seria esencial para poder entender lo que ha ocurrido en caso de víctimas y facilitaría el establecimiento de responsabilidades.

En todo caso, el uso de sistemas de IA en equipos de armamento autónomos (las denominadas "LAWS" en inglés) nos lleva al máximo grado de problemas éticos. La escalada hacia los sistemas armados autónomos es ética y jurídicamente inaceptable, porque delegar en una máquina las decisiones de matar va en contra de la dignidad humana y de los derechos de las personas. Hay que situar el concepto de dignidad humana como límite insalvable,[6] marcando una "línea roja" más allá de la cual la autonomía en los sistemas de armas ya no puede ser aceptable.[7] Los drones letales autónomos no podrán tomar decisiones éticas complejas en campos de batalla dinámicos, ni podrán distinguir adecuadamente entre soldados y civiles, y tampoco podrán evaluar el grado de proporcionalidad de un ataque. Sin hablar de su comportamiento imprevisible, la posible pérdida de control, los accidentes "habituales" y los potenciales malos usos.

Notas:

1. Joaquín Rodríguez, Xavi Mojal, Tica Font y Pere Brunet (2019), "Nuevas armas contra la ética y las personas. Drones armados y drones autónomos", Informe 39, Centro Delàs de Estudios para la Paz: http://centredelas.org/wp-content/uploads/2019/11/informe39_DronesArmados_RE_CAST_web_DEF-1.pdf
2. Véase la respuesta a la pregunta 5.
3. Véase la respuesta a la pregunta 3.
4. El sesgo de automatización es la tendencia humana a dar por bueno aquello que nos proponen las máquinas. Noel Sharkey explica que hay que tener en cuenta este sesgo, que rebaja fuertemente el posible apoyo ético a estos sistemas. Este sesgo hace, según Sharkey, que "los operadores estén predispuestos a aceptar las recomendaciones informáticas sin buscar otras informaciones que permitan su confirmación. La presión temporal añadida hace que los operadores caigan en todas las trampas del razonamiento automático: en lugar de pensar, pasan a creer y aceptar aquello que la máquina los propone; ignoran la ambigüedad, suprimen la duda, inventan causas e intenciones, se centran en las pruebas existentes e ignoran las pruebas ausentes que ellos tendrían que buscar". Véase: Sharkey, Noel (2014): "Towards a principle for the human supervisory control of robot weapons". UNOG: https://www.unog.ch/80256EDD006B8954/(httpAssets)/2002471923EBF52AC1257CCC0047C791/$file/Article_Sharkey_PrincipleforHumanSupervisory.pdf
5. Alan Winfield and Marina Jirotka (2017): "The Case for an Ethical Black Box", Proc. of the Annual Conference Towards Autonomous Robotic Systems TAROS 2017, pp. 262-273: https://link.springer.com/chapter/10.1007/978-3-319-64107-2_21
6. Palmerini E., Azzarri F., et al.. (2016) "Robolaw: Guidelines on Regulating Robotics": https://www.researchgate.net/publication/322041670_Guidelines_on_Regulating_Robotics
7. Daan Kayser and Alice Beck (2018): "Crunch Time". PAX Report, Noviembre 2018: https://www.paxvoorvrede.nl/media/files/pax-rapport-crunch-time.pdf