Comisión Europea

*Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías*

Declaración sobre

# Inteligencia artificial, robótica y sistemas "autónomos"

*Investigación e Innovación*

Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías
**Inteligencia artificial, robótica y sistemas "autónomos"**

Comisión Europea
Dirección General de Investigación e Innovación
Unidad RTD.01 — Mecanismo de Asesoramiento Científico

Contacto Jim Dratwa, Director de la Oficina del GEE
E-mail EC-ETHICS-GROUP@ec.europa.eu
RTD-PUBLICATIONS@ec.europa.eu

Comisión Europea
B-1049 Bruselas

Impreso por OP en Luxemburgo.

Manuscrito completado en marzo 2018.



Puede consultar más información sobre la Unión Europea en línea (http://europa.eu).

Luxemburgo: Oficina de Publicaciones de la Unión Europea, 2018.

| | | | |
|---|---|---|---|
| Versión original impresa | ISBN 978-92-79-80328-4 | doi:10.2777/786515 | KI-04-18-224-EN-C |
| Versión original PDF | ISBN 978-92-79-80329-1 | doi:10.2777/531856 | KI-04-18-224-EN-N |

COMISIÓN EUROPEA

Declaración sobre

# Inteligencia artificial, robótica y sistemas "autónomos"

***Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías***

*Bruselas, 9 marzo 2018*

2018 Dirección General de Investigación e Innovación

# Índice

## Resumen

*Los avances en inteligencia artificial, robótica y las llamadas tecnologías "autónomas"* [1] *han originado una serie de dilemas morales cada vez más urgentes y complejos. Actualmente se están haciendo esfuerzos para orientar estas tecnologías hacia el bien común y para encontrar soluciones a los desafíos éticos, sociales y legales que generan. Sin embargo, estos esfuerzos han resultado ser un mosaico de iniciativas dispares. Esta situación genera la necesidad de implementar un proceso colectivo, amplio e inclusivo de reflexión y diálogo. Este diálogo debe estar basado en los valores en los que queremos fundamentar nuestra sociedad y en el papel que queremos que la tecnología desempeñe.*

*Esta declaración hace un llamado para iniciar la construcción de un marco ético y legal común e internacionalmente reconocido para el diseño, producción, uso y gobernanza de la inteligencia artificial, la robótica y los sistemas "autónomos". Además, esta declaración propone un conjunto de principios éticos fundamentales que pueden servir de guía para el desarrollo de este marco ético y legal. Estos principios están basados en los valores establecidos en los Tratados de la UE y en la Carta de Derechos Fundamentales de la UE.*

[1] Esta declaración se enfoca en el conjunto de tecnologías digitales inteligentes. Actualmente estas tecnologías convergen y a menudo están interrelacionadas, conectadas o totalmente integradas. Algunos ejemplos son la inteligencia artificial clásica, los algoritmos de aprendizaje automático (en inglés Machine Learning) y aprendizaje profundo (en inglés Deep Learning), las redes conexionistas y generativas antagónicas, la mecatrónica y la robótica. Entre los casos más conocidos que ilustran la convergencia de estas tecnologías podemos mencionar los automóviles que prescinden de un conductor, los sistemas de armas robóticas, los *bots* conversacionales (*chatbots*) y los sistemas de reconocimiento de voz e imagen.

## Antecedentes

Durante las dos primeras décadas del siglo XXI hemos sido testigos de aplicaciones sorprendentes de la "tecnología autónoma" y la "inteligencia artificial". Entre algunos de los ejemplos más destacados se encuentran los automóviles que no necesitan de un conductor, los drones que pueden volar por sí mismos, los robots para las exploraciones marítimas y espaciales, los sistemas de armas, los agentes de software como los *bots* financieros, y el diagnóstico médico asistido por aprendizaje profundo (en inglés *Deep Learning*). Importantes impulsores de estos avances son la inteligencia artificial (IA), especialmente en la forma de aprendizaje automático (en inglés *Machine Learning*), y la disponibilidad cada vez mayor de conjuntos de datos masivos generados a partir de diferentes ámbitos de la vida. Estas tecnologías digitales confluyen en diversas aplicaciones, lo que rápidamente las hace más poderosas. Por ejemplo, estas son empleadas en un creciente número de novedosos productos y servicios en los sectores público y privado, con aplicaciones tanto militares como civiles. Posibles consecuencias de la integración de la IA en estos sistemas son la redefinición el concepto de trabajo, la mejora de las condiciones de trabajo, y la reducción del aporte y la interferencia humana durante las operaciones. Además, la IA puede ayudar mediante tecnología inteligente a asistir o reemplazar humanos en trabajos difíciles, sucios, aburridos o peligrosos, entre otros.

