ARI 52/2021
6 de mayo de 2021

Hacia un régimen europeo de control de la Inteligencia Artificial

Andrés Ortega | Investigador sénior asociado del Real Instituto Elcano | @andresortegak

Tema

Las propuestas de la Comisión Europea para regular la Inteligencia Artificial, prohibiendo algunas aplicaciones y limitando otras, son ambiciosas y aspiran a tener un impacto global. Tardarán en materializarse, con una dudosa efectividad.

Resumen

La Comisión Europea ha planteado una propuesta ambiciosa de regulación de la Inteligencia Artificial (IA) en la UE, prohibiendo algunas aplicaciones que pueden violar los derechos europeos, limitando otras de alto riesgo, con un sistema de supervisión nacional y comunitario, que incluye la posibilidad de multas cuantiosas. Puede tener un impacto global, pero también frenar las posibilidades de desarrollo propio de la IA en la UE. A la vez, ha presentado un Plan Coordinado sobre IA que viene a actualizar, post-pandemia, la estrategia al respecto que había presentado anteriormente, especialmente en el terreno de las inversiones, para las que propugna 20.000 millones de euros al año.

Análisis

Introducción

La Comisión Europea ha propuesto una ambiciosa legislación a nivel de la UE para defender los valores europeos en una tecnología, o grupo de tecnologías, la Inteligencia Artificial (IA), que está viviendo una explosión y se desarrolla muy rápidamente. Se trata de conseguir que la gente tenga confianza en esta tecnología que se expande, y en la que la UE se quiere convertir en un líder global. Es una regulación enfocada sobre todo al control de los riesgos que plantea la IA, para la seguridad de consumidores y usuarios y del respeto a los derechos fundamentales, y poder sacar más provecho con confianza de las ventajas que aporta. La búsqueda de la confianza en el usuario es un objetivo central. La Comisión plantea algunos riesgos inadmisibles. La IA que contradiga los valores de la UE será prohibida. Es una propuesta dirigida principalmente a proveedores y operadores, en el marco de la UE, pero con aspiraciones globales. Desde luego, va más allá de los principios éticos para la IA acordados en el marco de la OCDE o del G20. Y marca que la autorregulación de toda una industria está llegando a su fin.

En el mundo hay a grandes rasgos cinco regímenes de gobernanza de la digitalización, que incluye la IA, y se diferencian por quién y en qué grado se ejerce el control sobre el mercado:

(1) El estadounidense, o “capitalismo de vigilancia”, en el que un pequeño número de proveedores de servicios digitales disfruta de un poder de mercado preponderante y de ventajas informativas digitales, especialmente de control, sobre los datos de los usuarios y sobre la publicidad. Ello, como bien puso de manifiesto Edward Snowden con sus revelaciones, no excluye la intromisión de los servicios de inteligencia.

(2) El chino, o “Estado de vigilancia”, en el que el Estado goza de ventajas informativas preponderantes, apoyado por un pequeño número de proveedores de servicios digitales. En él, el Estado quiere centralizar los datos, porque son un instrumento de control político.

(3) El europeo, que busca una regulación centrada en las personas, o “humanista”. Los proveedores de servicios digitales se ven constreñidos por una regulación de la UE que se centra en la protección de los valores y derechos de los usuarios digitales, en primer lugar, la privacidad. El nuevo paso de la Comisión se sitúa en este marco. Es la propuesta de regulación de la IA más estricta que se haya puesto sobre la mesa.

(4) El de un conjunto de democracias –como el Reino Unido, Australia y Japón–, también preocupadas por el poder de las grandes plataformas estadounidenses y chinas, pero que carecen del poder de negociación de la UE.

(5) El del resto del mundo, compuesto principalmente por economías emergentes, con poco poder de mercado y poco poder regulador, que están entre unas (como África) y otras (como Rusia).

Tras otras anteriores, pre-pandemia, la Comisión Europea presentó a finales de 2020 dos iniciativas de gran calado: las propuestas de Reglamento de Servicios Digitales, esencialmente para responsabilizar a las plataformas sobre contenidos (con la amenaza de cuantiosas multas), y el Reglamento de Mercados Digitales para limitar las actividades de algunas de estas empresas, sobre todo contra los “guardianes”, los *gatekeepers*, pues otras compañías han de usar sus servicios para sus propios negocios, y son capaces de dictar cómo han de funcionar los mercados.

