COMISIÓN
EUROPEA

Bruselas, 8.4.2019
COM(2019) 168 final

COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES

Generar confianza en la inteligencia artificial centrada en el ser humano

1. INTRODUCCIÓN — ESTRATEGIA EUROPEA DE LA IA

La inteligencia artificial (IA) tiene potencial para transformar nuestro mundo para mejor: puede mejorar la asistencia sanitaria, reducir el consumo de energía, hacer que los vehículos sean más seguros y permitir a los agricultores utilizar el agua y los recursos de forma más eficiente. La IA puede utilizarse para predecir el cambio climático y medioambiental, mejorar la gestión del riesgo financiero y proporcionar las herramientas para fabricar, con menos residuos, productos a la medida de nuestras necesidades. La IA también puede ayudar a detectar el fraude y las amenazas de ciberseguridad y permite a los organismos encargados de hacer cumplir la ley luchar contra la delincuencia con más eficacia.

La IA puede beneficiar a la sociedad y a la economía en su conjunto. Es una tecnología estratégica que se está desarrollando y utilizando a buen ritmo en todo el mundo. No obstante, también trae consigo nuevos retos para el futuro del trabajo y plantea cuestiones jurídicas y éticas.

Para abordar estos retos y aprovechar al máximo las oportunidades que ofrece la IA, en abril de 2018 la Comisión publicó una estrategia europea[1]**.** La estrategia coloca a la persona en el centro del desarrollo de la IA — **es una IA centrada en el ser humano**. Adopta un planteamiento triple para potenciar la capacidad tecnológica e industrial de la UE e impulsar la adopción de la IA en todos los ámbitos de la economía, prepararse para las transformaciones socioeconómicas y garantizar el establecimiento de un marco ético y jurídico apropiado.

Para concretizar la estrategia en materia de IA, **la Comisión desarrolló junto con los Estados miembros un plan coordinado sobre la inteligencia artificial**[2], que presentó en diciembre de 2018, para crear sinergias, reunir datos —la materia prima de numerosas aplicaciones de IA— e incrementar las inversiones conjuntas. El objetivo es fomentar la cooperación transfronteriza y movilizar a todos los agentes con el fin de aumentar las inversiones públicas y privadas hasta **un mínimo de 20 000 millones EUR** anuales durante la próxima década[3]. La Comisión ha duplicado sus inversiones en IA en Horizonte 2020 y tiene previsto invertir cada año 1 000 millones EUR de Horizonte Europa y del programa Europa Digital, especialmente para espacios comunes de datos en salud, transporte y fabricación, y grandes instalaciones de experimentación, como hospitales inteligentes e infraestructuras para vehículos automatizados y una agenda de investigación estratégica.

Para implementar esta agenda estratégica común de investigación, innovación y despliegue, la Comisión ha intensificado su **diálogo con todas las partes interesadas relevantes** de la industria, institutos de investigación y autoridades públicas. El nuevo programa Europa Digital será también determinante para contribuir a que la IA esté a disposición de las pequeñas y medianas empresas en todos los Estados miembros a través de polos de innovación digital, instalaciones de ensayo y experimentación reforzadas, espacios de datos y programas de formación.

Sobre la base de su reputación de productos seguros y de calidad, el enfoque ético de Europa con respecto a la IA refuerza la confianza de los ciudadanos en el desarrollo digital y pretende generar una ventaja competitiva para las empresas europeas de IA. El objetivo de la presente Comunicación es poner en marcha una fase piloto global en la que participen las partes

---

1 COM(2018) 237.

2 COM(2018) 795.

3 Para contribuir a alcanzar este objetivo, la Comisión propuso, en el próximo período de programación 2021-2027, que la Unión asigne al menos 1 000 millones EUR anuales de los fondos delos programas Horizonte Europa y Europa Digital para invertir en IA.

interesadas a la escala más amplia posible con el fin de ensayar la implementación práctica de la orientación ética para el desarrollo y el uso de la IA.

2. GENERAR CONFIANZA EN LA IA CENTRADA EN EL SER HUMANO

La Estrategia europea de IA y el plan coordinado dejan claro que **la confianza es un requisito previo para garantizar un enfoque de la IA centrado en el ser humano**: la IA no es un fin en sí mismo, sino un medio que debe servir a las personas con el objetivo último de aumentar su bienestar. Para ello, **la fiabilidad de la IA debe estar garantizada**. Los valores en los que se basan nuestras sociedades han de estar plenamente integrados en la evolución de la IA.