Actualmente, y sin intervención humana directa o control externo, los sistemas inteligentes entablan diálogos con clientes en centros de llamadas en línea, manejan incesantemente y con gran precisión manos robóticas que recogen y manipulan objetos, compran y venden grandes cantidades de acciones en milisegundos, maniobran o frenan automóviles para prevenir choques, clasifican personas y su comportamiento, e imponen multas.

Lamentablemente, podemos observar que las herramientas cognitivas más poderosas son también las más opacas, puesto que sus acciones han dejado de ser programadas linealmente por humanos. En este sentido, podrían destacarse algunos ejemplos. Google Brain desarrolla una IA que al parecer construye otras de su misma naturaleza mejor y más rápidamente que los humanos. AlphaZero puede autoejecutarse y pasar en cuatro horas de ser completamente ignorante de las reglas del ajedrez a alcanzar el nivel del campeón del mundo. De hecho, es imposible comprender exactamente cómo logró AlphaGo vencer al campeón mundial de Go. Estos dos casos ilustran que el aprendizaje profundo y los llamados "enfoques de redes generativas antagónicas" (en inglés *generative adversarial networks*) hacen posible que las máquinas se "enseñen" a sí mismas nuevas estrategias y adquieran nuevos elementos para ser incorporados en sus análisis. De esta forma, las acciones de estas máquinas se vuelven indescifrables y escapan del escrutinio humano. Esto se debe, en primer lugar, a que resulta imposible averiguar cómo se generan los resultados más allá de los algoritmos iniciales. En segundo lugar, porque el rendimiento de estas máquinas se basa en los datos utilizados durante el proceso de aprendizaje y estos pueden no estar disponibles o ser inaccesibles. Esto también implica que si estos sistemas usan datos con sesgos y errores, estos dos últimos quedarán enraizados en el sistema.

A menudo, cuando los sistemas pueden aprender a realizar este tipo de tareas complejas sin la instrucción o supervisión humana, se les califica como "autónomos". Dichos sistemas pueden manifestarse en forma de sistemas robóticos de alta tecnología o software inteligente, como los *bots*. En muchos casos, estos sistemas autónomos son lanzados y liberados en nuestro mundo sin supervisión, a pesar de que poseen el potencial de alcanzar objetivos que no fueron previstos por sus diseñadores o propietarios humanos.

Por lo tanto, consideramos relevantes los siguientes avances tecnológicos:

(1) La IA en la forma de aprendizaje automático (especialmente "aprendizaje profundo") y que se alimenta de datos masivos, se está volviendo cada vez más poderosa. Asimismo, la IA se está aplicando en mayor medida a nuevos productos y servicios digitales, en los

sectores público y privado, y en contextos tanto militares como civiles. Como se señaló anteriormente, el funcionamiento interno de la IA puede ser extremadamente difícil o imposible de monitorear, explicar y evaluar críticamente. Además, las capacidades avanzadas de la IA se están acumulando sobre todo en manos del sector privado y, por lo general, con derechos exclusivos.

(2) La mecatrónica avanzada (una combinación de IA, aprendizaje profundo, ciencia de datos, tecnología de sensores, internet de las cosas y las ingenierías mecánica y eléctrica) ofrece una amplia gama de sofisticados sistemas robóticos y de alta tecnología de aplicación práctica. Por ejemplo, en los sectores de servicios y producción industrial, asistencia sanitaria, comercio minorista, logística, domótica (automatización del hogar), seguridad y protección. Dos campos de particular importancia en los debates públicos son los sistemas robóticos aplicados a las armas y los vehículos "autónomos".