Ya antes había presentado unas directrices para la IA, según las cuales una IA confiable debe ser legal (respetar todas las leyes y regulaciones aplicables), ética (respetar los principios y valores éticos), y robusta (tanto desde una perspectiva técnica como teniendo en cuenta su entorno social). Las directrices presentaban un conjunto de seis requisitos clave que los sistemas de IA deben cumplir para ser considerados confiables: (1) agencia y supervisión humana; (2) robustez técnica y seguridad; (3) privacidad y gobernanza de datos; (4) diversidad, no discriminación y equidad; (5) bienestar social y ambiental; y (6) responsabilidad (la IA y sus resultados deben rendir cuentas ante auditores externos e internos).

En octubre de 2020 la Presidencia alemana del Consejo de la UE publicó unas conclusiones sobre la Carta de Derechos Fundamentales en el contexto de la IA y el cambio digital. Proporcionaron orientaciones sobre la dignidad, las libertades, la

igualdad, la solidaridad, los derechos de los ciudadanos y la justicia ante los retos de la IA.

La propuesta de *Artificial Intelligence Act* (AIA), Reglamento de Inteligencia Artificial, presentada el pasado 21 de abril, concreta y avanza en algunas de estas líneas. Se presenta como unas "nuevas reglas y acciones para la excelencia y la confianza en la IA" y contiene algunas complejas medidas preventivas en el desarrollo de la IA. O para ser precisos, de las aplicaciones de la IA, no de la IA en sí. Es una propuesta muy ambiciosa con aspiraciones de universalidad.

El proceso para su elaboración se ha desarrollado a diversos niveles. La Comisión impulsó un primero Grupo de Alto Nivel de Expertos de 52 miembros, luego ampliado a "Alianza IA", que desembocó en un Libro Blanco sobre IA, que recibió numerosos comentarios *online*. La Comisión ha avanzado de un marco a una propuesta plenamente regulatoria. La propuesta de Reglamento de la Comisión será sometida al Consejo de Ministros y al Parlamento Europeo, y pueden cambiar muchas cosas en el largo y complejo procedimiento legislativo. Se calcula que en el mejor de los casos no se cerrará la nueva reglamentación hasta 2023. Para la aprobación del Reglamento General de Protección de Datos (RGPD) pasaron cuatro años entre la propuesta de la Comisión y su entrada en vigor en 2018.

Reflejo de la complejidad de que se trata es la dificultad del texto para llegar a una definición de qué es la IA lo más neutra posible, que lo cubra todo, incluido la IA con lógica simbólica, el aprendizaje máquina y los sistemas híbridos. Para ello ha tenido que recurrir a un anexo con una lista de técnicas utilizadas, consciente, además, que esa lista requerirá actualizaciones:

"(art.3): Un sistema de inteligencia artificial es software desarrollado con una o varias de las técnicas y enfoques enumerados en el anexo I y que puede, para un conjunto determinado de objetivos definidos por el ser humano, generar resultados como contenidos, predicciones, recomendaciones o decisiones que influyen en los entornos con los que interactúan."

No es, precisamente, claro.

El anexo se refiere a: (1) enfoques de aprendizaje automático, incluido el aprendizaje supervisado, no supervisado y de refuerzo, utilizando una amplia variedad de métodos, incluido el aprendizaje profundo; (2) enfoques basados en la lógica y el conocimiento, incluida la representación del conocimiento, la programación (lógica) inductiva, las bases de conocimiento, los motores de inferencia y deducción, el razonamiento (simbólico) y los sistemas expertos; y (3) enfoques estadísticos, estimación bayesiana, métodos de búsqueda y optimización.

Cabe considerar que son estas herramientas, más que elementos, objeto de leyes.

Regulación de riesgos

Esta nueva propuesta parte de que las aplicaciones de la IA aportarán muchos beneficios, pero pueden, si no se regulan, conllevar riesgos que minarán la confianza de los usuarios/ciudadanos. Y de eso se trata, de regular esos riesgos, diferenciando entre los directamente prohibidos, los de alto riesgo y los de riesgo mediano o bajo. Ahora bien, de esta legislación están exentas todas las aplicaciones militares de la IA, y deja muchas excepciones para las policiales.

Cuando la Comisión presentó el Libro Blanco sobre IA en febrero de 2020, planteó solo dos niveles de riesgo (alto y bajo), que generaron controversia. Se veían como un binomio muy limitado y que no representaba la realidad en su totalidad. Además, los "requisitos de determinación del nivel de riesgo" eran realmente flojos y abarcaban mucho. Daban un elevado margen de maniobra para etiquetar con cierta laxitud a una aplicación como de riesgo bajo. Tampoco quedaba claro *quién* determinaría dicho nivel.

