Universidad Rey Juan Carlos

CÁTEDRA SERVICIOS DE INTELIGENCIA Y SISTEMAS DEMOCRÁTICOS

INFORME DE INTELIGENCIA

"IMPACTO SOCIAL DE LA INTELIGENCIA ARTIFICIAL"

Grupo 3

INVIA – Proyecto de Inteligencia Visual Analítica
Máster en Análisis de Inteligencia
Universidad Rey Juan Carlos

INDICE

1. Inteligencia Artificial y Gobernanza

1.1. Auditorías y responsabilidad proactiva

Una de las herramientas normativas que se han establecido para "garantizar y demostrar" el cumplimiento del Reglamento General de Protección de Datos (en adelante, RGPD) en el uso de Inteligencia Artificial (IA) es la realización de auditoria de los tratamientos. Asimismo, se plantea incorporar procesos de supervisión de códigos de conducta. Ambos instrumentos jurídicos requieren disponer de criterios objetivos para la evaluación de la adecuación normativa.

Desde el punto de vista del ciclo de vida del componente IA, podemos encontrarnos con los siguientes tratamientos de datos personales:

- Cuando se utilizan datos personales en la etapa de desarrollo del componente IA.
- Cuando se utilizan datos personales en las etapas de verificación o validación del componente IA.
- Cuando se incluye el componente IA en un tratamiento de datos personales (etapa de explotación), como podría ser un tratamiento de control de seguridad (que incluye reconocimiento facial) o de atención al ciudadano (que incluye en *chatbot*).
- En cualquier otra etapa del ciclo de vida que involucre datos personales.

Dichos tratamientos podrían utilizar conjuntos de datos (*datasets*) e inferir nuevos datos personales. Todos ellos, directos o indirectos, originales o derivados, son datos personales en tanto hagan referencia a un individuo identificado o identificable, objeto de protección de acuerdo con el artículo 1 del RGPD. Los datos personales se clasifican en identificadores, cuasi-identificadores y categorías especiales de datos.

1.2. Cinco modelos de gobernanza de la IA en el mundo

En el mundo existen actualmente cinco regímenes diferenciados 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:

- 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.

- 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.

- 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.

- 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.

- 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).

1.3. Datos, materia prima y regulación

Una de las principales características del marco normativo es la distinta densidad y carácter vinculante de las disposiciones en función de la cercanía a la protección de datos, o bien al de las herramientas de IA.

En la primera área, nos encontramos con normativa vinculante, como el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo (Reglamento General de Protección de Datos, RGPD), y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LO 3/2018) en el caso de España. En ambas normas, su aplicación objetiva corresponde al tratamiento total o parcialmente automatizado de datos personales, así como el tratamiento no automatizado de datos personales contenidos o destinados a ser incluidos en un fichero. La Ley española añade, respecto al RGPD, la protección de derechos digitales en sus artículos 80 a 97, que no están incluidos en el ámbito objetivo de aplicación del RGPD. Otro de los elementos incluidos por la ley nacional respecto al Reglamento europeo, es la protección de datos personales de personas fallecidas. Así, en su artículo 3 se establece que los familiares del fallecido y los herederos pueden solicitar el acceso a los datos personales del causante, salvo que lo hubiese prohibido expresamente.

En cuanto al ámbito subjetivo del RGPD, es importante destacar el hecho de que se aplique, aunque el tratamiento de datos personales no tenga lugar en el territorio de la Unión, e incluso aunque los responsables del tratamiento de datos no estén establecidos en la Unión en caso de bienes y servicios de conformidad a las reglas del Derecho internacional público. Esto supone una protección extra para los usuarios de IA que estén establecidos en la UE, ya que los datos de los usuarios quedarán protegidos por el RGPD independientemente del lugar en el que se contraten los servicios.

No obstante, no toda la materia relacionada con los datos trata sobre su protección o está relacionada con los derechos fundamentales. En este sentido y para mejorar la competitividad de la economía europea, el Parlamento Europeo y el Consejo aprobaron el Reglamento (UE) 2018/1807, de 14 de noviembre de 2018, relativo a un marco para la libre circulación de datos no personales en la Unión Europea. La UE pretende así fomentar la libre circulación de este tipo de datos al no afectar a los derechos fundamentales establecidos en los artículos 7 y 8 de la Carta de los Derechos Fundamentales de la Unión Europea. La libertad de elección de los proveedores de servicios beneficiaría considerablemente tanto al sector público como al sector privado, se conseguirían precios más competitivos y mejores prestaciones de servicios a los ciudadanos. De acuerdo con su artículo 3, los datos de carácter no personal son aquéllos no incluidos en el RGPD. Debido a las dificultades que se pueden ocasionar a la hora de identificar correctamente los datos no personales, el propio Reglamento en su considerando (9) menciona algunos ejemplos específicos entre los que se encuentran “los conjuntos de datos agregados y anonimizados utilizados para análisis de datos a gran escala, los datos sobre agricultura de precisión que pueden ayudar a controlar y optimizar la utilización de plaguicidas y de agua, o los datos sobre las necesidades de mantenimiento de máquinas industriales”.

2. Legislación Europea

2.1. Sumario: Normativa regulatoria IA

El pasado 21 de abril fue publicada la versión definitiva de la propuesta de Reglamento (UE) de la Comisión Europea sobre el marco jurídico aplicable a los sistemas de Inteligencia Artificial (IA). Dicha propuesta regula los sistemas de IA de alto riesgo y contiene, además, reglas de transparencia armonizadas para aquellos sistemas dirigidos a interactuar con personas físicas para generar o manipular imágenes, sonidos o contenidos de video. Se trata del primer marco legal Europeo sobre esta tecnología, que además llega acompañada de otra normativa sobre maquinaria y robots. No obstante, el texto se encuentra todavía en estado muy preliminar y, hasta su aprobación final, su contenido puede variar.

La combinación del primer marco jurídico sobre la IA de la Historia y de un nuevo plan coordinado con los Estados miembros, garantizará la seguridad y los derechos fundamentales de las personas y las empresas, al tiempo que reforzará la adopción, la inversión y la innovación en materia de IA en toda la UE. Al mismo tiempo, se han elaborado unas nuevas normas sobre maquinaria que utiliza inteligencia artificial, para dar mayor confianza a los usuarios en la nueva y versátil generación de productos.

2.2. La propuesta normativa

2.2.1. Una normativa global en la Unión Europea

Los beneficios potenciales de la IA para nuestras sociedades son múltiples, desde la mejora de la atención médica a la educación. Ante el rápido desarrollo tecnológico de la IA, la UE debe actuar a una para aprovechar estas oportunidades.

Si bien la mayoría de los sistemas de IA plantean un riesgo bajo o nulo, algunos de esos sistemas entrañan peligros que deben abordarse para evitar situaciones no deseadas. Por ejemplo, la opacidad de muchos algoritmos puede crear incertidumbre y obstaculizar la aplicación efectiva de la legislación vigente en materia de seguridad y derechos fundamentales. Para responder a estos retos, es necesaria una actuación legislativa para garantizar el correcto funcionamiento del mercado interior de los sistemas de IA, con una ponderación adecuada de los beneficios y de los riesgos. Esto abarca aplicaciones como los sistemas de identificación biométrica o decisiones de IA que afectan a intereses personales importantes, por ejemplo, en los ámbitos de la contratación, la educación, la asistencia sanitaria o la aplicación de la ley.

La propuesta de la Comisión de una normativa en materia de IA tiene por objeto garantizar la protección de los derechos fundamentales y la seguridad de los usuarios, a fin de que haya confianza en el desarrollo y la adopción de la IA.

2.2.2. Tipos de riesgos que aborda la normativa y marco de aplicación

Hay casos en los que las características específicas de determinados sistemas de IA pueden dar lugar a nuevos riesgos relacionados con la seguridad de los usuarios y los derechos fundamentales. Esto genera inseguridad jurídica para las empresas y una adopción potencialmente más lenta de las tecnologías de IA por parte de aquellas y los ciudadanos, debido a la falta de confianza. La disparidad de las respuestas reglamentarias de las autoridades nacionales entrañaría el riesgo de fragmentar el mercado interior.

El marco jurídico se aplicará a los agentes tanto públicos como privados de dentro y fuera de la UE, en la medida en que el sistema de IA se introduzca en el mercado de la Unión o su uso afecte a personas establecidas en ella. Puede afectar tanto a los proveedores (por ejemplo, un programador de una herramienta de evaluación de resúmenes curriculares) como

a los usuarios de sistemas de IA de alto riesgo (por ejemplo, un banco que compre esa herramienta). No se aplica a los usos privados no profesionales.

Los sujetos obligados son los siguientes:

- Los usuarios de sistemas de IA situados en la Unión;
- Los prestadores y usuarios de sistemas de IA que están situados en un tercer país cuando el resultado producido por el sistema se utiliza en la Unión.
- Los prestadores que introducen en el mercado o ponen en servicio sistemas de IA en la Unión.

Sin perjuicio de la regulación que recoge esta nueva norma y la clasificación en función del riesgo que para los derechos fundamentales puede comportar, resulta igualmente aplicable lo establecido en el RGPD y demás normativa comunitaria, normativa de consumidores y usuarios, seguridad de productos y servicios, salud, transporte o responsabilidad por productos entre otras.

2.2.3.Categorías de riesgo

La Comisión propone un planteamiento basado en el riesgo, con cuatro niveles de riesgo:

- Riesgo inadmisible: Se prohibirá un conjunto muy limitado de usos especialmente nocivos de la IA que contravienen los valores de la Unión al violar los derechos fundamentales (por ejemplo, puntuación social por parte de los Gobiernos, explotación de los puntos débiles de los niños, uso de técnicas subliminales y, salvo contadas excepciones, determinados sistemas de identificación biométrica remota en directo en espacios públicos con fines policiales).
- Alto riesgo: Se considera de alto riesgo un número limitado de sistemas de IA definidos en la propuesta y que tienen un impacto negativo en la seguridad de las personas o en sus derechos fundamentales (protegidos por la Carta de los Derechos Fundamentales de la UE). Se adjunta a la propuesta la lista de sistemas de IA de alto riesgo, que puede revisarse para adaptarla a la evolución de los casos de uso de la IA.

A fin de garantizar la confianza y un nivel elevado y coherente de protección de la seguridad y los derechos fundamentales, se proponen requisitos obligatorios para todos los sistemas de IA de alto riesgo. Esos requisitos se refieren a: la calidad de los conjuntos de datos utilizados; la documentación técnica y la llevanza de registro; la transparencia y la divulgación de información a los usuarios; la supervisión humana, la solidez, la precisión; y la ciberseguridad. En caso de infracción, los requisitos permitirán a las autoridades nacionales acceder a la información necesaria para investigar si el uso de la IA cumplió la legislación.

La normativa propuesta es coherente con la Carta de los Derechos Fundamentales de la Unión Europea y está en consonancia con los compromisos internacionales de la Unión en materia de comercio.