(3) Se están creando sistemas cada vez más inteligentes que dicen exhibir un alto grado de "autonomía", lo que significa que son sistemas que de forma independiente desarrollan y realizan tareas, sin necesidad de operadores o de control humano.

(4) Parece haber una tendencia a incrementar la automatización y "autonomía" en robótica, IA y mecatrónica. Naciones y grandes empresas han hecho enormes inversiones en este campo y las superpotencias del mundo aspiran a asegurarse una posición de liderazgo en la investigación de la IA.

(5) Existe un desarrollo dirigido a estrechar aún más la interacción entre los humanos y las máquinas (como en el caso de los *cobots*, *cybercrews, digital twins*, interfaces cerebro-computadora y cíborgs, estos dos últimos ejemplos integran máquinas inteligentes en el cuerpo humano). Avances similares se pueden observar en la IA. Por ejemplo, equipos compuestos por sistemas de IA y profesionales humanos obtienen mejor rendimiento en algunos campos que cuando trabajan por separado.

## Reflexiones morales

### Preguntas clave

Los actuales sistemas y software de alta tecnología emergentes son merecedores de un análisis particular. Específicamente se hace referencia a sistemas y software que se vuelven progresivamente independientes de los humanos y ejecutan tareas que convencionalmente requieren inteligencia humana. La importancia de realizar este análisis reside en las relevantes y complejas cuestiones morales que estos sistemas suscitan.

Primero, estos sistemas generan preguntas sobre seguridad, protección, prevención de daños y mitigación de riesgos. ¿Cómo podemos construir un mundo con IA y dispositivos "autónomos" interconectados que sea seguro y cómo podemos estimar los riesgos involucrados?

Segundo, originan preguntas sobre responsabilidad moral humana. Tomemos como punto de partida los sistemas sociotécnicos dinámicos y complejos que incorporan IA y componentes robóticos avanzados. En esos casos, ¿dónde se sitúa el agente moralmente relevante? ¿Cómo se debe atribuir y distribuir la responsabilidad moral? ¿Quién es responsable de resultados no deseados, y en qué sentido es responsable? ¿Es razonable hablar de "control compartido" y "responsabilidad compartida" entre humanos y máquinas inteligentes? ¿Formarán los humanos parte de los ecosistemas de dispositivos "autónomos" solamente para funcionar como "zonas de absorción de impacto" moral, para asumir obligaciones, o estarán realmente en una posición que les permita responsabilizarse de lo que hacen?

Tercero, estos sistemas generan preguntas sobre gobernanza, regulación, diseño, desarrollo, inspección, monitoreo, prueba y certificación. ¿Cómo se deben nuestras instituciones y leyes ser rediseñadas para que estén al servicio del bienestar de las personas y la sociedad, y para hacer de la sociedad un lugar seguro ante la aplicación de estas tecnologías?

Cuarto, hay preguntas sobre la toma de decisiones democráticas que incluyen aquellas que atañen a instituciones, políticas y valores. Estas significativas cuestiones apuntalan todas las anteriores. Por ejemplo, en todo el mundo se llevan a cabo investigaciones para determinar en qué medida terceros se están aprovechando de los ciudadanos mediante técnicas avanzadas de provocación (en inglés *nudging*). Estas son técnicas que están basadas en aprendizaje automático, datos masivos y ciencias del comportamiento. Haciendo uso de estas herramientas, es posible crear perfiles personales detallados, aplicar protocolos de microfocalización (en inglés *microtargeting*), y adaptar y manipular las arquitecturas de toma de decisiones según fines comerciales o políticos.