La Unión se fundamenta en **los valores de respeto de la dignidad humana, la libertad, la democracia, la igualdad, el Estado de Derecho y el respeto de los derechos humanos**, incluidos los derechos de las personas pertenecientes a minorías[4]. Estos valores son comunes a las sociedades de todos los Estados miembros, en las que prevalecen el pluralismo, la no discriminación, la tolerancia, la justicia, la solidaridad y la igualdad. Además, la **Carta de los Derechos Fundamentales de la UE** reúne, en un único texto, los derechos individuales, civiles, políticos, económicos y sociales de que gozan los ciudadanos de la UE.

La UE se asienta sobre un **sólido marco normativo**, que constituirá la referencia mundial para la IA centrada en el ser humano. El Reglamento general de protección de datos garantiza un elevado nivel de protección de los datos personales y requiere la implementación de medidas que garanticen la protección de datos desde la fase de diseño y por defecto[5]. El Reglamento relativo a la libre circulación de datos no personales suprime barreras a la libre circulación de este tipo de datos y garantiza el tratamiento de todas las categorías de datos en cualquier lugar de Europa. El recientemente adoptado Reglamento de Ciberseguridad contribuirá a reforzar la confianza en el mundo digital; el Reglamento sobre la privacidad y las comunicaciones electrónicas propuesto[6] persigue también este mismo objetivo.

No obstante, la IA conlleva nuevos retos, ya que permite a las máquinas «aprender» y tomar decisiones y ejecutarlas sin intervención humana. No falta mucho para que este tipo de funcionalidad sea lo habitual en muchos tipos de bienes y servicios, desde los teléfonos inteligentes hasta los vehículos automatizados, los robots y las aplicaciones en línea. Ahora bien, las decisiones adoptadas mediante algoritmos pueden dar datos incompletos y, por tanto, no fiables, que pueden ser manipulados por ciberataques, pueden ser sesgados o simplemente estar equivocados. Aplicar de forma irreflexiva la tecnología a medida que se desarrolla produciría, por tanto, resultados problemáticos, así como la renuencia de los ciudadanos a aceptarla o utilizarla.

La tecnología de IA debería, más bien, desarrollarse de manera que las personas sean su centro y se gane así la confianza del público. Esto implica que las aplicaciones de IA no solo

[4] La UE ha ratificado así mismo la Convención de las Naciones Unidas sobre los derechos de las personas con discapacidad.

[5] Reglamento (UE) 2016/679. El Reglamento general de protección de datos (RGPD) garantiza la libre circulación de datos personales dentro de la Unión. Contiene disposiciones sobre la adopción de decisiones basada únicamente en el tratamiento automatizado, lo que abarca la elaboración de perfiles. Las personas afectadas tienen derecho a ser informadas de la existencia de toma de decisiones automatizada y a recibir información significativa sobre la lógica aplicada en la misma, así como sobre la importancia y las consecuencias previstas de este tratamiento para ellas. En tales casos, también tienen derecho a obtener intervención humana, a expresar su punto de vista y a impugnar la decisión.

[6] COM(2017) 10.

deben ajustarse a la ley, sino también respetar unos principios éticos y garantizar que su implementación evite daños involuntarios. En cada una de las fases de desarrollo de la IA debe estar garantizada la diversidad en cuanto al género, el origen racial o étnico, la religión o las creencias, la discapacidad y la edad. Las aplicaciones de IA deben empoderar a los ciudadanos y respetar sus derechos fundamentales. Su objetivo debe ser mejorar las capacidades de las personas, no sustituirlas, y permitir también el acceso de las personas con discapacidad.

Por consiguiente, son necesarias unas **directrices éticas** que se basen en el marco regulador existente y que sean aplicadas por desarrolladores, proveedores y usuarios de la IA en el mercado interior, estableciendo unas condiciones de competencia éticas en todos los Estados miembros. Por esta razón, la Comisión ha creado un **grupo de expertos de alto nivel sobre la IA**[7] que representa a toda una serie de partes interesadas, al que ha encomendado la elaboración de unas directrices éticas en materia de IA, así como la preparación de una serie de recomendaciones para una política más amplia en este ámbito. Al mismo tiempo, se ha creado la **Alianza europea de la IA**[8], una plataforma multilateral abierta con más de 2 700 miembros, para aportar una contribución más amplia a la labor del grupo de expertos de alto nivel sobre la IA.