Los sistemas de IA considerados de alto riesgo abarcan las tecnologías de IA empleadas en: Infraestructuras críticas (por ejemplo, transportes), que pueden poner en peligro la vida y la salud de los ciudadanos; Formación educativa o profesional, que pueden determinar el acceso a la educación y la carrera profesional de una persona; Componentes de seguridad de los productos (ej. aplicación de IA en cirugía asistida por robots); Empleo, gestión de trabajadores y acceso al trabajo por cuenta propia (ej. programas informáticos de clasificación de CV); Servicios públicos y privados esenciales (ej. sistemas de calificación crediticia que priven a los ciudadanos de la oportunidad de obtener un préstamo); Aplicación de las leyes, que pueden interferir con los derechos fundamentales de las personas (ej. evaluación de la fiabilidad de las

pruebas); Gestión de la migración, el asilo y el control de las fronteras (ej. comprobación de documentos); Administración de justicia y procesos democráticos

2.2.4. Requisitos previos a la comercialización

Los sistemas de IA de alto riesgo estarán sujetos a obligaciones estrictas antes de que puedan comercializarse, debiendo garantizar:

- Sistemas adecuados de evaluación y mitigación de riesgos;
- Alta calidad de los conjuntos de datos que alimentan el sistema para minimizar los riesgos y los resultados discriminatorios;
- Registro de la actividad para garantizar la trazabilidad de los resultados;
- Documentación detallada que aporte toda la información necesaria sobre el sistema y su finalidad para que las autoridades evalúen su conformidad;
- Información clara y adecuada al usuario;
- Medidas apropiadas de supervisión humana para minimizar el riesgo;
- Alto nivel de solidez, seguridad y precisión.

Riesgo limitado: En el caso de determinados sistemas de IA se imponen obligaciones específicas de transparencia, por ejemplo, cuando exista un riesgo claro de manipulación (por ejemplo, mediante el uso de robots conversacionales). Los usuarios deben ser conscientes de que están interactuando con una máquina.

Riesgo mínimo: Todos los demás sistemas de IA pueden desarrollarse y utilizarse con arreglo a la legislación vigente sin obligaciones jurídicas adicionales. La inmensa mayoría de los sistemas de IA utilizados actualmente en la UE pertenece a esta categoría. De forma voluntaria, los proveedores de estos sistemas pueden optar por aplicar los requisitos de una IA digna de confianza y adherirse a códigos de conducta voluntarios.

Junto con una definición clara de lo que es de «alto riesgo», la Comisión presenta un método riguroso para contribuir a detectar los sistemas de IA de alto riesgo dentro del marco jurídico. El objetivo es aportar seguridad jurídica a las empresas y otros agentes económicos.

La clasificación del riesgo se basa en la finalidad prevista del sistema de IA, en consonancia con la legislación vigente de la UE en materia de seguridad de los productos. Esto significa que la clasificación del riesgo depende de la función desempeñada por el sistema de IA y de la finalidad y las modalidades específicas para las que se utilice dicho sistema.

Entre los criterios de esta clasificación se cuentan el alcance del uso de la aplicación de IA y su finalidad prevista, el número de personas potencialmente afectadas, la dependencia respecto al resultado y la irreversibilidad de los daños, así como la medida en que la legislación vigente de la Unión prevé medidas eficaces para prevenir o minimizar sustancialmente esos riesgos.

2.2.5. Peligros inherentes: Identificación biométrica remota ¿qué dice la normativa?

Con arreglo a las nuevas normas, todos los sistemas de IA destinados a utilizarse para la identificación biométrica remota de personas se considerarán de alto riesgo y estarán sujetos a evaluaciones de la conformidad *ex ante* y por terceros, incluida la documentación y los requisitos de supervisión humana desde su diseño. Ensayos y conjuntos de datos de alta calidad contribuirán a garantizar que estos sistemas sean precisos y que no se deriven efectos discriminatorios en la población afectada.

El uso de la identificación biométrica remota en tiempo real en lugares públicos con fines coercitivos plantea especiales riesgos desde el punto de vista de los derechos fundamentales y sobre todo de la dignidad humana, el respeto de la vida privada y familiar, la

protección de los datos personales y la no discriminación. Por lo tanto, está prohibido en principio, con algunas excepciones estrictas, estrictamente definidas, limitadas y reguladas, tales como su empleo por parte de la policía para la búsqueda selectiva de víctimas potenciales específicas de delitos, incluidos menores desaparecidos; la respuesta al peligro inminente de un atentado terrorista; o la detección e identificación de autores de delitos graves.

Por último, todos los sistemas de reconocimiento emocional y categorización biométrica estarán siempre sujetos a requisitos específicos de transparencia. También se considerarán aplicaciones de alto riesgo si entran en los casos de uso definidos como tales, por ejemplo, en los ámbitos del empleo, la educación, la aplicación de la ley, la migración y el control de fronteras.

La identificación biométrica puede adoptar distintas formas. Puede utilizarse para la autenticación del usuario, es decir, para desbloquear un teléfono inteligente o para la verificación o autenticación en los pasos fronterizos a fin de comprobar si la identidad de una persona coincide con la que figura en sus documentos de viaje (búsqueda de correspondencias uno a uno). También puede servir para identificar a distancia personas en una multitud, por ejemplo, mediante la comparación de la imagen de una persona con las contenidas en una base de datos (búsqueda de correspondencias de uno a varios).

La precisión de los sistemas de reconocimiento facial puede variar considerablemente en función de factores muy diversos, tales como la calidad de la cámara, la luz, la distancia, la base de datos, el algoritmo y la etnia, edad o sexo del sujeto. Lo mismo se aplica a los sistemas de reconocimiento vocal y de la forma de andar y otros sistemas biométricos. Unos sistemas muy avanzados están reduciendo sin cesar sus tasas de falsa aceptación. Si bien un índice de precisión del 99 % puede parecer bueno en general, se trata de un riesgo considerable cuando pueda llevar a sospechar de una persona inocente. Incluso un porcentaje de error del 0,1 % es alto si afecta a decenas de miles de personas.

2.2.6.Obligaciones de los proveedores de sistemas de IA de alto riesgo

Antes de comercializar un sistema de IA de alto riesgo en el mercado de la Unión o de ponerlo en servicio de otra forma, los proveedores deberán someterlo a una evaluación de la conformidad.

Esto les permitirá demostrar que su sistema cumple los requisitos obligatorios de una IA digna de confianza (por ejemplo, calidad de los datos, documentación y trazabilidad, transparencia, supervisión humana, exactitud y solidez). En caso de que el sistema en sí o su finalidad se modifiquen sustancialmente, deberá repetirse la evaluación. En el caso de determinados sistemas de IA, también deberá participar en este proceso un organismo notificado independiente. Los sistemas de IA que sean componentes de seguridad de productos contemplados en la legislación sectorial de la Unión siempre se considerarán de alto riesgo cuando sean objeto de una evaluación de la conformidad por terceros con arreglo a dicha legislación sectorial. También en el caso de los sistemas de identificación biométrica, siempre hará falta una evaluación de la conformidad por terceros.

Los proveedores de sistemas de IA de alto riesgo también tendrán que aplicar sistemas de calidad y gestión de riesgos para garantizar su conformidad con los nuevos requisitos y minimizar los riesgos para los usuarios y las personas afectadas, incluso después de que un producto se haya comercializado. Las autoridades de vigilancia del mercado apoyarán el seguimiento posterior a la comercialización mediante auditorías y ofreciendo a los proveedores la posibilidad de informar sobre incidentes graves o violaciones de los derechos fundamentales de que hayan tenido conocimiento.

Los Estados miembros desempeñan un papel clave en la aplicación y el cumplimiento del presente Reglamento. A este respecto, cada Estado miembro deberá designar una o varias

autoridades nacionales competentes para supervisar la aplicación y ejecución, así como para realizar actividades de vigilancia del mercado. A fin de aumentar la eficiencia y establecer un punto de contacto oficial con la población y otros homólogos, cada Estado miembro deberá designar una autoridad nacional de supervisión, que también representará al país en el Comité Europeo de Inteligencia Artificial.

Cuando se comercialicen o utilicen sistemas de IA que no respeten los requisitos del Reglamento, los Estados miembros deberán establecer sanciones efectivas, proporcionadas y disuasorias, incluidas multas administrativas, por las infracciones y notificarlas a la Comisión.

- El Reglamento fija umbrales que deberán tenerse en cuenta:
    - Hasta treinta millones de euros o el 6 % del volumen de negocios anual total a escala mundial del ejercicio financiero anterior, optándose por la de mayor cuantía, por las infracciones por incumplimiento o prácticas prohibidas en relación con los requisitos sobre los datos;
    - Hasta veinte millones de euros o el 4 % del volumen de negocios anual total a escala mundial del ejercicio financiero anterior por el incumplimiento de cualquier otro requisito obligación del Reglamento;
    - Hasta diez millones de euros o el 2 % del volumen de negocios total anual a escala mundial del ejercicio anterior por el suministro de información incorrecta, incompleta o engañosa a los organismos notificados y a las autoridades nacionales competentes en respuesta a una solicitud.
- Como las instituciones, agencias y organismos de la UE deben dar ejemplo, también estarán sujetas a las normas y a las posibles sanciones; el Supervisor Europeo de Protección de Datos estará facultado para imponerles multas.

2.2.7.¿Qué es el Comité Europeo de Inteligencia Artificial?

El Comité Europeo de Inteligencia Artificial estará formado por representantes de alto nivel de las autoridades nacionales de supervisión competentes, el Supervisor Europeo de Protección de Datos y la Comisión. Su función consistirá en facilitar una aplicación fluida, eficaz y armonizada del nuevo Reglamento sobre la IA.

2.2.8.¿Cómo protegerán las normas los derechos fundamentales?

Ya existe una sólida protección de los derechos fundamentales y la no discriminación a escala de la UE y de los Estados miembros, pero la complejidad y la opacidad de determinadas aplicaciones de IA («cajas negras») plantean problemas. Un planteamiento centrado en el ser humano en materia de IA supone velar por que las aplicaciones de IA cumplan la legislación en materia de derechos fundamentales. Los requisitos de rendición de cuentas y transparencia para el uso de sistemas de IA de alto riesgo, combinados con mejores capacidades de ejecución, garantizarán que se tenga en cuenta el cumplimiento de la legislación en la fase de desarrollo.

¿Cómo hace frente este Reglamento a los sesgos raciales y de género en la IA?

- Es muy importante que los sistemas de IA no creen ni reproduzcan sesgos. Al contrario, si se diseñan y utilizan adecuadamente, los sistemas de IA pueden contribuir a reducir los prejuicios y la discriminación estructural existente y, por tanto, facilitar decisiones más equitativas y no discriminatorias (por ejemplo, en la contratación de personal).
- Los nuevos requisitos obligatorios aplicables a todos los sistemas de IA de alto riesgo servirán para este fin. Los sistemas de IA deberán ser técnicamente sólidos a fin de garantizar que la tecnología se ajuste a su finalidad y los falsos resultados positivos o negativos no afecten de manera desproporcionada a los grupos protegidos (por ejemplo, origen racial o étnico, sexo, edad, etc.).
- Los sistemas de alto riesgo tendrán que entrenarse y ensayarse con conjuntos de datos lo suficientemente representativos como para reducir al mínimo el riesgo de sesgos

injustos incorporados al modelo y velar por que puedan solventarse mediante las medidas apropiadas de detección y corrección de sesgos y otras medidas paliativas.
- También deberán poder rastrearse y auditarse, garantizando la conservación de la documentación pertinente, incluidos los datos utilizados para entrenar el algoritmo, algo que es clave en las investigaciones *a posteriori*.