Meses antes, el Consejo Alemán de Ética de Datos había publicado un informe donde planteaba que debía haber cinco niveles de riesgo. Que ahora la Comisión Europea plantee cuatro niveles es positivo porque va más allá que su propia propuesta de 2020, pero sigue siendo un enfoque limitado si se compara con el detalle, precisión y *checks and balances* que aportaba el informe del Consejo Alemán.

La tipología de riesgo planteada desde la Comisión es la siguiente:

(a) Riesgos prohibidos: la Comisión plantea que algunos riesgos son inadmisibles. La IA que contradice los valores de la UE será prohibida, lo que incluye los sistemas de IA que se consideren una clara amenaza para la seguridad, los medios de subsistencia y los derechos de las personas. Esto abarca los sistemas o las aplicaciones de IA que manipulan el comportamiento humano para eludir la voluntad de los usuarios (por ejemplo, manipulación subliminal o juguetes que utilicen asistencia vocal para incitar a comportamientos peligrosos a los menores, la explotación de niños o personas con discapacidad mental, resultando en daños físicos/psicológicos, el reconocimiento facial o biométrico en espacios públicos (con excepciones, como las medidas antiterroristas, o la policía predictiva) y los sistemas que permitan la "puntuación social" por parte de las autoridades públicas (aunque no de las privadas). En cuanto al reconocimiento facial, esta prohibición, incluso con algunas excepciones policiales para casos antiterroristas y otros, resulta poco creíble, dada la proliferación de cámaras –públicas y privadas– que va en aumento en nuestras ciudades. Hay, además, empresas privadas que se dedican al reconocimiento facial, o a la puntuación social, que no estarán prohibidas, aunque sí consideradas dentro de la siguiente categoría.

(b) De alto riesgo: está permitida, pero sujeta al cumplimiento de los requisitos de IA y la evaluación de la conformidad *ex ante*. Los sistemas de IA considerados de alto riesgo abarcan las tecnologías empleadas en: las infraestructuras críticas (que pueden poner en peligro la vida y la salud de los ciudadanos como, por ejemplo, los transportes); la formación educativa o profesional, que puede determinar el acceso a la educación y la carrera profesional de una persona (por ejemplo, la puntuación en exámenes); los componentes de seguridad de los productos (por ejemplo, la

aplicación de la IA en cirugía asistida por robots); el reclutamiento de empleados (por ejemplo, los programas informáticos de clasificación de CV para procedimientos de contratación); los servicios públicos y privados esenciales (por ejemplo, los sistemas de calificación crediticia que priven a los ciudadanos de la oportunidad de obtener un préstamo); la aplicación de las leyes, que pueden interferir con los derechos fundamentales de las personas (por ejemplo, la evaluación de la fiabilidad de las pruebas); la gestión de la migración, asilo y control de las fronteras (por ejemplo, la comprobación de la autenticidad de los documentos de viaje); y la administración de justicia y procesos democráticos (por ejemplo, la aplicación de la ley a un conjunto concreto de hechos).

En especial, se consideran de alto riesgo y estarán sujetos a requisitos estrictos todos los sistemas de identificación biométrica remota. Su uso estará sujeto a la autorización de un órgano judicial u otro organismo independiente y a los límites adecuados desde el punto de vista de la duración, el alcance geográfico y las bases de datos exploradas. En este caso, los sistemas de IA de alto riesgo estarán sujetos a obligaciones estrictas antes de que puedan comercializarse, entre los que se cuentan la necesidad de una alta calidad de los datos que alimentan el sistema para minimizar los riesgos y los resultados discriminatorios, un registro de la actividad para garantizar la trazabilidad de los resultados, una documentación detallada que aporte toda la información necesaria sobre el sistema y su finalidad para que las autoridades evalúen su conformidad y una información clara y adecuada al usuario, medidas apropiadas de supervisión humana para minimizar el riesgo. Y se exige un alto nivel de solidez, seguridad y precisión de estas aplicaciones.

(c) De riesgo limitado: se consideran sistemas de IA con obligaciones específicas de información/transparencia. Por ejemplo, en robots conversacionales los usuarios deberán ser conscientes de que están interactuando con una máquina para poder tomar una decisión informada de continuar o no.