El grupo de expertos de alto nivel sobre la IA publicó un primer borrador de las directrices éticas en diciembre de 2018. Tras una **consulta a las partes interesadas**[9] y **reuniones con representantes de los Estados miembros**[10], el grupo de expertos sobre la IA presentó un documento revisado a la Comisión en marzo de 2019. En sus reacciones hasta la fecha, los interesados en general se han mostrado satisfechos con la naturaleza práctica de las directrices y la orientación concreta que ofrecen a desarrolladores, proveedores y usuarios de la IA sobre cómo garantizar la fiabilidad.

2.1. Directrices para una IA fiable elaboradas por el grupo de expertos de alto nivel sobre la IA

Las directrices elaboradas por el grupo de expertos de alto nivel sobre la IA, a las que se refiere la presente Comunicación[11], se basan en particular en el trabajo realizado por el Grupo europeo de ética de la ciencia y de las nuevas tecnologías y la Agencia de los Derechos Fundamentales.

Las directrices propugnan que, para lograr una «IA fiable», son necesarios tres componentes: 1) debe ser conforme a la ley, 2) debe respetar los principios éticos y 3) debe ser sólida.

Sobre estos tres componentes y los valores europeos expuestos en la sección 2, las directrices señalan siete requisitos esenciales que deben respetar las aplicaciones de IA para ser

---

7 https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

8 https://ec.europa.eu/digital-single-market/en/european-ai-alliance

9 A la consulta respondieron 511 organizaciones, asociaciones, empresas, institutos de investigación, particulares y otros. Puede consultarse un resumen de las respuestas en: https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf

10 El trabajo del grupo de expertos fue acogido favorablemente por los Estados miembros, y el Consejo, en sus conclusiones adoptadas el 18 de febrero de 2019, tomó nota, *inter alia*, de la próxima publicación de las directrices éticas y apoyó el esfuerzo de la Comisión de llevar el enfoque ético de la UE a la escena mundial: https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/es/pdf

11 https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

consideradas fiables. Las directrices también incluyen una lista para ayudar a comprobar si se cumplen estos requisitos.

Los siete requisitos esenciales son los siguientes:

- Intervención y supervisión humanas
- Solidez y seguridad técnicas
- Privacidad y gestión de datos
- Transparencia
- Diversidad, no discriminación y equidad
- Bienestar social y medioambiental
- Rendición de cuentas

Aunque estos requisitos están pensados para ser aplicados a todos los sistemas de IA en diferentes entornos y sectores, el contexto específico en el que se apliquen debe tenerse en cuenta para su implementación concreta y proporcionada, adoptando un enfoque basado en el impacto. A modo de ejemplo, una aplicación de IA que sugiere un libro de lectura inadecuado es mucho menos peligrosa que otra que diagnostique erróneamente un cáncer y, por tanto, puede estar sujeta a una supervisión menos estricta.

Las directrices elaboradas por el grupo de expertos de alto nivel sobre la IA no son vinculantes y, como tales, no crean nuevas obligaciones legales. No obstante, muchas disposiciones vigentes del Derecho de la Unión (y a menudo de uso o ámbito específico) ya reflejan uno o varios de estos requisitos esenciales, por ejemplo, las normas de seguridad, de protección de los datos personales, de privacidad o de protección del medio ambiente.

La Comisión se congratula del trabajo del grupo de expertos de alto nivel sobre la IA y lo considera una valiosa aportación para su elaboración de políticas.

2.2. Requisitos esenciales para una IA fiable

La Comisión apoya los siguientes requisitos esenciales para una IA fiable, que están basados en valores europeos. Anima a las partes interesadas a aplicarlos y a comprobar la lista que los lleva a la práctica con el fin de crear el entorno adecuado de confianza para un desarrollo y un uso provechosos de la IA. La Comisión acoge favorablemente las reacciones de las partes interesadas para evaluar si esta lista facilitada en las directrices requiere de otros ajustes.

I. <u>Intervención y supervisión humanas</u>

Los sistemas de IA deben ayudar a las personas a elegir mejor y con más conocimiento de causa en función de sus objetivos. Deben actuar como facilitadores de una sociedad floreciente y equitativa, apoyando la intervención humana y **los derechos fundamentales**, y no disminuir, limitar o desorientar la autonomía humana. El **bienestar global del usuario** debe ser primordial en la funcionalidad del sistema.