El sistema de cumplimiento antes y después de su comercialización habrá de garantizar que estos sistemas sean objeto de un seguimiento periódico y que se solventen rápidamente los riesgos potenciales.

Códigos de conducta voluntarios:

Los proveedores de aplicaciones de alto riesgo podrán garantizar la fiabilidad de su sistema de IA formulando sus propios códigos de conducta voluntarios o adhiriéndose a códigos de conducta adoptados por otras asociaciones representativas, que se aplicarán simultáneamente a las obligaciones de transparencia de determinados sistemas de IA. La Comisión animará a las asociaciones del sector y a otras organizaciones representativas a adoptar códigos de conducta voluntarios.

Otras normativas: importaciones de sistemas y aplicaciones de IA

Los importadores de sistemas de IA tendrán que cerciorarse de que el proveedor extranjero ya haya completado el procedimiento de evaluación de la conformidad pertinente y de que disponga de la documentación técnica exigida por el Reglamento. Además, los importadores deberán asegurarse de que su sistema lleve un marcado europeo de conformidad (CE) y vaya acompañado de la documentación y las instrucciones de uso necesarias.

2.3. Plan coordinado: actualización de 2021

2.3.1. Objetivos

El plan coordinado presenta una serie concreta de actuaciones conjuntas de la Comisión Europea y los Estados miembros sobre la manera de alcanzar un liderazgo mundial de la UE en materia de IA digna de confianza. Las acciones clave propuestas responden a la idea de que, para tener éxito, la Comisión Europea, junto con los Estados miembros y los agentes privados, debe: acelerar las inversiones en tecnologías de IA para impulsar una recuperación económica y social resiliente facilitada por la adopción de «nuevas» soluciones digitales; actuar en materia de estrategias y programas de IA mediante su plena y oportuna aplicación para velar por que la UE aproveche plenamente la ventaja que conlleva ser pioneros, y armonizar la estrategia en materia de IA para eliminar la fragmentación y hacer frente a los desafíos mundiales.

2.3.2. ¿Cómo impulsará la UE la excelencia desde el laboratorio hasta el mercado?

El plan coordinado revisado contempla la idea de cofinanciar las instalaciones de ensayo y experimentación, que pueden convertirse en un recurso común y altamente especializado a escala de la UE que fomente el rápido despliegue y una mayor aceptación de la IA.

Además, la Comisión también va a crear una red de centros europeos de innovación digital, que son «ventanillas únicas» para ayudar a las pymes y a las Administraciones públicas a ser más competitivas en este ámbito.

2.3.3. ¿Cómo va a alcanzar la UE el liderazgo estratégico en sectores de gran impacto?

Para adaptarse a la evolución del mercado y a las actuaciones en curso en los Estados miembros y reforzar la posición de la UE a escala mundial, el plan coordinado propone siete nuevos ámbitos de acción sectoriales.

2.3.4. ¿Cómo invertirán los Estados miembros en IA?

Maximizar los recursos y coordinar las inversiones es vital y un componente esencial de la estrategia de la Comisión en materia de IA. Con cargo al programa Europa Digital, el primer instrumento financiero de la UE centrado en la tecnología digital, y el programa Horizonte Europa, la Comisión tiene previsto invertir mil millones de euros al año en IA. El objetivo es movilizar inversiones adicionales del sector privado y de los Estados miembros para alcanzar un volumen de inversión anual de 20 000 millones de euros a lo largo de esta década. El Mecanismo de Recuperación y Resiliencia recientemente adoptado, que es el mayor paquete de estímulo jamás financiado con cargo al presupuesto de la UE, destina 134 000 millones de euros al sector digital. Supondrá un punto de inflexión que permitirá a Europa tener mayores ambiciones y convertirse en líder mundial en el fomento de una IA de vanguardia y digna de confianza.

2.3.5. Nuevo Reglamento sobre máquinas

Cómo se relaciona el Reglamento sobre máquinas con la IA

El Reglamento sobre máquinas velará por que la nueva generación de maquinaria garantice la seguridad de los usuarios y consumidores, y estimule la innovación. La maquinaria abarca una amplia gama de productos de consumo y profesionales, desde robots hasta cortadoras de césped, impresoras 3D, máquinas de construcción y líneas de producción industrial.

Cómo beneficiará a las empresas, especialmente a las pymes

Las empresas solo tendrán que realizar una única evaluación de la conformidad para ambos Reglamentos (IA y máquinas). La nueva legislación reducirá la carga administrativa y financiera de los fabricantes al permitir formatos digitales para las instrucciones y la declaración de conformidad, y al solicitar una adaptación de las tasas para las pymes cuando se necesite un tercero para la evaluación de la conformidad de la máquina.

2.4. Otras áreas legales importantes

2.4.1. Aspectos éticos

La propuesta persigue el respeto a la dignidad humana, los derechos fundamentales, la autonomía y determinación de la personal, así como otros factores tales como la prevención de daños, la promoción de la equidad, la inclusión y la transparencia y la eliminación de sesgos y discriminación.

La propuesta parte de que el ser humano debe ser el centro de esta tecnología, estableciendo obligaciones específicas para los sistemas de IA que supongan un “alto riesgo”. Estas tecnologías de “alto riesgo” deberán ser claramente delimitadas y exhaustivamente listadas en la regulación. Así, deberán basarse en criterios objetivos relativos tales como: (i) la capacidad que tiene la tecnología de causar un daño o violar un derecho fundamental o la normativa de seguridad aplicable; y (ii) al sector y la finalidad para la cual esta se utilice. Este listado deberá ser revisado periódicamente.

La propuesta abarca diferentes escenarios y crea un “test de responsabilidad ética” que deberá ser aprobado por las empresas que quieran utilizar un sistema de IA y presenten un riesgo mayor. En definitiva, se trata de una evaluación ex ante imparcial, regulada y externa, realizada por un órgano público y apoyada en criterios concretos y definidos.

Entre las obligaciones, se encontrarían también el cumplimiento de los requisitos ya recomendados por el *High Level Expert Group*en IA, creado por la Comisión Europea, en su “Guía para conseguir una IA confiable” (no discriminación, factor humano, seguridad, transparencia, *accountability* y trazabilidad) y la obligación de realizar auditorías periódicas y solicitar certificaciones.

Para aquellas tecnologías que no suponen un "alto riesgo", el esquema propuesto sería, por el momento, voluntario. No obstante, ya ha habido algún estado (como Alemania) que se han pronunciado mostrando su preocupación a esta regulación "voluntaria".

En materia de protección de datos, el PE se remite al cumplimiento del RGPD (en materia de IA y protección de datos, os remitimos a nuestra entrada de blog)

2.4.2. Responsabilidad civil

El PE propone ajustar ciertas cuestiones tanto en la regulación relativa a la responsabilidad de productos (*Product Liability Directive*) como en la relativa a la seguridad de productos (*Product Safety Directive*).

Esto se debe a que existen diversos factores relacionados con la tecnología que incorpora IA que deben ser considerados y que la diferencia del resto de productos ofrecidos en el mercado (por ejemplo, su complejidad, su grado de conectividad, su posible falta de transparencia, sus vulnerabilidades, su capacidad de aprendizaje autónomo o nivel de autonomía).

Se enfatiza el hecho de que estas tecnologías no tienen personalidad jurídica y que su único objetivo es servir a la humanidad. Siguiendo los principios de responsabilidad, todo aquel que crea, mantiene, controla o interfiere debe ser responsable en cierta medida y según los distintos criterios propuestos, por los daños que haya podido causar la actividad.

Por tanto, se propone distribuir la responsabilidad entre todos los actores que participan en las cadenas de valor de estas tecnologías (incorporando a los desarrolladores, fabricantes, programadores, operadores, etc.). Una novedad importante es que se diferencia entre la responsabilidad del llamado operador inicial (*frontend operato*r) y el operador final (*backend operator*). Pese a que en teoría el operador inicial es el que decide el uso de un sistema de IA, el operador final podría tener un alto nivel de control si se le considera un "productor" bajo el Artículo 3 de la Directiva de Responsabilidad por Productos Defectuosos, y consecuentemente tener un nivel de responsabilidad alto.

Se propone revertir las reglas de la carga de la prueba del daño causado en ciertos casos definidos, proponiendo, por tanto:

- Un sistema de responsabilidad objetiva para operadores de sistemas de IA de alto riesgo, que serán responsables de los daños y perjuicios causados por una actividad física o virtual, un dispositivo o un proceso gobernado por dicho sistema de IA. No podrá eludir la responsabilidad alegando que actuaron con diligencia debida o que el daño o perjuicio fue causado por una actividad gobernada por su sistema de IA. La responsabilidad se incluirá en los casos de fuerza mayor.
- Un sistema de responsabilidad subjetiva para sistemas que no constituyan un alto riesgo y que deberán responder de los daños provocados (salvo que puedan demostrar la falta de culpa con base en motivos tasados). No podrá eludir la responsabilidad argumentando que el daño o perjuicio fue causado por una actividad, dispositivo o proceso autónomo gobernado por su sistema de IA.
- Un sistema de responsabilidad solidaria, en caso de que haya más de un operador de la IA.

2.4.3. Propiedad intelectual e industrial

El PE considera que es necesario diferenciar entre las creaciones que hayan sido generadas por la IA y aquellas otras creaciones humanas en las que haya intervenido un sistema de IA.

En las creaciones humanas en las que haya participado la IA, el régimen jurídico actual de propiedad intelectual e industrial seguiría siendo aplicable, siendo el autor el titular de los derechos.

No obstante, el PE considera que las creaciones generadas autónomamente por la IA no podrían ser protegidas bajo este derecho, para garantizar la observancia del principio de originalidad (ligado a la personalidad del autor y a la naturaleza humana).

En todo caso, el PE entiende que estas creaciones generadas por la IA deberían ser de alguna forma protegidas para fomentar la inversión y mejorar la seguridad jurídica. Propone evaluar la posibilidad de conceder los derechos de autor sobre este tipo de "creación obra" a la persona física que la edite y la haga pública de forma lícita (siempre que el diseñador/es de la tecnología subyacente no se opongan a dicho uso).

Por otro lado, el pasado 20 de enero de 2021 fue publicado un informe de propuesta de resolución del Parlamento Europeo en materia de inteligencia artificial y la interpretación y aplicación del Derecho internacional en los ámbitos de los usos civil y militar. A continuación, se resumen algunos de los principales aspectos que dicha propuesta cubre:

2.4.4. Aspectos de uso militar y supervisión humana

El Parlamento Europeo aboga por sistemas que permitan al humano ejercer un alto nivel de control sobre los sistemas de IA de modo que se tenga en todo momento los medios para corregir su curso, detenerla o desactivarla en caso de comportamiento imprevisto, intervención accidental, ciberataque o interferencia de terceros con tecnología pasada en IA.

Se defiende que los "Sistemas autónomos armamentísticos letales" (SAAL), deben emplearse únicamente cómo último recurso y solo son lícitos si están sujetos a un estricto control humano. Los sistemas sin control humano ("*human off the loop*") ni supervisión humana de ningún tipo deben prohibirse sin excepciones.

El texto también hace un llamamiento a promover un marco global para el uso militar de la IA, junto con la comunidad internacional.

2.4.5. Aspectos de la Inteligencia Artificial en el sector público

El incremento en el uso de sistemas de IA en salud y justicia nunca debería substituir al contacto humano; todo individuo debería ser informado cuando una decisión está siendo tomada por una IA y tener la opción de una segunda opinión.