Finalmente, también se plantean preguntas sobre la transparencia de la IA y los sistemas "autónomos" y la capacidad que tenemos para explicarlos. ¿Que valores sustentan estos sistemas de forma efectiva y demostrable? ¿En cuáles valores se fundamenta la forma en la que diseñamos nuestras políticas y nuestras máquinas? ¿En cuáles valores queremos basar nuestras sociedades? Además, ¿cuáles valores estamos permitiendo que sean socavados, abierta o silenciosamente, en aras del progreso tecnológico? Por último, ¿qué concesiones estamos dispuestos a hacer para obtener utilidades? Un caso para analizar estas preguntas es la aplicación de la IA para la "optimización" de procesos sociales mediante sistemas de puntuación social con los que algunos países experimentan. Este tipo de actividades violan la idea fundamental de igualdad y libertad de la misma forma que lo hacen los sistemas de castas. Ambos construyen "diferentes tipos de personas", cuando en realidad solo hay personas con "características diferentes". En vista de estos eventos, ¿cómo se puede prevenir que estas poderosas tecnologías sean utilizadas como herramientas para socavar sistemas democráticos y como mecanismos de dominación?

## Consideraciones clave

Desde una perspectiva ética, es importante tener en cuenta que:

El concepto "autonomía" tiene un origen filosófico y se refiere a la capacidad que tienen las personas humanas para legislarse a sí mismas, para formular, pensar y elegir normas, reglas y leyes que ellos mismos deben cumplir. Este concepto abarca el derecho a ser libre para establecer estándares, objetivos y propósitos de vida propios. Notablemente, aquellos procesos cognitivos que sustentan y facilitan la autonomía están entre los más estrechamente relacionados con la dignidad de las personas, la agencia humana y la actividad humana por excelencia. Por lo general, estos procesos comprenden las capacidades de autoconocimiento y autoconciencia, que a su vez están íntimamente relacionadas con motivos y valores personales. Por lo tanto, la autonomía, en el sentido éticamente relevante de la palabra, solo puede ser atribuida a los seres humanos. De ahí que resulte inapropiado utilizar el término "autonomía" para referirse a meros artefactos, aunque se trate de sistemas adaptativos complejos muy avanzados o incluso "inteligentes". Sin embargo, el término sistemas "autónomos" ha ganado gran aceptación en la literatura científica y en los debates públicos. El término se utiliza para hacer referencia al grado más alto de automatización y de independencia de los seres humanos en términos de "autonomía" operativa y de toma de decisiones. Pero la autonomía, en su sentido original, es un aspecto importante de la dignidad humana que no debe relativizarse.

Dado que ningún artefacto o sistema inteligente puede ser considerado "autónomo" en el sentido ético original, tampoco puede ser considerado titular de la moralidad y dignidad humanas. Esto sin importar lo avanzados o sofisticados que sean. La dignidad humana es el fundamento de los derechos humanos. Esto implica que se debe garantizar el control humano significativo y la participación humana en aquellos ámbitos que conciernen a los seres humanos y su entorno. Por lo tanto, a diferencia de lo que ocurre en campo de la automatización de la producción, no es apropiado administrar ni decidir sobre los seres humanos de la misma forma en la que se administra y decide sobre objetos o datos, incluso si resulta técnicamente concebible. La gestión "autónoma" aplicada a los seres humanos va en contra de consideraciones éticas y menoscaba los valores fundamentales europeos tan profundamente enraizados. Los seres humanos deben ser capaces de decidir sobre cuestiones tan importantes como los valores que fundamentan la tecnología, aquello que debe ser considerado moralmente relevante, y los objetivos últimos y los conceptos de lo que es bueno que son dignos de ser perseguidos. Este tipo de cuestiones no pueden dejarse en manos de las máquinas, no importa lo poderosas que sean.

La habilidad y la voluntad de asumir y atribuir responsabilidad moral son parte integral de la concepción de la persona. Además, en ellas se basan todas nuestras instituciones morales, sociales y legales. La responsabilidad moral se interpreta aquí en sentido amplio, haciendo referencia a varios aspectos de la agencia humana. Por ejemplo, causalidad, rendición de cuentas (obligación de dar cuentas), responsabilidad (obligación de compensar daños), actitudes reactivas como elogiar y culpabilizar (la idoneidad de diversas emociones morales) y deberes propios de roles sociales específicos. La responsabilidad moral, cualquiera que sea el sentido pertinente, no puede ser asignada o trasladada a la tecnología "autónoma".