La supervisión humana ayuda a garantizar que un sistema de IA no socave la autonomía humana ni cause otros efectos adversos. Dependiendo del sistema específico de IA y de su

ámbito de aplicación, deben garantizarse los grados adecuados de **medidas de control**, incluida la adaptabilidad, la exactitud y la explicabilidad de los sistemas de IA[12]. **La supervisión** debe lograrse a través de mecanismos de gobernanza, tales como el enfoque de la participación humana (*human-in-the-loop*), la supervisión humana (*human-on-the-loop*), o el control humano (*human-in-command*).[13] Hay que garantizar que las autoridades públicas tengan la capacidad de ejercer sus competencias de supervisión conforme a sus mandatos. En igualdad de condiciones, cuanto menor sea la supervisión que puede ejercer un ser humano sobre un sistema de IA, más extensas tendrán que ser las pruebas y más estricta la gobernanza.

II. Solidez y seguridad técnicas

La fiabilidad de la IA requiere que los algoritmos sean suficientemente seguros, fiables y sólidos para resolver errores o incoherencias durante todas las fases del ciclo vital del sistema de IA y hacer frente adecuadamente a los resultados erróneos. Los sistemas de IA deben ser **fiables**, lo bastante seguros para ser **resilientes,** tanto frente a los ataques abiertos como a tentativas más sutiles de manipular datos o los propios algoritmos, y deben garantizar un **plan de contingencia** en caso de problemas. Sus decisiones deben ser **acertadas** o, como mínimo, reflejar su nivel de acierto, y sus resultados, **reproducibles**.

Además, los sistemas de IA deben integrar mecanismos de seguridad y de seguridad desde el diseño para garantizar que sean **verificablemente seguros** en cada fase, teniendo muy presente la seguridad física y psicológica de todos los afectados. Esto incluye la minimización y, cuando sea posible, la reversibilidad de las consecuencias no deseadas o errores en el funcionamiento del sistema. Deben instaurarse procesos para aclarar y evaluar los riesgos potenciales asociados al uso de los sistemas de IA, en diversos ámbitos de aplicación.

III. Privacidad y gestión de datos

Deben garantizarse la privacidad y la **protección de datos** en **todas las fases** del ciclo vital del sistema de IA. Los registros digitales del comportamiento humano pueden permitir que los sistemas de IA infieran no solo las preferencias, la edad y el sexo de las personas, sino también su orientación sexual o sus opiniones religiosas o políticas. Para que las personas puedan confiar en el tratamiento de datos, debe garantizarse que tienen el pleno control sobre sus propios datos, y que los datos que les conciernen no se utilizarán para perjudicarles o discriminarles.

Además de salvaguardar la privacidad y los datos personales, deben cumplirse requisitos en cuanto a garantizar la calidad de los sistemas de IA. La calidad de los conjuntos de datos utilizados es primordial para el funcionamiento de los sistemas de IA. Cuando se

[12] El Reglamento general de protección de datos da a las personas el derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado cuando produzca efectos jurídicos en los usuarios o les afecte significativamente de modo similar (artículo 22 del RGPD).

[13] *Human-in-the-loop* (HITL) se refiere a la intervención humana en cada ciclo de decisión del sistema, lo que en muchos casos no es posible ni deseable. *Human-on-the-loop* (HOTL) se refiere a la capacidad de la intervención humana durante el ciclo de diseño del sistema y a la supervisión del funcionamiento del sistema. *Human-in-command* (HIC) se refiere a la capacidad de supervisar la actividad global del sistema de IA (incluido su impacto más amplio económico, social, jurídico y ético) y a la capacidad de decidir cuándo y cómo utilizar el sistema en cada situación determinada. Esto puede incluir la decisión de no utilizar un sistema de IA en una situación concreta, establecer niveles de discreción humana durante el uso del sistema o garantizar la capacidad de imponerse a una decisión tomada por el sistema.

recopilan datos, pueden reflejar sesgos sociales, o contener inexactitudes o errores. Esto debe resolverse antes de entrenar un sistema de IA con un conjunto de datos. Además, debe garantizarse la **integridad** de los datos. Los procesos y conjuntos de datos utilizados deben ponerse a prueba y documentarse en cada fase, como la planificación, el entrenamiento, el ensayo y el despliegue. Esto debe aplicarse también a los sistemas de IA que no han sido desarrollados internamente, sino que se han adquirido fuera. Por último, el **acceso** a los datos debe estar adecuadamente regulado y controlado.