Cuando la IA se utiliza en asuntos de salud pública (por ejemplo, cirugía asistida por robots, prótesis inteligentes, medicina predictiva), los datos personales de los pacientes deben ser protegidos y se debe mantener el principio de igualdad de trato.

Si bien el uso de tecnologías de IA en el sector de la justicia puede ayudar a acelerar los procedimientos y a tomar decisiones más racionales, las decisiones judiciales finales deben ser tomadas por humanos, ser estrictamente verificadas por una persona y estar sujetas a un proceso debido.

2.4.6. Aspectos de vídeovigilancia masiva y *deepfakes*

Los reguladores están también altamente preocupados por las amenazas de una violación de los derechos humanos y soberanía estatal que se pueda llevar a cabo con el uso de las tecnologías de IA.

Piden que se prohíba a las autoridades el uso de «*aplicaciones de calificación social masiva altamente intrusivas*» (para controlar y calificar a los ciudadanos) dado que ponen en grave peligro el respeto de los derechos fundamentales.

El informe también plantea la preocupación por las tecnologías de ultra falsificación (*deepfake*), que permiten falsificar de manera cada vez más realista fotos, audio y video que podrían utilizarse para chantajear, crear bulos, mermar la confianza de la ciudadanía e influir sobre el discurso público. Solicita que se obligue a todo documento ultra falseado o video realista realizado mediante técnicas de síntesis que incluya una etiqueta que lo califique como "no original".

3. Inteligencia Artificial y el ámbito laboral

En relación con el impacto sobre el empleo que puede producirse por el desarrollo de este tipo de sistemas, se ha considerado, por una parte, la amenaza de la destrucción de puestos de trabajo que se desempeñan en la actualidad y qué sectores pudieran verse afectados más profundamente. También, se han abordado factores como la formación necesaria para asumir el cambio de paradigma en el empleo al objeto de adaptarnos y anticiparnos a la evolución de la IA y también las oportunidades que se nos presentan para la mejora de la calidad de vida de los ciudadanos a nivel global y especialmente en el ámbito laboral.

Es evidente que millones de puestos de trabajo dejarán de existir, pero se está produciendo demanda en nuevas profesiones, principalmente con un perfil tecnológico, en las que se ofrece una mejor remuneración y que presentan un escaso desempleo.

Entre los sectores con mayor impacto en el empleo existe un amplio consenso que se verán afectados aquellos relacionados con tareas administrativas, el transporte y la logística. Por otra parte, la IA no sólo afectará a la disponibilidad y características de los empleos, también debe señalarse que la propia gestión de los recursos humanos de una empresa, incluyendo la evaluación del desempeño de los trabajadores, no serán ajenos a la utilización de estos sistemas.

Por tanto, las nuevas generaciones de jóvenes que se forman para entrar el mercado de trabajo e incluso los que ya tienen voluntad de hacerlo, deben asumir este cambio en el paradigma el empleo y no perder más tiempo en adaptarse a la influencia de la IA.

Este trayecto no es mi mucho menos apocalíptico, de hecho, nuestra necesidad de interacción social será un elemento que asegurará la continuidad de muchos empleos que podrían hacerse con IA pero que seguirá siendo desempeñado por personas. También hay estimaciones de que el balance será finalmente positivo en creación de empleo neto y de qué en determinados entornos laborales se liberará a personas de tareas físicas y se reducirán los riesgos para la seguridad y salud de los trabajadores.

De la misma forma se han señalado impactos positivos de la IA en áreas radicalmente distintas como pueden ser en materia de Medicina Preventiva, en la contribución a la eliminación de sesgos en la toma de decisiones, e incluso en la reducción de la brecha de género.

Para afrontar con éxito el reto reducir los riesgos que puedan producirse por los avances de la IA deben potenciarse especialmente los aspectos éticos y morales, apostarse decididamente en el ámbito de la formación por el pensamiento computacional y abarcando también aspectos psicosociales que incluyan la potenciación de la inteligencia emocional y la creatividad.

3.1. El Informe sobre Sociedad Digital en España: Habilidades de IA demandadas

En el informe “Sociedad Digital en España. El año en que todo cambió” de la Fundación Telefónica, se estima que la digitalización permitiría aumentar el PIB entre un 1,5 y 2,5 puntos porcentuales hasta 2025 e incrementar la productividad de las PYMES entre un 15 y un 25% y también se incluye la propuesta de una iniciativa para apoyar el aprendizaje de por vida por el futuro del trabajo para permitir la recualificación a escala de los ciudadanos y equipar a todos los ciudadanos con habilidades cognitivas y digitales que necesitan para triunfar en el contexto del futuro trabajo.

Según se señala en el informe, en una encuesta realizada a los empresarios por la Comisión Europea relativa a las habilidades de inteligencia artificial demandadas, muestra que España está por encima de la media en relación con la demanda de expertos en aprendizaje automático y modelización y en robótica, así como en la necesidad de trabajadores con

conocimientos en programación, con una menor diferencia respecto del resto de países europeos en la demanda *big data* y *cloud computing.*

HABILIDADES DE INTELIGENCIA ARTIFICIAL DEMANDADAS

Aprendizaje Automático y Modelación	Cloud Computing	Gestión de Big Data	Programación	Robótica
ESPAÑA 46% EUROPA (27) 39%	ESPAÑA 33% EUROPA (27) 33%	ESPAÑA 49% EUROPA (27) 43%	ESPAÑA 60% EUROPA (27) 52%	ESPAÑA 50% EUROPA (27) 31%

Fuente: European Commission (2020), *European enterprise survey on the use of technologies based on artificial intelligence.*

En el documento de la Fundación Telefónica también se hace referencia a otros estudios en los que más que una pérdida de empleos se abren oportunidades para la colaboración entre personas y máquinas, lo que se conoce como *Fusion Skills.*

"No obstante, solamente el 2,5% de los empleos comprenden un elevado un elevado número de tareas susceptibles de ser realizadas por sistemas basados en el aprendizaje automático. La innovación requerirá la relocalización de recursos y la recualificación de los trabajadores, no solamente en temas técnicos, sino en habilidades tan «humanas» como pueden ser la creatividad, la comunicación, y la capacidad para juzgar y tomar decisiones".

En relación con las *habilidades de fusión* se introduce una lista de ocho puntos elaborada por Paul Daugherty y James Wilson:

- Rehumanizar el tiempo.
- Normalización responsable.
- Integración de juicio.
- Interrogatorio inteligente.
- Empoderamiento basado en robots.
- Fusión holística.
- Aprendizaje recíproco.
- Reinvención continua.

3.2. La transformación del empleo y creación neta de puestos de trabajo

En la misma línea de análisis en la que se estima una transformación del empleo más que una pérdida masiva de puestos de trabajo, se encuentra la Directora global en ciencia de datos para Vodafone, Nuria Oliver:

"Todos los estudios anticipan una transformación radical en la que la se van a destruir millones de puestos de empleo, pero se van a crear muchos más millones de puestos de trabajo. Según el foro económico mundial se anticipa que **habrá una creación neta de 58 millones de puestos de trabajo**, pero lo importante es entender que estos puestos de trabajo nuevos van a ser radicalmente diferentes de los puestos que se han destruido…Yo propongo dos cosas: primero que se incorpore una asignatura troncal de **pensamiento computacional**…pero también propongo que se refuerce mucho más la creatividad y las inteligencias social y emocional que creo que tampoco estamos reforzando…"

Según Nuria Oliver, la formación en *pensamiento computacional* abarcaría cinco áreas de competencia:

- Pensamiento algorítmico.
- Ciencia de datos.
- Redes.
- Programación.
- Hardware y electrónica.

La empresa Repsol, en su metodología para la Industria 5.0, nos presenta la automatización de procesos como un aporte de seguridad y flexibilidad para los trabajadores y un reenfoque de estos en tareas de mayor añadido. Profundizando en los argumentos que hemos venido exponiendo, la compañía energética pone en valor las competencias en conocimiento de robótica y de tecnologías digitales para la experimentación, el uso de herramientas de gestión de la información, trazabilidad de flujos de información y automatización de procesos con RPA.

3.3. Sociedad 5.0

Para poner al ser humano en el centro del desarrollo tecnológico y aumentar su bienestar, Japón ya está trabajando en lo que se denomina Sociedad 5.0, un modelo de desarrollo social apoyado en la Inteligencia Artificial con el objetivo de crear ciudades más sostenibles. A la vanguardia de este proyecto se encuentran las localidades de Aizuwakamatsu y Arao.

3.4. Retos y oportunidades de la IA en el mercado laboral

3.4.1. La brecha entre los líderes en talento en IA (casi todos ellos, países con ingresos altos) y el resto del planeta está creciendo.

El talento en IA es escaso y está distribuido de forma desigual en las diferentes industrias, sectores y países. Más de la mitad de la población mundial en desarrollo carece de competencias digitales básicas. En la era de la IA, esta diferencia en competencias digitales está aumentando, con unos pocos países que progresan rápidamente, mientras la mayoría del mundo en desarrollo se queda atrás.

3.4.2. Convertir la IA en una fuerza positiva requiere un enfoque proactivo y colaborativo.

La IA puede tener una función clave en la prestación de soluciones para ayudar a la humanidad a alcanzar los Objetivos de Desarrollo Sostenible (ODS) de la ONU: la educación (con programas digitales personalizados) y la salud (con diagnósticos y seguimiento remotos personalizados, además de análisis de Big Data para supervisar y reducir las enfermedades endémicas y epidemias) son dos de los ejemplos más inmediatos.

Las tecnologías digitales basadas en la IA pueden permitir que segmentos más amplios del mercado laboral mejoren su productividad y tengan acceso a empleos mejor remunerados, lo cual, a su vez, puede contribuir a promover el crecimiento inclusivo.

3.4.3. Los países en desarrollo podrían beneficiarse de la IA

La gran reducción en costes de capital propiciada por las aplicaciones de IA, junto al hecho de que la dirección del cambio tecnológico es, al menos en parte, orientada por la oferta relativa de trabajadores poco cualificados en comparación con los muy calificados, lo que supondría un beneficio para los trabajadores de los países en desarrollo.

3.4.4. Las políticas en materia de competencias profesionales son indispensables, mas no suficientes.

Es necesario que la educación y la formación vayan mucho más allá de los años escolares, de manera que los trabajadores puedan capacitarse o reciclarse profesionalmente cuando sea necesario a lo largo de sus carreras. El aprendizaje permanente debe convertirse en

una realidad para que el mundo del trabajo pueda beneficiarse de estas nuevas tecnologías, ahora y en el futuro.

También es relevante garantizar la difusión de las nuevas tecnologías en todo el mundo y permitir el acceso a los datos. Los responsables de la toma de decisiones y los interlocutores sociales deben además garantizar que ciertas empresas no logren dominar el mercado y así excluyan a otras. El aumento observado en la concentración del mercado entre las empresas digitales es motivo de preocupación y deben tomarse medidas decisivas.

4. Inteligencia Artificial y la desinformación: desafíos y oportunidades

4.1. Desafíos

La Inteligencia Artificial también ha tenido influencia en la forma en la que se produce y difunde la información. Su desarrollo nos enfrenta a nuevos riesgos y cuestiones éticas que trataremos a continuación.