(d) De riesgo mínimo o nulo: la inmensa mayoría de los sistemas de IA entra en esta categoría. La propuesta permite el uso de aplicaciones tales como videojuegos basados en la IA o filtros de correo basura. El proyecto de Reglamento no interviene aquí, ya que estos sistemas de IA sólo representan un riesgo mínimo o nulo para los derechos o la seguridad de los ciudadanos.

Todas estas limitaciones se aplicarían a los programas y dispositivos fabricados en la UE o importados. Esto va a plantear problemas a la hora de controlar de dónde proceden componentes muy concretos de la IA.

La regulación no se aplica a los usos privados no profesionales (como Amazon Echo en el domicilio, Siri, etc.).

Transparencia

Se exige un alto grado de entendibilidad o transparencia de lo ocurre durante el funcionamiento de los algoritmos. Este es uno de los requisitos más básicos planteado por la Comisión. La IA debe de ser transparente, lo que supone poder reconstruir cómo

y por qué se comporta de una determinada manera. Quienes interactúen con esos sistemas deben de saber que se trata de IA así como qué personas son sus responsables. Los sistemas deben notificar a los seres humanos que están interactuando con un sistema de IA a menos que sea evidente, que se les aplican reconocimiento emocional o sistemas de categorización biométrica, e incluso avisar de falsificaciones profundas (*deep fakes*), lo que debería hacerse extensivo a toda la desinformación.

Se abre la puerta a medidas adicionales voluntarias de transparencia. Los modelos de negocio de datos también deben ser transparentes.

No es nada seguro que se pueda garantizar técnicamente esta transparencia, pues cada vez se sabe menos de lo que pasa en el interior de algunos sistemas de IA. En este punto de transparencia la Comisión se centra demasiado en empresas privadas y no en el sector público, a diferencia de otras regulaciones nacionales o locales.

Instrumentos de control

En términos de gobernanza, la Comisión propone que las autoridades nacionales de vigilancia del mercado controlen la aplicación de las nuevas normas. También sugiere la creación de un Comité Europeo de IA que facilitará su aplicación e impulsará la creación de normas en materia de IA. Además, se proponen códigos de conducta voluntarios para la IA que no entrañe un alto riesgo, así como espacios controlados de pruebas para facilitar la innovación responsable.

Se trata de no sofocar la creatividad, especialmente ante la aspiración de que la UE se convierta en una potencia global en materia tecnológica. Se propone la puesta en operación de “areneros” (*sandboxes*) para el uso de datos por las *startups*. Pero la Comisión no precisa aún qué tipo de datos se usarán para esta experimentación con nuevos programas, destinados, por ejemplo, a mejorar el sistema de justicia o el sistema sanitario, entre otros, sin miedo a efectos de posibles errores. Es decir, ensayos en los campos de alto riesgo.

Se proponen unas obligaciones para los operadores (establecer e implementar un sistema de gestión de calidad en su organización, de documentación técnica, de registro para permitir a los usuarios supervisar el funcionamiento del sistema de IA de alto riesgo, y se les obliga a someterse a una evaluación de la conformidad y potencialmente reevaluación del sistema, en caso de modificaciones significativas, además de registrar el sistema de IA en la base de datos de la UE, realizar una monitorización post-mercado y colaborar con las autoridades de vigilancia del mercado. También se introduce para la IA un “marcado europeo”, una indicación de que un producto cumple los requisitos de una legislación pertinente de la Unión que regula el producto en cuestión. Los usuarios tienen asimismo una serie de obligaciones.

La propuesta abre la posibilidad de importantes multas por incumplimiento de esta nueva reglamentación, como el caso del RGPD, de hasta 30 millones de euros o el 6% de los ingresos globales de la empresa afectada, lo que, por ejemplo, en el caso de Facebook, podría llegar a más de 4.000 millones de dólares.

Inversiones

Junto con esta propuesta legislativa, la Comisión ha presentado un Plan Coordinado sobre IA, que viene a actualizar, post-pandemia, las estrategias al respecto que se había presentado. El Plan Coordinado describe los cambios e inversiones políticas necesarios a nivel de los Estados miembros y de la propia UE para fortalecer la posición de liderazgo de Europa en el desarrollo de la IA centrada en el ser humano, sostenible, segura, inclusiva y confiable. El objetivo es aumentar gradualmente la inversión pública y privada en IA en la UE hasta un total de 20.000 millones de euros anuales a lo largo de esta década. Es una cifra significativa. Se calcula que empresas de IA de EEUU atrajeron en 2019 25.000 millones de dólares, de un total de 40.000 millones en términos globales. A lo que en estos próximos años habrá que sumar el nuevo empuje a este respecto incluido en el plan de "infraestructuras" de 2-3 billones de dólares de la Administración Biden, con una parte importante para la I+D de tecnología emergentes, muy especialmente la IA.