IV. Transparencia

Debe garantizarse la **trazabilidad** de los sistemas de IA; es importante registrar y documentar tanto las decisiones tomadas por los sistemas como la totalidad del proceso (incluida una descripción de la recogida y el etiquetado de datos, y una descripción del algoritmo utilizado) que dio lugar a las decisiones. A este respecto, en la medida de lo posible debe aportarse la **explicabilidad** del proceso de toma de decisiones algorítmico, adaptada a las personas afectadas. Debe proseguirse la investigación en curso para desarrollar mecanismos de explicabilidad. Además, deben estar disponibles las explicaciones sobre el grado en que un sistema de IA influye y configura el proceso organizativo de toma de decisiones, las opciones de diseño del sistema, así como la justificación de su despliegue (garantizando, por tanto, no solo la transparencia de los datos y del sistema, sino también la transparencia del modelo de negocio).

Por último, es importante **comunicar** adecuadamente las capacidades y limitaciones del sistema de IA a las distintas partes interesadas afectadas de una manera adecuada al caso de que se trate. Por otra parte, los sistemas de IA deben ser identificables como tales, garantizando que los usuarios sepan que están interactuando con un sistema de IA y qué personas son responsables del mismo.

V. Diversidad, no discriminación y equidad

Los conjuntos de datos utilizados por los sistemas de IA (tanto para el entrenamiento como para el funcionamiento) pueden verse afectados por la inclusión de sesgos históricos involuntarios, por no estar completos o por modelos de gobernanza deficientes. La persistencia en estos sesgos podría dar lugar a una discriminación (in)directa. También pueden producirse daños por la explotación intencionada de sesgos (del consumidor) o por una competencia desleal. Por otra parte, la forma en la que se desarrollan los sistemas de IA (por ejemplo, la forma en que está escrito el código de programación de un algoritmo) también puede estar sesgada. Estos problemas deben abordarse desde el inicio del desarrollo del sistema.

También puede ayudar a resolver estos problemas establecer **equipos de diseño diversificados** y crear mecanismos que garanticen la **participación,** en particular de los ciudadanos, en el desarrollo de la IA. Es conveniente consultar a las partes interesadas que puedan verse afectadas directa o indirectamente por el sistema a lo largo de su ciclo de vida. Los sistemas de IA deberían tener en cuenta toda la gama de capacidades, habilidades y necesidades humanas y garantizar la accesibilidad mediante un enfoque de diseño universal para tratar de lograr la igualdad de acceso para las personas con discapacidades.

VI. Bienestar social y medioambiental

Para que la IA sea fiable, debe tomarse en cuenta su impacto sobre el **medio ambiente y sobre otros seres sensibles**. Idealmente, todos los seres humanos, incluso las generaciones futuras, deberían beneficiarse de la biodiversidad y de un entorno habitable. Debe, por tanto, fomentarse la sostenibilidad y la **responsabilidad ecológica** de los sistemas de IA. Lo mismo puede decirse de las soluciones de IA que abordan ámbitos de interés mundial, como por ejemplo los objetivos de desarrollo sostenible de las Naciones Unidas.

Por otra parte, el impacto de los sistemas de IA debe considerarse no solo desde una perspectiva individual, sino también desde la perspectiva de la **sociedad en su conjunto**. Debe prestarse especial atención al uso de los sistemas de IA, particularmente en situaciones relacionadas con el proceso democrático, incluida la formación de opinión, la toma de decisiones políticas o en el contexto electoral. También debe tenerse en cuenta el **impacto social** de la IA. Si bien los sistemas de IA pueden utilizarse para mejorar las habilidades sociales, de la misma manera pueden contribuir a su deterioro.

VII. Rendición de cuentas

Deben instaurarse mecanismos que garanticen la responsabilidad y la rendición de cuentas de los sistemas de IA y de sus resultados, tanto antes como después de su implementación. La **posibilidad de auditar** los sistemas de IA es fundamental, puesto que la evaluación de los sistemas de IA por parte de auditores internos y externos, y la disponibilidad de los informes de evaluación, contribuye en gran medida a la fiabilidad de la tecnología. La posibilidad de realizar auditorías externas debe garantizarse especialmente en aplicaciones que afecten a los derechos fundamentales, por ejemplo las aplicaciones críticas para la seguridad.