La IA aplicada al sector de la comunicación está siendo utilizada para finalidades diversas. Así, se han desarrollado tecnologías que son capaces de realizar recopilaciones masivas de datos que permiten conocer características, necesidades, creencias y vulnerabilidades de grupos de personas, creando perfilados de forma automática que pueden ser empleados por partidos políticos en período electoral o por compañías para lanzar campañas de publicidad dirigida.

La IA también está siendo utilizada para generar contenido automáticamente de forma masiva a través de modelos GPT-3. Para ello, se hace uso de los perfilados automáticos generados por la recopilación masiva de datos para dirigir la información a individuos objetivo con la intención de intervenir intencionadamente en la formación de la opinión popular y sus procesos de toma de decisión. Las problemáticas que acontecen en estas prácticas son, por un lado, los mecanismos a través de los cuales se recopilan los datos de los usuarios, que en muchas ocasiones representan una amenaza desde el punto de vista de la protección de datos personales y, por otro, que el contenido que genere la IA tenga como fin la desinformación.

La influencia de los algoritmos ha permitido lanzar operaciones de influencia o campañas de desinformación para distorsionar de forma intencional la percepción pública sobre un asunto en particular. La European Comission´s High Level Expert Group on Fake News and Online Disinformation define desinformación como toda información falsa, imprecisa o engañosa, diseñada y promovida para provocar un daño público con fines de lucro. Las operaciones de desinformación intervienen y manipulan la formación de la opinión pública, generan confusión y degradan la confianza en los medios de comunicación, las instituciones e incluso en los procesos democráticos.

Otro de los desafíos que presenta la aplicación de la IA es la producción de contenido en formato audio y vídeo. Estas prácticas se conocen como '*deep fakes'*, que consisten en manipular digitalmente material audiovisual de forma que adquiera apariencia realista para inducir a engaño al que consume dicho contenido. El problema principal que se ha planteado acerca de esta tecnología es el acceso a un conjunto de datos de entrenamiento en formato vídeo y audio lo suficientemente grande para que la representación parezca genuina. En 2017, científicos de la universidad de Washington produjeron un vídeo falso en el que aparecía Barack Obama. Este tipo de prácticas contribuyen a la desinformación y llevan consigo el riesgo de desestabilizar la confianza popular hacia las instituciones, enfatizando la polarización en las sociedades, profundizando tensiones y socavando la credibilidad de los ciudadanos en los procesos democráticos. Así lo ha expresado la Comisión Europea en su estrategia para dar forma al futuro de Europa. Esta institución considera que la desinformación erosiona la confianza en las instituciones, medios de comunicación y los procesos de toma de decisión de los ciudadanos. En la misma línea, la Asamblea Parlamentaria del Consejo de Europa (PACE) expresa en la Resolución 2326 (2020) su preocupación "por la escalada de contaminación de la información en un mundo conectado digitalmente y cada vez más polarizado".

4.2. Oportunidades

Con motivo de combatir la desinformación se han puesto en marcha numerosas iniciativas para verificar datos, aunque la mayor parte de ellas son manuales. *Duke Reporters'*

Lab ha identificado 290 proyectos activos de verificación de datos en el censo que realizaron en junio de 2020 (Stencel & Luther, 2020). Este mismo informe indica que, tras años de crecimiento constante, hay indicios de que la tendencia de nuevos verificadores se esté desacelerando desde julio de 2019, a pesar de que el contenido engañoso cada vez desempeña un papel más relevante en la formación de la opinión pública.

Ante el constante incremento de información falsa en el ciberespacio, la verificación manual es cada vez más ineficaz e ineficiente. Por esta razón, se plantea la IA como la solución definitiva para contrarrestar la amenaza de la desinformación y evitar la distorsión de la opinión pública, detectando y eliminando de forma automatizada contenido con información falsa. Sin embargo, esta tecnología se encuentra actualmente en pleno desarrollo y la verificación automatizada de la información que recibimos supone un objetivo lejano aún. Las principales limitaciones que presenta la verificación de hechos a través de la IA es su capacidad para evaluar la precisión de las declaraciones de los individuos. Los modelos de IA son propensos a falsos negativos dado que el sarcasmo, la ironía o los entornos culturales y políticos específicos de cada país son elementos complejos que la IA aún no domina. Asimismo, esta tecnología contiene también sesgos cognitivos que puede afectar a la desinformación, ya que está influenciada por las personas que la diseñan y entrenan, así como también por los datos poco representativos, erróneos o imprecisos a través de los que se va alimentando. Esto es lo que se denomina el sesgo de los algoritmos.

5. Sistemas de armadas autónomos letales (LAWS, por sus siglas en inglés)

Los Sistemas de Armas Autónomos (LAWS) suponen un cambio de paradigma en la manera de hacer la guerra, de magnitud similar a la aparición de la pólvora o el arma atómica. Su implementación incrementa la eficacia y la efectividad de los ejércitos que las poseen hasta niveles inéditos, en comparación con los sistemas operados por humanos, hasta tal punto que se prevé una nueva carrera armamentística a nivel global por la incorporación de esta tecnología, ligada a la del desarrollo de la Inteligencia Artificial.

El desarrollo de estos sistemas armamentísticos supone igualmente grandes desafíos tanto a la seguridad global, como en el plano de la ética. La adquisición de LAWS por parte de estados autoritarios, totalitarios y/o actores no estatales supone un riesgo de primer orden para la vida humana y la estabilidad global. Es probable que a medio plazo se fomente una regulación internacional, para hacer frente a las dificultades que presenta esta tecnología en cuanto a atribución, depuración de responsabilidades y falta de control humano.

5.1. Qué son las LAWS:

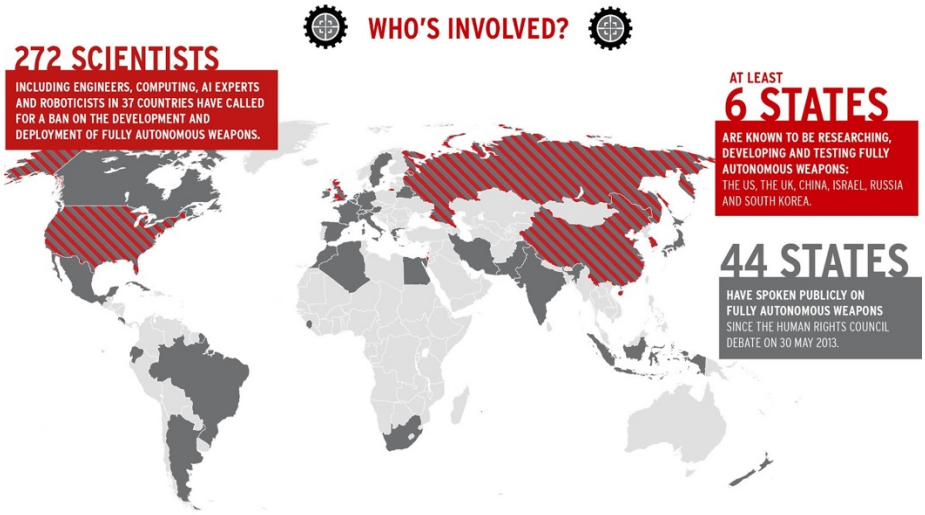
No existe una definición consensuada para los Sistemas de Armas Autónomos. El Departamento de Defensa de EEUU define, en su directiva 3000.09, los Sistemas Letales de Armas Autónomos (LAWS por sus siglas en inglés) como el sistema de armas que, una vez activado, puede seleccionar y atacar objetivos sin necesidad adicional de un operador humano.

En otras palabras, los LAWS están diseñados para tomar decisiones independientes que involucran el uso de fuerza letal. Para hacerlo, no requieren que ningún ser humano supervise el proceso (lo que se denomina “humano fuera del bucle” o “human out of the loop”), ya que se sirven de una inteligencia artificial para el procesado de gran cantidad de datos recogidos por sus sensores para identificar objetivos e incluso diseñar tácticas de grupo o enjambre.

Los sistemas de armas autónomos pueden ser clasificados, en primer lugar, por su grado de autonomía. Así, encontramos los siguientes tres sistemas: Los primeros son los sistemas tele-operados (como los drones Reaper o Predator) de uso generalizado en la actualidad, operados mediante control remoto por humanos. Les siguen los sistemas automatizados o semiautónomos (dron de vigilancia Global Hawk) con parámetros preprogramados que no requieren una orden o comando humano, aunque la monitorización humana e mantiene. Por último, encontramos los sistemas lenamente autónomos (Sistema de combate AEGIS) con el poder de decidir por sí mismos sobre sus operaciones, siendo incluso capaces de aprender gracias a la Inteligencia Artificial, y adaptarse a nuevas informaciones.

5.1.1. Países con tecnología LAWS:

Diversos sistemas clasificados como LAWS son producidos por múltiples estados en la actualidad. Sin embargo, EEUU, China, Rusia, Israel, Reino Unido y Corea del Sur, acompañados más recientemente por Turquía, son los estados que más invierten en la investigación y el desarrollo de esta tecnología en la actualidad.



Fuente: World Economic Forum & Campaign to Stop Killer Robots

5.1.2. Estado de la legislación:

Fue en 2013 cuando la comunidad internacional comenzó a considerar el asunto de los LAWS. Ese año, la ONU se hizo eco del asunto de los Sistemas Letales de Armas Autónomos, con una reunión informal de expertos, que tuvo lugar durante la Convención de Ciertas armas Convencionales (CCW por sus siglas en inglés). En 2016 se estableció un Grupo de Expertos Gubernamentales (GGE), que en 2017 se reunieron por primera vez para analizar el asunto en sus aspectos militar, tecnológico, legal y ético. La reunión terminó con poco consenso, pero se concluyó que los LAWS deberían estar sujetos plenamente al Derecho Internacional Humanitario (DHI).

En marzo de 2019, Antonio Guterres, Secretario General de la ONU, declaró que las armas automáticas con poder y discrecionalidad para seleccionar objetivos sin intervención humana son políticamente inaceptables, moralmente repugnantes y deberían ser prohibidos por el Derecho Internacional, y apremió al GGE a llegar a una solución acorde a la magnitud del desafío. Éstos publicaron un nuevo informe en septiembre 2019 en el que reafirmaban la necesidad de los LAWS de someterse al Derecho Internacional y sobre los peligros de la potencial adquisición de esta tecnología por parte de entidades terroristas.

En septiembre de 2020, el Instituto para la investigación del Desarme de la ONU (UNIDIR) publicó el informe “Black Box, Desbloqueada” en el que describen la predictibilidad y la comprensibilidad de la IA como factores clave para el desarrollo de los LAW.

Naciones con proyectos de armas autónomas en desarrollo han estado invirtiendo esfuerzos para evitar un veto o prohibición al desarrollo y despliegue de LAWS. Entre ellas encontramos a China, EE. UU., Israel, Rusia y Corea del Sur. A favor de la no prohibición, EE. UU. mantiene argumentos como que a más inteligentes se vuelvan las armas, mejor equipados estarán los contendientes para mantenerse en el Derecho Humanitario Internacional, yendo incluso más lejos al afirmar que no implementar el uso de IA en el campo de batalla, si funciona, podría considerarse una violación del DHI. Desarrollan sus argumentos en torno a la idea de los “robots ´éticos” que operasen de manera más ética que los soldados humanos en el campo de

batalla, reduciendo la cadena de responsabilidades a un único humano, el autor de la orden; la utilidad militar, por las posibilidades revolucionarias del desarrollo del armamento con IA, por ejemplo en cuanto a la posibilidad de acometer tácticas de enjambre.