Durante recientes debates sobre los Sistemas de Armas Letales Autónomos (LAWS por sus siglas en inglés) y vehículos autónomos, parece haberse llegado al amplio consenso de que el control humano significativo es esencial para la responsabilidad moral. En efecto, el principio de control humano significativo fue inicialmente propuesto con el objetivo de restringir el desarrollo y la utilización de los sistemas de armas del futuro. Esto significa que los humanos, y no las computadoras y sus algoritmos, deben mantener el control sobre estos sistemas y deben ser moralmente responsables.[2]

[2] ONG Artículo 36, 2015.

## Más allá de un marco ético limitado

El desarrollo de sistemas "autónomos" ya ha dado lugar a debates éticos de alto perfil en lo que se refiere a vehículos que no necesitan de un conductor y los Sistemas Autónomos de Armas Letales (LAWS por sus siglas en inglés). Aunque todavía no hay en el mercado automóviles que prescindan completamente de un conductor, muchos países en todo el mundo se están preparando para la posibilidad legal de permitir vehículos "autónomos" en vía pública. De hecho, la primera persona que murió en un accidente automovilístico mientras conducía en modo "autónomo" causó gran controversia moral en 2016. Actualmente, los debates morales sobre vehículos "autónomos" se limitan a la discusión de casos excepcionales, experimentos mentales usualmente llamados "dilemas del tranvía". Estos dilemas plantean accidentes inevitables en los que las únicas opciones posibles conllevan la pérdida de vidas humanas. Esta interpretación ultrarestrictiva de los problemas éticos suscita un enfoque calculador, que generalmente aplica parámetros excesivamente simplistas a las realidades humanas. Pareciera que este enfoque se ocupa principalmente de la responsabilidad de los sistemas "autónomos", de sus efectos y de cómo deberían ser programados para que su implementación conduzca a resultados moralmente aceptables en términos de vidas perdidas respecto a vidas salvadas. Todo esto ignora preguntas más profundas como, "¿qué decisiones relativas al diseño se tomaron en el pasado que condujeron a este dilema moral?", "¿qué valores deben contribuir al diseño?", "¿cómo se deben sopesar estos valores en caso de conflicto y quién debe sopesarlos?", "¿qué indican los abundantes datos empíricos que se están acumulando sobre la forma en la que las personas deciden en los casos del dilema del tranvía y cómo se traducen esos resultados a las configuraciones automáticas para vehículos?".

Un segundo campo de debate y controversia son los sistemas de armas "autónomos". Aunque estos sistemas militares pueden estar equipados con armas letales, en lo que respecta a su software, estos no son muy diferentes de los sistemas "autónomos" que se encuentran en diversos contextos civiles que nos son familiares. Una parte importante de este debate tiene lugar en la Conferencia sobre Ciertas Armas Convencionales que se celebra en Ginebra. Esta conferencia está dedicada a discutir la admisibilidad moral de las armas "autónomas" y la responsabilidad legal y moral del despliegue de estos sistemas. El debate debe ahora dar paso a responder preguntas sobre la naturaleza y el significado del "control humano significativo" y sobre cómo instituir formas de control moralmente deseables.

Una tercera área de aplicación relevante es el software "autónomo", por ejemplo los *bots*. En este campo, algoritmos y software ya manejan en gran medida el comercio, las finanzas y los mercados bursátiles. Asimismo, sistemas inteligentes actuales mantienen diálogos con clientes en centros de atención telefónica, esto sin intervención humana o control externo. Otro ejemplo son las interfaces de reconocimiento de voz y los sistemas de recomendación de plataformas en línea que hacen sugerencias a sus usuarios, como ya lo hacen Siri, Alexa y Cortana. Más allá de los cuestionamientos sobre protección de datos y privacidad, nos podríamos preguntar si las personas tienen derecho a saber si están tratando con un ser humano o con un artefacto de IA. Además, surgen dudas sobre si limitar las recomendaciones que hacen los sistemas de IA a sus usuarios. Esto porque estas sugerencias individuales están basadas en una personalidad construida por el sistema de IA, el cual utiliza la noción que la persona en cuestión tiene sobre sí misma.