Los potenciales impactos negativos de los sistemas de IA deben señalarse, evaluarse, documentarse y reducirse al mínimo. El uso de las evaluaciones de impacto facilita este proceso. Estas evaluaciones deben ser proporcionales a la magnitud de los riesgos que plantean los sistemas de IA. Los **compromisos** entre los requisitos —que a menudo son inevitables— deben abordarse de una manera racional y metodológica, y ser tenidos en cuenta. Por último, cuando se produzcan efectos adversos injustos, deben estar previstos mecanismos accesibles que garanticen una **reparación adecuada**.

2.3. Próximas etapas: una fase piloto global en la que participen las partes interesadas a la escala más amplia posible

Alcanzar un consenso sobre estos requisitos esenciales para un sistema de IA es un primer hito importante en el camino hacia las directrices para una IA ética. Como siguiente paso, la Comisión garantizará que esta orientación pueda probarse e implementarse en la práctica.

Para ello, la Comisión pondrá en marcha una fase piloto específica concebida para obtener respuestas estructuradas de las partes interesadas. Este ejercicio se centrará, en particular, en la lista elaborada por el grupo de expertos de alto nivel para cada uno de los requisitos esenciales.

Este trabajo seguirá dos líneas: i) una fase piloto para las directrices en la que participen las partes interesadas que desarrollan o utilizan IA, incluidas las administraciones públicas, y ii) un proceso permanente de consulta a las partes interesadas y de sensibilización entre los

Estados miembros y diferentes grupos de partes interesadas, entre otras el sector de la industria y los servicios:

i) A partir de junio de 2019, se pedirá a todas las partes interesadas y los particulares que prueben la lista y den su opinión sobre cómo mejorarla. Además, el grupo de expertos de alto nivel sobre la IA realizará un examen exhaustivo con las partes interesadas del sector público y privado para recopilar puntos de vista más detallados sobre cómo pueden implementarse las directrices en una amplia gama de ámbitos de aplicación. Todas las respuestas sobre la viabilidad de las directrices se evaluarán para finales de 2019.

ii) Al mismo tiempo, la Comisión organizará otras actividades de divulgación, dando la oportunidad a los representantes del grupo de expertos de alto nivel sobre la IA de presentar las directrices a las partes interesadas relevantes en los Estados miembros, incluida la industria y el sector servicios, y brindando a estas partes interesadas la oportunidad adicional de formular observaciones sobre las directrices de IA y de contribuir a ellas.

La Comisión tendrá en cuenta el trabajo del grupo de expertos sobre ética para la conducción conectada y automatizada[14] y el trabajo con proyectos de investigación financiados por la UE y con asociaciones relevantes público-privadas sobre la implementación de los requisitos esenciales[15]. Por ejemplo, la Comisión apoyará, en coordinación con los Estados miembros, el desarrollo de una base de datos común de imaginería médica inicialmente dedicada a las formas más comunes de cáncer, de manera que puedan entrenarse los algoritmos para diagnosticar síntomas con gran precisión. De forma similar, la cooperación de la Comisión y los Estados miembros permite multiplicar los corredores transfronterizos para probar vehículos conectados y automatizados. Las directrices deben aplicarse en estos proyectos y someterse a ensayo y los resultados alimentarán el proceso de evaluación.

La fase piloto y la consulta a las partes interesadas se beneficiarán de la contribución de la Alianza europea de la IA y de la AI4EU, la plataforma de IA a la demanda. El proyecto AI4EU[16], puesto en marcha en enero de 2019, reúne algoritmos, herramientas, conjuntos de datos y servicios para ayudar a las organizaciones, en particular a las pequeñas y medianas empresas, a implementar soluciones de IA. La Alianza europea de la IA, junto con la AI4EU, continuará movilizando el ecosistema de IA en toda Europa, también con vistas a poner a prueba las directrices éticas en materia de IA y fomentar el respeto de la IA centrada en el ser humano.