Por parte de la UE, en 2021 se publicó el informe *Guidelines for military and non-military use of Artificial Intelligence 2021,* que se configura como un código ético para la regulación del uso de la IA en el ámbito militar y de los sistemas de armas autónomos y trata de limitar sus potenciales riesgos asociados.

5.1.3.Tipos de armas:

El desarrollo de sistemas de armas autónomos no es algo exclusivamente reciente. Ya en 1980, encontramos el sistema de cañón US Phalanx, precursor de los LAWS, y en lo referente a vehículos de combate aéreos no tripulados (UCAV por sus siglas en inglés), el Boeing X-45, desarrollado en 2005. En la actualidad, se considera que existen 25 tipos de sistemas, incluidos en 8 grandes categorías:

- *Loitering Munitions*: conocidos popularmente como "drones kamikaze", como por ejemplo los israelíes IAI Harop, que ya han sido empleados por Azerbaiyán en la guerra de Nagorno-Karabaj en 2020.
- *Unmanned Combat Aircraft:* Vehículos aéreos de combate no tripulados
- *Precision Guided Munitions:* Municiones guiadas
- *Unmanned Ground Vehicles:* Vehículos de combate terrestre autónomos
- *Unmanned Marine Vehicles:* Vehículos de combate acuáticos autónomos
- *Border Control:* Sistemas autónomos de control de fronteras.
- *Counter Terrorism and Law Enforcement:* Sistemas autónomos de lucha antiterrorista.
- *Antianimal:* por ejemplo, el COSTBOT, arma no militar australiana para proteger el gran arrecife de una especie de pez invasora.

5.1.4.Ventajas frente a otros sistemas:

Los defensores de los Sistemas Letales Autónomos señalan múltiples ventajas de su implementación. En el ámbito militar, destaca el potencial de ser más rápidas, mejores y más baratas que otros sistemas. Su éxito dependerá en buena medida del nivel de IA empleada. Sus costes suponen un tercio del empleado en sistemas con efectivos humanos.

Además, no están sujetos a las limitaciones de los sistemas operados por humanos en cuanto a peso y dimensiones, con lo que pueden ser mucho más pequeños e indetectables. Otra de sus principales ventajas es que excede sustancialmente en rango de alcance, en durabilidad y resistencia los actuales sistemas compuestos por efectivos humanos, y disminuye la huella logística.

Las máquinas dotadas de Inteligencia Artificial son además mucho más rápidas que los humanos en tomar decisiones basadas en grandes cantidades de datos e inteligencia. Al mismo tiempo, son indiferentes al entorno de operaciones y sus condiciones físicas, climáticas o de seguridad.

Asimismo, los LAWS están mucho menos expuestos a fallos de interpretación y errores cometidos por la necesidad de una respuesta rápida, superando ampliamente las capacidades para discernir objetivos que posee el ser humano. Esto llevaría a una potencial reducción de las víctimas colaterales en una situación de conflicto o que requiera el uso de la fuerza.

Igualmente, al no estar operadas por humanos, están exentas de emocionalidad, con lo cual quedarían fuera de la ecuación emociones como la ira o el odio que pudieran llevar a contendientes humanos a cometer matanzas, genocidios o crímenes de lesa humanidad, movidos por estos factores.

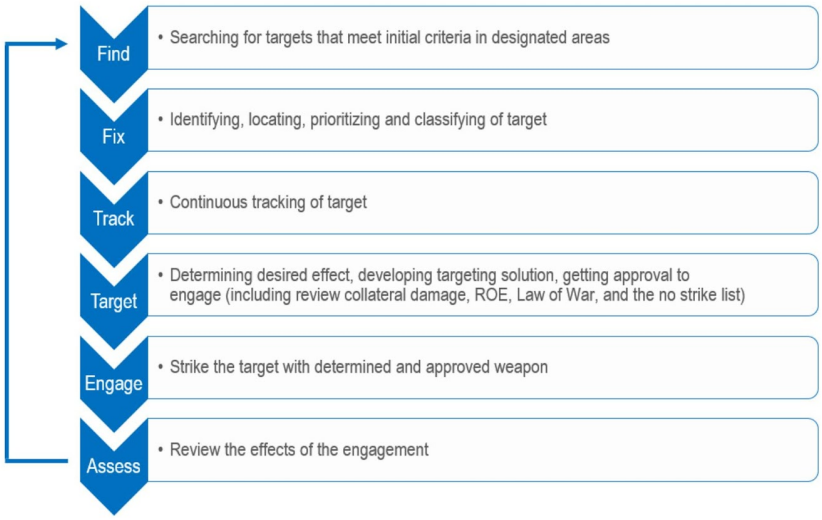


Figure 1: Dynamic Targeting Cycle

Proceso de selección de objetivo. Fuente: International panel for the regulation of Autonomous Weapons

5.1.5. Riesgos y desafíos:

Por parte de los detractores de estos sistemas, son múltiples los desafíos y peligros que su implementación no regulada podría conllevar a corto o medio plazo.

En primer lugar, los sistemas LAWS aumentarían enormemente la dificultad de atribución de un ataque, con lo que podría imbuir de una gran sensación de impunidad a determinados actores para llevar a cabo actos de enorme gravedad. El sentimiento de impunidad se vendría reforzado a causa de la facilidad para eludir responsabilidades por parte del estado u organización causante del ataque, al no haber una persona directamente involucrada en la decisión de quitar vidas, pues sería decidido por un algoritmo.

Otro de los serios desafíos que plantea sería una carrera armamentística a nivel mundial en la que ningún estado podría no invertir en el desarrollo o adquisición de estos sistemas, ya que supondría quedar a merced de estados, organizaciones o actores rivales.

El uso por estados no democráticos para controlar a su población, por parte de regímenes totalitarios, o estados fallidos para deshacerse de minorías étnicas, es otro de sus grandes riesgos. El potencial de esta tecnología para acabar con vidas humanas basándose en los parámetros que establezca el algoritmo que los dirigen, podría dotarles de autonomía para discernir e identificar objetivos en base a parámetros raciales, religiosos o de cualquier otra índole, la convierte potencialmente en una seria amenaza para la vida de las personas y la estabilidad global.

La posibilidad de adquisición y uso por parte de actores no estatales y organizaciones terroristas supone, en potencia, otro de los grandes desafíos a la seguridad mundial. La adquisición y manejo de estos sistemas podría permitir que actores relativamente pequeños pudieran acometer ingentes daños de diversa índole.

Asimismo, al estar estos sistemas dotados de una inteligencia artificial e interconectados, sería técnicamente posible que distintos actores los manipulasen mediante ciberataques, bien neutralizando las defensas de un país que se fundamenten en este sistema, o incluso volviéndolas contra el propio país u organización.

La carrera armamentística que se avecina a partir del aumento en la implantación y desarrollo de esta tecnología repercutirá igualmente, de manera negativa e impredecible, en la

seguridad global, pudiendo alterar el balance de poderes al permitir, teóricamente, que estados o actores más débiles acaben adquiriendo ventajas competitivas frente a otros actores que no tengan esta tecnología tan implementada.

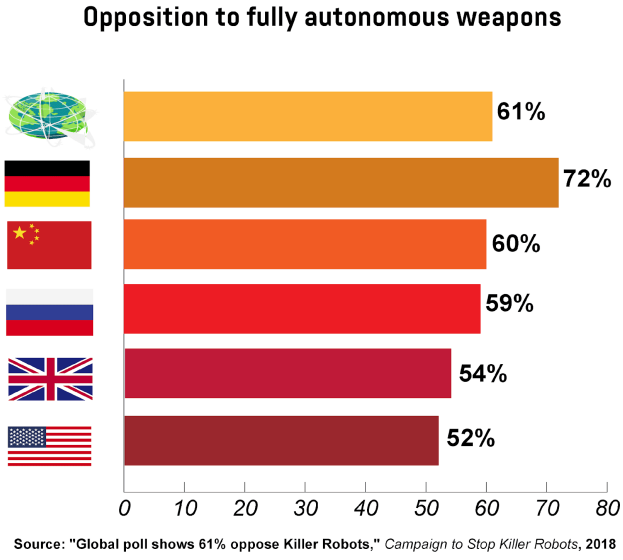
5.1.6. Implicaciones éticas y legales:

Los Sistemas de Armas Autónomos plantean una serie de desafíos desde el punto de vista ético y humanitario:

En cuanto al Derecho Internacional Humanitario, los LAWS levantan dudas en torno a distintos principios. En cuanto al principio de Distinción, se considera que, en el momento presente, es difícil para una IA discernir con eficacia entre objetivos civiles y militares, y dentro de estos últimos, entre combatientes activos y efectivos heridos. Esto entraría en conflicto con el artículo 51 de la Convención de Ginebra. Otro de los principios con los que colisiona es el de proporcionalidad (artículos 51 y 57 de la Convención de Ginebra), que versa sobre la prevención de ataques indiscriminados. La implementación y obligación del uso de Control Humano Consciente e IA Explicable en los sistemas LAWS podría mitigar estos puntos.

En cuanto al Derecho Penal Internacional, los LAWS plantean un serio desafío a la hora de establecer la autoría y la responsabilidad, siendo este uno de los principales retos. No existe un criterio unificado en la actualidad a la hora de depurar responsabilidades en cuanto a potenciales víctimas producidas por un ataque de estos sistemas, ya que, una vez desplegados, tienen la capacidad y el poder para decidir por sí mismos, y sin control humano, sobre la vida o la muerte de personas.

Varios estados encabezan la oposición pública al desarrollo e implementación de estos sistemas sin una regulación estricta, encabezando Alemania la lista los estados occidentales.



Fuente: Campaign to Stop Killer Robots

5.1.7. Perspectivas de Futuro LAWS:

El desarrollo de la IA en el ámbito militar y armamentístico es una realidad que no va a desaparecer, y supone toda una revolución en la manera de hacer la guerra, comparable a la aparición de la pólvora y a la de las armas nucleares. Este cambio de paradigma no sólo afectará a la vertiente militar, sino por consiguiente también tendrá gran impacto en los ámbitos político y social a nivel global. Su implementación será cada vez mayor e incluso los estados que en el momento presente no hacen uso de esta tecnología, tendrán que hacerse con ella para poder defenderse frente a los actores que sí la poseen. Esto conllevará una intensificación en la carrera armamentística que dirigentes, como el ministro de exteriores alemán, alertan que ya está

teniendo lugar. Los propios EEUU justifican su necesidad de desarrollo en este campo ante el avance de China y Rusia en la implementación de la Inteligencia Artificial en el sector militar.

Se prevé que en dos décadas habrá enjambres de sistemas autónomos en tierra, mar y aire capaces de operar coordinadamente de manera absolutamente independiente basándose en inteligencia artificial. Países como Turquía, Israel, Reino Unido, Estados Unidos y otras naciones ya están desarrollando tecnologías de sistemas de armas autónomas que incorporan este tipo de tácticas. La rapidez con que estos sistemas reaccionarán entre sí dejará al margen la capacidad de control humano, que sólo mediante sistemas similares podrá hacerle frente.

