"Garantías frente al sesgo y discriminación algorítimicas"

Prof. Dr. Lorenzo Cotino Hueso

Entre los riesgos e impactos de la IA destacan los errores, sesgos y discriminación masivos. Por lo general suelen darse por malas elecciones de datos, mal recopilados o de mala calidad. Las mismas series históricas de datos que alimentan al sistema de IA pueden ser reflejo de una realidad social discriminatoria. Lo peor, además, es que se pueden generar espirales de sesgo, error y discriminación pues los sistemas muy posiblemente acentuarán sus decisiones al nutrirse de nuevos datos cada vez más negativos para los sectores perjudicados. Las discriminaciones también pueden ser causa de un erróneo diseño y elección de algoritmos, del peso que atribuyen a unos u otros factores o a los errores en el desarrollo de los sistemas de aprendizaje automático. La detección de las causas de sesgos y discriminaciones puede ser realmente compleja y pasan por las garantías de transparencia y caja blanca y debido proceso. la posibilidad de corrección de datos y de recurrir decisiones algorítmicas.

Más allá de errores, el sistema de IA y algoritmos puede estar diseñado para tener en cuenta circunstancias especialmente prohibidas (sexo, raza, religión, salud, etc.). En muchos casos, supondrá un tratamiento de datos especialmente protegidos (art. 9 RGPD-UE), con particulares garantías en las decisiones sólo automatizadas significativas (art. 22. 3º RGPD-UE). Además, los tratamientos diferentes basados en circunstancias especialmente prohibidas son especialmente sospechosos de discriminación. Generalmente habrá partir de la presunción de nulidad de estos tratamientos. Y aún contarán con más garantías Y prohibiciones respecto de tratamientos automatizados en materia penal y de justicia (artículo 11. 3º de la Directiva (UE) 2016/680)

Hay que tener en cuenta el fenómeno de los datos llamados *proxies* o indirectos, esto es, datos que en principio no son datos sensibles; no obstante de los mismos pueden derivar factores especialmente prohibidos o datos especialmente protegidos. Así sucede por ejemplo a partir de datos como los gustos, el tipo de compras, barrios dónde se mueve, etc. habrá que analizar posibles "enmascaramientos" y la elección intencional de factores que están cerca de los prohibidos o de datos afines a los especialmente protegidos. Y también los propios algoritmos se pueden programar para ignorar o minimizar la importancia que los factores prohibidos en sus decisiones.

La resolución del Parlamento UE sobre macrodatos "insta" a "minimizar la discriminación y el sesgo algorítmicos" (nº 20) y afirma también la necesaria "mitigación algorítmica" (nº 21, ver también 32). Ahora bien las medidas de corrección de posibles sesgos han de estar bien justificadas en su necesidad, así como en su razonabilidad y proporcionalidad. Y lo que es peor. la reducción de la discriminación puede dar lugar a resultados absolutamente ineficaces e incluso más discriminatorios.

Frente a la discriminación algorítmica hay que extender el modelo de la responsabilidad proactiva en protección de datos, la no discriminación en el diseño y por defecto, así como medidas concretas en los estudios de impacto. Te conté europeo de protección de datos ha detallado importantes buenas prácticas contra la discriminación. Asimismo no pocas de las 150 cuestiones del *checklist* en las Directrices para la ética en el diseño en la UE 2019 lo son para lograr la exactitud y fiabilidad, integridad, calidad de los datos y para evitar el sesgo y la discriminación.