Aspiración global

La propuesta legislativa y el Plan Coordinado es el mayor paso que ha tomado la Comisión Europa para crear un liderazgo global de la EU en materia de IA confiable. La UE se considera una superpotencia reguladora. Sobre todo, por el impacto global que ha tenido el RGPD. Dentro de este espíritu del "Efecto Bruselas", como lo bautizó Anu Bradford, la Comisión sabe que los fabricantes prefieren no tener que desarrollar varios estándares según los mercados. Países democráticos como Japón, Canadá y Australia están mostrando un gran interés por estas propuestas.

Pero EEUU se resistirá. Aunque hay nuevos vientos federales y a nivel de algunos estados, como California, para controlar el desarrollo tecnológico, EEUU intentará que no se repita el caso del RGPD, que le cogió bastante desprevenido. Y esto debería ser objeto, al menos, de un diálogo transatlántico, e incluso con China.

Son importantes a este respecto algunos movimientos que están surgiendo de las propias empresas, como el de Apple para limitar la captura de datos en los iPhones por las Apps (aunque hay un claro elemento de negocio en estos pasos).

Y es sabido que los árbitros no ganan partidos. Sin una base industrial fuerte para producir IA, la UE tendrá problemas para imponer sus normas.

Conclusiones

Críticas

Las reacciones han sido dispares. La Asociación de Consumidores Europeos (BEUC) la considera débil en términos de protección de estos, al ser demasiado dependiente en las propias valoraciones de la industria, y contemplar demasiadas excepciones. Desde el punto de vista de los fabricantes de *software*, las salvaguardias parecen adecuadas a la hora de maximizar las ventajas de la IA. Para la industria, en general estas limitaciones pueden representar una carga excesiva para muchos proveedores.

Estas nuevas medidas pueden ser visionarias, pero pueden frenar el desarrollo de la IA en Europa, mientras se amplía en EEUU, o, aún más, en China, con investigaciones e inversiones yéndose a lugares más permisivos al respecto como estos países, que ya son superpotencias en la materia, o incluso el Reino Unido, fuera de la UE pero potencia en IA y a favor de no limitar el reconocimiento facial con estas tecnologías en labores policiales.

Se intenta regular sobre consecuencias que aún no se conocen. Y se regula más lo privado que lo público. Se asume que todos los desarrolladores de IA en la UE van a tener los mismos tiempos, recursos y capacidades para demostrar que cumplen con todos los requisitos. Las empresas europeas ni tienen en general el mismo tamaño que las chinas o estadounidenses, ni la misma cantidad de personal para demostrar, a cada ocasión, todos estos requisitos.

Además, en este plan de la Comisión Europea para la IA, aunque en términos éticos vaya más lejos que muchos planteamientos, hasta ahora no aborda dos temas de, también, alto contenido ético: (1) el impacto de la, o las, IA en el empleo, parte esencial de la dimensión social y de la confianza de los ciudadanos que temen ante todo el efecto de la automatización en el trabajo; y (2) la ausencia del impacto en la emisión de gases de efecto invernadero, para casar la transición digital con la ecológica ante una IA que cada vez consume más electricidad.

Pero hay una crítica más fundamental que se puede hacer a la propuesta de la Comisión. ¿Es realmente regulable la IA con un reglamento general? Diversos expertos opinan que no. Un problema es que no tenemos una ley general que regule las bases de datos, sino leyes que abordan cuestiones específicas. Sin olvidar la complejidad de cumplir con estos requisitos legales cuando la IA se entremezcle con otras tecnologías (como el Internet de las Cosas). Una ley sobre crédito al consumo no es lo mismo que una ley sobre historias clínicas. Las técnicas de IA serán parte de cada pieza de *software*: escribir una ley para cubrir todo eso parece un nivel incorrecto de abstracción. No hay leyes que digan que no se le permite tener errores. Esto es como intentar escribir una ley única que cubra “los “automóviles”, es decir, a la vez la conducción en estado de ebriedad, los estándares de emisiones de GEI, el estacionamiento, el tratamiento fiscal de las carreteras, etc. La Comisión ha lanzado su plan para regular la IA, con un borrador de reglas y definiciones que seguramente resulten inútiles en pocos años.