Los Sistemas de armas autónomos, dotados de inteligencias artificiales capaces de procesar ingentes cantidades de datos y actuar en consecuencia en cuestión de milisegundos, podrían trasladar la guerra a unas escalas de velocidad inabarcables para el ser humano. Así, cuando estos sistemas se enfrenten entre sí, la velocidad de sus interacciones podría llevar a reacciones en cadena ultrarrápidas y de efectos y magnitud impredecibles, pudiendo llegar a desencadenar guerras relámpago entre sistemas (flash wars).

Otro de los cambios más significativos en la manera de hacer la guerra, es previsible que venga de la combinación y coordinación sin necesidad de intervención humana entre sistemas de armas autónomos y herramientas de ciberguerra con capacidad de autorreplicarse.

Ante la ingente serie de desafíos que se avecinan con la implementación y generalización de esta tecnología, es previsible que desde Naciones Unidas se termine regulando su uso y desarrollo de manera global. Es probable que determinados estados sigan manifestando su reticencia a una regulación global, lo que podría acabar traduciéndose en algún tipo de sanciones hacia ellas, llegado el caso.

6. El uso de la inteligencia artificial y la radicalización

La radicalización es un proceso que, históricamente, solía desencadenarse principalmente a través de interacciones sociales en lugares de culto, escuelas religiosas, prisiones, etc. Hoy en día, con el desarrollo de las nuevas tecnologías digitales, este proceso se inicia a menudo en Internet, donde los contenidos de radicalización se comparten fácilmente y tienen un mayor alcance más rápido y con mayor facilidad.

Se podría hablar de dos "usos" de la IA en el proceso de radicalización: el indirecto y el directo. Con respecto al uso indirecto, se encuentran los algoritmos empleados en los sistemas de recomendación de vídeos (como el de YouTube). La opacidad en la que estos algoritmos sugieren nuevos elementos en las redes sociales y en los sitios web de música, vídeo o cine, no da al consumidor ninguna preferencia sobre cómo se pueden recomendar el contenido. Según un informe del 2019, los algoritmos de YouTube contribuyen a reforzar las opiniones extremistas: una vez que se ha visto un vídeo con contenido extremista, se recomiendan contenidos similares al usuario.

En cuanto al uso directo, se encuentra el Generative Pre-trained Transformer 3 (mejor conocido por sus siglas GPT-3), un nuevo modelo de inteligencia artificial que permite generar lenguaje escrito. Gracias a este algoritmo, el usuario solo necesita escribir un párrafo para que el sistema se encargue de completar el resto de forma más o menos coherente e, incluso, es posible que el mismo algoritmo genere contenido de forma automática. Así pues, gracias a un entrenamiento previo —mediante textos de Wikipedia e internet—, este proyecto permite producir textos "falsos" con una base ideológicamente coherente, sin la necesidad de un gran corpus y horas de preajustes. El Center on Terrorism, Extremism and Counterterrorism (CTEC) de la Universidad de Middleburry fue quien se encargo de estudiar el mal uso que se le podría dar al GPT-3 y, entre los experimentos y análisis que llevaron a cabo, se demostró que se puede incitar al GPT-3 a realizar diversas tareas como reproducir hilos de foros falsos donde se discute el genocidio y se promueve el nazismo o producir textos extremistas multilingües (como contenido antisemita en ruso), incluso cuando las "indicaciones" están en inglés. El principal riesgo del GPT-3 se encuentra relacionado con su mal uso y con su capacidad de influencia que parece humana y que es muy difícil poder diferenciar el contenido producido por este algoritmo de un texto producido por una persona. En consecuencia, el mal uso de este algoritmo podría "alivianar" la carga de trabajo de los grupos extremistas al facilitarle el trabajo de producción de textos en foros informativos e interactivos.

Actualmente, el GPT-3 se encuentra disponible a través de una API, por lo que para poder acceder a él es necesario una invitación de OpenAI, los desarrolladores de los algoritmos. Según su página web, ellos mismos cancelarán el acceso a la API en caso de que se vaya a utilizar de forma perjudicial (como acoso, spam, radicalización, etc.), sin embargo, son conscientes de que no pueden anticipar todas las consecuencias de esta tecnología, por lo que, de momento, el GPT-3 se encuentra disponible en su versión beta y privada en lugar de encontrarse disponible al público en general. Así pues, en caso de que se cree un modelo que se base en el GPT-3 y lo imite —y no utilice ningún tipo de filtro o resguardo—, aquellos grupos con intenciones de "extender" su pensamiento radical a través de textos se beneficiaría, ya que la capacidad del GPT-3 de emular un entorno ideológicamente coherente e interactivo de las comunidades extremistas en línea plantea el riesgo de amplificar los movimientos extremistas que buscan radicalizar y reclutar individuos. De esta forma, los extremistas podrían producir fácilmente textos sintéticos y luego emplear la automatización para acelerar la difusión de este contenido fuertemente ideológico y emotivos en los foros, donde sería difícil distinguirlo de los contenidos generados por personas.

7. Los semiconductores y la Inteligencia Artificial

El marco de trabajo de la inteligencia artificial (IA) puede dividirse a grandes rasgos en tres capas: en primer lugar, la infraestructura (incluye los chips de IA y el *big data*), que apoya las capacidades de detección y cálculo cognitivo computacional de la capa tecnológica; en segundo lugar, el nivel de aplicación, es aquel que es "más superficial" ya que proporciona los servicios que utilizan a la IA como la conducción autónoma y la asistencia virtual; en tercer y último lugar, los chips de IA, que se encuentran en el "corazón" de la cadena tecnológica y que son fundamentales para el procesamiento de los algoritmos, especialmente para las redes neuronales profundas (DNN, por sus siglas en inglés). Así pues, con las aplicaciones de la inteligencia artificial (IA) cobrando impulso en diversas áreas, la demanda por sensores especializados, circuitos integrados, memoria y procesadores mejorados se encuentra en aumento.

Con respecto a las DDN, a lo largo de los años, la complejidad entre capas y números de nodos del modelo de red neuronal ha crecido exponencialmente, lo que supone un reto importante para la computación. Las unidades centrales de procesamiento tradicionales (CPU), se destacan en las cargas de trabajo generales —especialmente en aquellas basadas en reglas—, sin embargo, las CPU ya no pueden seguir el ritmo del paralelismo que requieren los algoritmos de IA. Hasta ahora, existen dos formas de hacer frente a este problema: por un lado, se puede añadir un acelerador basado en la arquitectura computacional existente (que es en donde entran en juego los semiconductores) y, por otro lado, se podría volver a desarrollar por completo esta arquitectura que simula las redes neuronales del cerebro humano (un enfoque que se encuentra en su fase inicial de desarrollo).

Así pues, como se mencionó previamente, para construir los aceleradores se necesita a los semiconductores. Estos son facilitadores tecnológicos esenciales que alimentan muchos de los dispositivos digitales de última generación que se utilizan hoy en día y, dentro del mundo de la IA, su mayor mercado se encuentra en la nube, ya que su adopción en los centros de datos sigue aumentando como medio para mejorar la eficiencia y reducir los costes operativos. Por otra parte, en el *deep learning*, la IA utiliza al *big data* como base para "entrenar" a los DNN con el fin de desarrollar la capacidad de "inferir" a partir de un conjunto de datos; esta fase de entrenamiento requiere una enorme cantidad de potencia de cálculo y esto requiere de servidores de alta gama que tengan un rendimiento paralelo avanzado para poder procesar el conjunto de datos, por lo tanto, aquí también son indispensables los aceleradores.

Con respecto a los fabricantes de semiconductores, la empresa surcoreana Samsung es la líder mundial en volumen de ventas y, entre las diez primeras hay seis estadounidenses —como Intel o Qualcomm—, y otras de Corea del Sur, Taiwán y Japón y para su producción, uno de los materiales más utilizados es el silicio (preferido por su precio y ventaja, aunque existen otros como el germanio y azufre). A pesar de que China no posee ninguna de las mayores empresas productoras de semiconductores, sí que es el primer productor de silicio (5,4 millones de toneladas anuales, correspondiente al 70% de la producción total) (El Orden Mundial, 2021).

Los principales actores del sector de los semiconductores se encuentran en Asia Oriental (China continental, Japón, Corea del Sur y Taiwán). La región se ha convertido en un punto caliente para la industria de los semiconductores debido a su floreciente economía, el aumento de las comunicaciones móviles y el crecimiento de la computación en nube y la importancia de estos semiconductores ha hecho de la producción del silicio y el resto de estos materiales un campo de competición geopolítica.

En el caso de la Unión Europea (UE), esta carece de actores relevantes en la cadena de producción de los semiconductores y, en la última hoja de ruta digital de la UE, la Brújula Digital, se le presta atención al papel de los semiconductores, sin embargo, con unos objetivos vagos y

cuestionables desde el punto de vista tecnológico. Mientras la UE se está poniendo al día, EE. UU. ya se encuentra formulando el siguiente paso en su camino para reducir la dependencia de la cadena de suministro de China mediante una mayor disociación del proceso de fabricación, por lo que la UE tendrá que decidir si quiere unirse al esfuerzo estadounidense o decide trazar su propio camino independiente. Así pues, la Unión Europea quiere duplicar su producción de chips hasta alcanzar el 20% del mercado mundial en 2030. Este objetivo pretende impulsar la "soberanía digital" mediante la financiación de diversas iniciativas de alta tecnología, disminuyendo el nivel de dependencia de países orientales y también no verse afectada por la guerra tecnológica entre China y Estados Unidos.

Con respecto a la producción de los semiconductores, esta industria funciona bajo un modelo de "justo a tiempo", en el que los semiconductores se producen según demanda. Este tipo de producción conlleva a que la industria se vea afectada periódicamente por la escasez provocada por catástrofes naturales, acontecimientos de origen humano o, como en el último caso, por la pandemia causada por la COVID-19.

Con la primera oleada a nivel mundial de casos de COVID, la escasez de semiconductores se vio agravada por dos factores: en primer lugar, la industria se volcó a producir chips para las portátiles y móviles; en segundo lugar, hubo una "doble reserva" de los compradores de chips, ansiosos por asegurarse el inventario, lo que a su vez fue consecuencia de la incertidumbre en la cadena de suministro creada por la guerra tecnológica entre EE. UU y China. Nadie cree que la escasez termine este año y muchos esperan que se prolongue hasta el 2022 y, según el director general de Intel, Pat Gelsinger, se espera que la escasez dure un par de años.

8. CONCLUSIONES

Se ha demostrado que los sistemas de IA cuentan con capacidad para vulnerar los derechos a la protección de datos personales y de la privacidad y la no discriminación. Debido a que esta tecnología puede conllevar un riesgo para la seguridad y el funcionamiento eficaz del régimen de responsabilidad civil, la Comisión Europea ha comenzado a elaborar un Reglamento sobre la IA, planteando incluso la creación de un Comité Europeo de Inteligencia Artificial, que se encargaría de garantizar el funcionamiento adecuado del mercado de los sistemas que utilicen IA a fin de que haya confianza en el desarrollo y la adopción de este tipo de tecnología.

La transformación y el impacto de la IA en el mercado laboral presentan una serie de oportunidades para las personas físicas y jurídicas que, una vez han sido identificadas, tienen que ser puestas en acción sin más demora. Los países que sean capaces de adaptarse de manera ágil en el ámbito educativo para poder formar futuros trabajadores que tengan las habilidades que son demandadas por las empresas darán un salto cualitativo como sociedad y se producirá un incremento significativo en el nivel de bienestar personal, laboral y global.