A comienzos de 2020, a partir de la evaluación de las reacciones recibidas durante la fase piloto, **el grupo de expertos de alto nivel sobre la IA revisará y actualizará las directrices**. Sobre la base de la revisión y de la experiencia adquirida, **la Comisión evaluará los resultados y propondrá las próximas etapas.**

Con la propuesta de directrices éticas para la IA todos salen ganando. Garantizar el respeto de los valores y derechos fundamentales no solo es esencial en sí mismo, sino que también facilita la aceptación por parte del público y aumenta la ventaja competitiva de las empresas europeas de IA al establecer un planteamiento de IA centrada en el ser humano, fiable, reconocida por sus productos éticos y seguros. Más en general, esto se basa en la sólida

[14] Véase la Comunicación de la Comisión sobre la movilidad conectada y automatizada, COM(2018) 283.

[15] En el marco del Fondo Europeo de Defensa, la Comisión desarrollará también orientaciones éticas específicas para la evaluación de propuestas de proyectos en el ámbito de la IA para la defensa.

[16] https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019

reputación de las empresas europeas por sus productos seguros y de gran calidad. La fase piloto contribuirá a garantizar que los productos de IA cumplan esta promesa.

2.4. Hacia unas directrices éticas en materia de IA internacionales

Las conversaciones internacionales sobre la ética en materia de IA se han intensificado después de que la presidencia japonesa del G7 diera gran importancia al tema en la agenda de 2016. Dadas las interrelaciones internacionales en el desarrollo de la IA en cuanto a circulación de datos, desarrollo de algoritmos e inversiones en investigación, **la Comisión seguirá esforzándose por llevar el enfoque de la Unión a la escena mundial y establecer un consenso sobre una IA centrada en el ser humano**[17].

El trabajo realizado por el grupo de expertos de alto nivel sobre la IA, y más en concreto la lista de requisitos y el proceso de participación con las partes interesadas, ofrece a la Comisión una valiosa aportación adicional para contribuir a los debates internacionales. La Unión Europea puede desempeñar un papel de liderazgo en el desarrollo de directrices internacionales sobre IA y, si es posible, un mecanismo de evaluación al respecto.

Por consiguiente, la Comisión:

Estrechará la cooperación con socios de ideas afines:

- explorando hasta qué punto puede lograrse la convergencia con los proyectos de directrices éticas de terceros países (por ejemplo, Japón, Canadá, Singapur) y, apoyándose en este grupo de países de ideas afines, preparar un debate más amplio, respaldado por acciones que implementen el Instrumento de Colaboración para la cooperación con terceros países[18]; y
- explorando cómo pueden contribuir las empresas de países no pertenecientes a la UE y las organizaciones internacionales a la fase piloto de las directrices mediante la realización de ensayos y validaciones.

Continuará desempeñando un papel activo en las conversaciones e iniciativas internacionales:

- contribuyendo a foros multilaterales como el G7 y el G20;
- participando en diálogos con países no pertenecientes a la UE y organizando reuniones bilaterales y multilaterales para llegar a un consenso sobre la IA centrada en el ser humano;

---

17 La alta representante de la Unión para Asuntos Exteriores y Política de Seguridad, con el apoyo de la Comisión, se basará en consultas con las Naciones Unidas, el Panel de Tecnología Global y otros organismos multilaterales, y en particular, coordinará propuestas para hacer frente a los complejos desafíos de seguridad que se plantean.

18 Reglamento (UE) n.º 234/2014 del Parlamento Europeo y del Consejo, de 11 de marzo de 2014, por el que se establece un Instrumento de Colaboración para la cooperación con terceros países (DO L 77 de 15.3.2014, p. 77). Por ejemplo, el proyecto previsto sobre «Una alianza internacional para un enfoque centrado en el ser humano para la IA» facilitará iniciativas conjuntas con socios de ideas afines, con el fin de promover unas directrices éticas y adoptar principios comunes y conclusiones operativas. Permitirá a la UE y países de ideas afines debatir conclusiones operativas derivadas de las directrices éticas sobre la IA propuestas por el grupo de expertos de alto nivel para alcanzar un enfoque común. Además, permitirá seguir el despliegue de la tecnología en materia de IA a nivel mundial. Por último, el proyecto prevé organizar actividades de diplomacia pública durante actos internacionales, por ejemplo, del G7, G20 y la Organización para la Cooperación y el Desarrollo Económicos.

- contribuyendo a actividades de normalización relevantes en organizaciones de desarrollo de normas internacionales para promover esta visión; y
- reforzando la recogida y difusión de puntos de vista sobre políticas públicas, trabajando conjuntamente con organizaciones internacionales relevantes.