La formación continúa en pensamiento computacional y la adquisición de habilidades en aprendizaje automático y modelación, *cloud computing*, gestión de Big Data, programación y robótica son ya imprescindibles en la actualidad, pero además es indispensable mantener una monitorización permanente de las necesidades que tienen las empresas y configurar un sistema de formación y especialización continua con una fuerte conectividad con el mercado laboral.

La alternativa a la flexibilidad y formación continua es la pérdida progresiva de empleos en virtud de su desaparición o de la reestructuración de tareas y modalidades de desempeño, lo que implicaría un fracaso colectivo que podría poner en peligro el equilibrio social e incluso la propia supervivencia de esos países.

Por otra parte, la verificación precisa de información a través de un sistema de inteligencia artificial para combatir la desinformación continúa siendo un objetivo lejano dado que, tanto los mecanismos de verificación automática a través de la IA, como la verificación manual tienen aún sus limitaciones. Por un lado, la revisión humana de la información que recibimos a través de fuentes abiertas no es eficiente, puesto que estamos cada vez más expuestos a mayores cantidades de información. A su vez, el filtro de tal cantidad de datos requeriría de un gran número de personas dedicadas a realizar las verificaciones y esto resultaría muy costoso. Por otro lado, aunque la tecnología es capaz de procesar y verificar automáticamente grandes cantidades de información al mismo tiempo, la IA, al igual que la revisión manual, también incluye sesgos cognitivos que contribuyen a la formación de la opinión pública. No obstante, la IA tiene un gran potencial como un complemento más rápido y rentable a la supervisión humana, pero la supervisión humana es necesaria hasta que la IA pueda identificar con precisión la desinformación en diferentes contextos lingüísticos, culturales y políticos.

La implementación de sistemas LAWS supondrá muy probablemente todo un cambio de paradigma a nivel global en el modo de hacer la guerra. Las elevadas posibilidades inéditas y el grado de eficiencia y eficacia que ofrece con respecto a las unidades de efectivos humanos, harán de su implantación una revolución similar a la del uso de la pólvora o la aparición del arma atómica, de efectos y derivaciones aún por esclarecer.

Los LAWS seguirán evolucionando y pronto su uso será generalizado. Ante las múltiples ventajas competitivas que ofrecen a quienes las poseen frente a quienes no, es muy probable que se dé un escenario de una nueva carrera armamentística entre estados, con elevado potencial de suponer un nuevo desafío disruptivo a la seguridad y a la estabilidad global. En última instancia, este escenario podría suponer una reconfiguración en el equilibrio de poderes actual en el plano militar, condicionado por el grado de inversión en esta tecnología.

Es probable que a medio plazo se legisle hacia un control humano responsable al otro lado de las acciones de los sistemas. Es muy probable que desde la ONU se legisle para evitar el desarrollo de estos sistemas hacia un punto que les permitiera actuar con total impunidad y no atribución. Sin embargo, es muy probable que las principales potencias ligadas al desarrollo de esta tecnología ralenticen, o detengan, la implantación de una regulación global que pueda repercutir negativamente en el aumento de su ventaja estratégica y militar, aún en detrimento de la seguridad global y del Derecho Internacional Humanitario.

Ante los múltiples desafíos éticos que plantea un sistema con potencial de regir sobre la vida de las personas de manera autónoma, la falta de una regulación a nivel internacional que impida la elusión de responsabilidades, los riesgos ante la dificultad de atribución, y el potencial de devastación en manos de determinados actores, supone en sí misma un riesgo global de magnitud comparable al propio desarrollo de esta tecnología bélica.

La adquisición por parte de estados con regímenes autoritarios, totalitarios, inmersos en luchas étnicas supone un serio desafío en el medio plazo a la seguridad de grupos enteros de población, así como para la seguridad global. Regímenes autoritarios como Turquía ya operaron con éxito drones letales autónomos, por ejemplo, en la guerra de Libia, y en el Nagorno Karabaj por parte de Azerbaiyán.

Ante un elevado grado de desarrollo de la tecnología LAWS, y una ausencia de regulación internacional, la adquisición de estos sistemas por parte de actores no estatales, como grupos armados y organizaciones terroristas, supone un riesgo real de primer orden, con altas probabilidades de materializarse. Esto plantearía un desafío a la seguridad global de magnitud y consecuencias impredecibles a nivel mundial, en caso de no ser prevenido convenientemente por parte de la Comunidad Internacional.

La aplicabilidad de la IA a la radicalización plantea ciertos interrogantes. Por un lado, uno de ellos es si habría que “censurar” mediante leyes cierto tipo de contenido que sirve para “nutrir” estas tecnologías y, en caso de que caigan en manos equivocadas, podrían hacer más mal que bien. En este punto también entraría una cuestión filosófica sobre dónde está el límite de aquello considerado como “extremismo” y aquello que no. Además, esta tecnología podría utilizar textos con fines académicos, educativos o de investigación para nutrirse. Por otro lado, está el interrogante de si las autoridades debiesen regular el uso, venta y adquisición de APIs como la del GPT-3 para que estas no deban ser comercializadas libremente a cualquier tipo de actores. En este punto también entraría en conflicto la libertad de cada persona de poder desarrollar la tecnología que quiera, así como también el hecho de qué uso quieran darles los gobiernos a estas tecnologías.

Por último, los semiconductores son indispensables para el desarrollo de la tecnología, sin embargo, la producción mundial se encuentra reducida a un par de actores, la mayoría ubicados en Asia oriental. Además, a pesar de que China no es un productor de semiconductores, sí que acapara el 70% de la producción mundial de silicio, por lo que, si en un futuro decide no vender este material a otros países para dedicarlo a su producción nacional, los otros países no podrán desarrollar ni avanzar en el campo de la IA. Ahora bien, con respecto a Europa, aún se está por ver si logra cumplir sus objetivos planteados para el 2030 y, de forma que no pueda aumentar su producción de semiconductores, seguirá dependiendo de terceros para poder abastecerse.

9. Bibliografía

McGuffie, K., & Newhouse, A. (2020). The Radicalization Risks of GPT-3 and Advanced Neural Language Models. Middlebury Institute of International Studies, Center on Terrorism, Extremism, and Counterterrorism, , Monterey.

OpenAI API. (s.f.). Obtenido de OpenAI API: https://openai.com/blog/openai-api/

Schroeter, M. (2020). Artificial Intelligence and Countering Violent Extremism: A Primer. London: Report-Gnet.

Amorós, M., & Évole, J. (2018). Fake News: La verdad de las noticias falsas. Barcelona: Plataforma Editorial.

Asociación de la Prensa de Madrid. (2019). Obtenido de https://www.apmadrid.es/comunicado/informe-de-la-profesion-periodistica-2019-aumenta-un-26-el-paro-de-los-periodistas-tras-6-anos-de-descensos/

BBC Mundo. (10 de diciembre de 2016). Rusia "intervino en las elecciones para promover la victoria de Donald Trump", dicen agencias de inteligencia de EE. UU.

Comisión Europea. (2021). Obtenido de https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation

European Commission . (2018). *A multi-dimensional approach to disinformation .* Belgium.

GAD3; Fundación AXA . (s.f.). *Foro Periodismo 2030.* Castilla y León .

GAD3; Fundación AXA. (2021). *Foro Periodismo 2030.*

Gartner. (3 de October de 2017). Obtenido de https://www.gartner.com/en/newsroom/press-releases/2017-10-03-gartner-reveals-top-predictions-for-it-organizations-and-users-in-2018-and-beyond

Lógica, S., & UCM. (2017). *"I Estudio sobre el Impacto de las Fake News en España"*. Obtenido de https://www.uoc.edu/portal/es/news/actualitat/2019/319-fakenews-comision-nacional.html

Pardo , R. S., & Pardo, J. (2018). *LA influencia del fenómeno "fake news" en la comunicación organizacional. La innovación de la innovación: del medio al contenido predictivo. Actas III Simposio Internacional sobre gestión de la comunicación.* Obtenido de https://xescom2018.wordpress.com/libro-de-acta/

Rodríguez-Fernández, L. (2019). Desinformación y comunicación organizacional: estudio sobre el impacto de las Fake News . *Revista Latina de Comunicación Social* , 1714-1728.

Vivar, J. M. (2019). *"Las fake news siempre han existido, pero hoy en día se han visto catapultadas por las redes sociales".* Madrid: Doxa Comunicación.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science Journals*, 359.

Stencel, M., Luther, J. (2020). Annual census finds nearly 300 fact-checking project around the word. https://reporterslab.org/annual-census-finds-nearly-300-fact-checking-projects-around-the-world/

Cassauwers, T. (2019) Can artificial intelligence help end fake news?. European Commission. https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news

"A legal perspective: Autonomous weapon systems under international humanitarian law" Neil Davison, Scientific and Policy Adviser, Arms Unit, Legal Division International Committee of the Red Cross.

Views of the ICRC on autonomous weapon systems, paper submitted to the Convention on Certain Conventional Weapons Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) , Comité Internacional de la Cruz Roja.

Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems, Servicio de Investigación del Congreso de los EEUU.

Autonomous Weapon Systems and International Crises Author(s): Nathan Leys Source: Strategic Studies Quarterly.

The Future of Wars: Artificial Intelligence (AI) and Lethal Autonomous Weapon Systems (LAWS). Autores: Stephanie Mae Pedron Georgia Southern University & Jose de Arimateia da Cruz Georgia Southern University & US Army War College.

Legal and Policy Implications of Autonomous Weapons Systems. Anoushka Soni, Elizabeth Dominic, The Centre for Internet and Society, India.

Military Applications of Artificial Intelligene. Forrester E. Morgan, Benjamin Burdreaus, Andrew J. Lohn, Mark Ashby, Christian Curriden, Kelly Klima, Derec Grossman: Rand Corporation.

McGuffie, K., & Newhouse, A. (2020). The Radicalization Risks of GPT-3 and Advanced Neural Language Models. Middlebury Institute of International Studies, Center on Terrorism, Extremism, and Counterterrorism, , Monterey.

OpenAI API. (s.f.). Obtenido de OpenAI API: https://openai.com/blog/openai-api/

Schroeter, M. (2020). Artificial Intelligence and Countering Violent Extremism: A Primer. London: Report-Gnet.

Deloitte. (2019). Semiconductors: the Next Wave Opportunities and winning strategies for semiconductor companies. Obtenido de El Orden Mundial. (2021). El Orden Mundial. Obtenido de ¿Qué son los semiconductores?: https://elordenmundial.com/que-son-los-semiconductores/

European Commission, Directorate-General for Communications Networks, Content and Technology. (2021). 2030 Digital Compass: the European way for the Digital Decade. Brussels.

King, I. (2021). Intel CEO Says Chip Shortage to Hit Bottom in Second Half. Bloomberg.

McKinsey & Company. (2019). McKinsey on Semiconductors: Creating value, pursuing innovation, and optimizing operations. McKinsey Practice Publications.

SCMP Reporters. (2021). Why there is a global semiconductor shortage, how it started, who it is hurting, and how long it could last. South China Morning Post.

Yishan, C., & Sundstrom, G. (2021). Global Chip Shortage: The Winners and Losers.