3. CONCLUSIONES

La UE reposa sobre un conjunto de valores fundamentales y ha construido un marco regulatorio sólido y equilibrado sobre estos cimientos. A partir de este marco regulatorio existente, son necesarias unas directrices éticas para el desarrollo y la utilización de la IA, dado lo novedoso de esta tecnología y los retos específicos que trae consigo. La IA solo podrá considerarse fiable si se desarrolla y utiliza de forma que respete unos valores éticos ampliamente compartidos.

Teniendo presente este objetivo, la Comisión acoge favorablemente la aportación preparada por el grupo de expertos de alto nivel sobre la IA. Sobre la base de los requisitos esenciales para que la IA se considere fiable, ahora la Comisión pondrá en marcha una fase piloto específica para garantizar que las directrices éticas derivadas para el desarrollo y la utilización de la IA puedan ser aplicadas en la práctica. La Comisión también trabajará para forjar un amplio consenso social sobre la IA centrada en el ser humano, incluyendo en ello a todas las partes interesadas y a nuestros socios internacionales.

La dimensión ética de la IA no es un lujo ni un algo accesorio: ha de ser parte integrante del desarrollo de la IA. Al tratar de lograr una IA centrada en el ser humano basada en la confianza, salvaguardamos el respeto de los valores esenciales de nuestra sociedad y forjamos una marca distintiva para Europa y su industria como líder de la IA de vanguardia en la que se puede confiar en todo el mundo.

Para garantizar el desarrollo ético de la IA en Europa en su contexto más amplio, la Comisión aplica un enfoque global que incluye, en particular, las siguientes líneas de acción para ser implementadas antes del tercer trimestre de 2019:

- Comenzará a poner en marcha un conjunto de **redes de centros de excelencia especializados en investigación sobre IA** a través de Horizonte 2020. Seleccionará un máximo de cuatro redes, centrándose en retos científicos o tecnológicos importantes, como la explicabilidad y la interacción avanzada entre los seres humanos y las máquinas, que son elementos clave para una IA fiable.
- Empezará a crear **redes de polos de innovación digital**[19] centrándose en la IA en la fabricación y en los macrodatos.
- Junto con los Estados miembros y las partes interesadas, la Comisión entablará conversaciones preparatorias para desarrollar y aplicar **un modelo para el intercambio de datos y para hacer el mejor uso de los espacios comunes de datos,** haciendo hincapié en el transporte, la atención sanitaria y la fabricación industrial[20].

Además, la Comisión está elaborando un informe sobre los retos que plantea la IA en relación con los marcos de seguridad y responsabilidad y un documento de orientación sobre la

[19] http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs

[20] Los recursos necesarios procederán de Horizonte 2020 (en virtud del cual cerca de 1 500 millones EUR están asignados a la IA durante el periodo 2018-2020) y su sucesor previsto Horizonte Europa, la parte digital del Mecanismo «Conectar Europa» y especialmente el futuro Programa Europa Digital. Los proyectos también utilizarán recursos del sector privado y de los programas de los Estados miembros.

implementación de la Directiva sobre responsabilidad por los daños causados por productos defectuosos[21]. Al mismo tiempo, la Empresa Común de Informática de Alto Rendimiento Europea (EuroHPC)[22] desarrollará la próxima generación de superordenadores, ya que la capacidad de computación es esencial para el tratamiento de datos y la formación en IA, y Europa necesita dominar la totalidad de la cadena de valor digital. La asociación en curso con los Estados miembros y la industria sobre componentes y sistemas microelectrónicos (ECSEL)[23], así como la iniciativa europea en materia de procesadores[24], contribuirán al desarrollo de una tecnología de procesadores de bajo consumo para una computación en el borde (*edge computing*) de alto rendimiento, fiable y segura.

Al igual que el trabajo sobre directrices éticas para la IA, todas estas iniciativas parten de la **estrecha cooperación con todas las partes afectadas**, Estados miembros, industria, agentes sociales y ciudadanos. En conjunto, el enfoque de Europa con respecto a la IA muestra cómo la competitividad económica y la confianza de la sociedad deben partir de los mismos valores fundamentales y reforzarse mutuamente.

21 Véase la Comunicación de la Comisión sobre la Inteligencia artificial para Europa, COM(2018) 237.

22 https://eurohpc-ju.europa.eu

23 www.ecsel.eu

24 www.european-processor-initiative.eu