En general, se reconoce que existe una creciente necesidad de abordar estas difíciles cuestiones éticas, legales y sociales. Sin embargo, la IA y la robótica actuales avanzan más rápidamente que el proceso para encontrar respuestas. Por si fuera poco, los presentes esfuerzos han resultado ser un mosaico de iniciativas descoordinadas, lo que genera la clara necesidad de desarrollar un marco ético común y reconocido internacionalmente para el diseño, la producción, el uso y la gobernanza de la IA, los robots y los sistemas "autónomos". Dicho marco debe estar fundamentado en un proceso colectivo, amplio e inclusivo.

Esta declaración hace una llamada para iniciar dicho proceso y propone un conjunto de principios éticos fundamentales y prerrequisitos democráticos, con el objetivo de que sean utilizados para orientar el análisis del derecho vinculante. El GEE opina que Europa debería desempeñar un papel activo y destacado en estos procesos. Consecuentemente, el GEE asume un rol supervisor en los debates sobre la responsabilidad moral de la IA y de la tecnología "autónoma", y aboga por planteamientos sistemáticos de pensamiento y de investigación. Dichos planteamientos deben tomar en cuenta aspectos éticos, legales y de gobernanza de los sistemas de alta tecnología que, independientemente de usuarios humanos, afectan nuestra realidad. Especialmente porque estos sistemas pueden ir en beneficio o en perjuicio del ser humano. El asunto aquí planteado, es un asunto de gran urgencia.

## Hacia un marco ético compartido para la inteligencia artificial, la robótica y los sistemas "autónomos"

Algunas de las iniciativas más destacadas que buscan la formulación de principios éticos para la IA y los sistemas "autónomos", provienen de la industria y de los profesionales y sus respectivas asociaciones. Entre estas iniciativas es importante destacar el tratado "Diseño Éticamente Alineado" del IEEE (Instituto de Ingenieros Eléctricos y Electrónicos),[3] la Cumbre Global "IA para el bien" [4] de la IUT (Unión Internacional de Telecomunicaciones), que se llevó a cabo en el verano de 2017, y el trabajo de la ACM (Asociación para Maquinaria Computacional), donde cabe destacar la conferencia AAAI/ACM "IA, Ética y Sociedad" [5], que se llevó a cabo en febrero de 2018. Otros ejemplos dentro del sector privado son los esfuerzos de las compañías IBM, Microsoft y DeepMind de Google, que han establecido sus propios códigos éticos para la IA y han unido esfuerzos para crear iniciativas de gran alcance. Entre ellas se encuentran la asociación "Partnership on AI" [6] y "OpenAI",[7] que generan coaliciones entre industria, organizaciones sin fines de lucro e instituciones académicas.

Una de las principales iniciativas que abogan por un desarrollo responsable de la IA está siendo liderada por el *Future of Life Institute*, desde donde se han formulado los "Principios de Asilomar para la IA". Estos son 23 principios considerados fundamentales para orientar la investigación y la aplicación de la IA. Esta lista ya ha sido firmada por cientos de partes interesadas.[8] Los signatarios son mayoritariamente científicos, investigadores de la IA y representantes del sector industrial. Un proceso participativo similar se dio durante el Foro sobre el Desarrollo Socialmente Responsable de la Inteligencia Artificial, celebrado por la Universidad de Montreal en noviembre de 2017. Este culminó con el borrador inicial de una potencial "Declaración para el Desarrollo Responsable de la Inteligencia Artificial". Este documento es de acceso abierto y en su plataforma en línea se invita a miembros de todos los sectores de la sociedad a realizar comentarios sobre el texto.[9]

Asimismo, la ONU y las conferencias de la Convención sobre Ciertas Armas Convencionales (CCW por sus siglas en inglés, Ginebra) han iniciado un debate mundial sobre las aplicaciones militares de la IA. En este debate, la mayoría de las Altas Partes Contratantes respaldaron el llamado principio de "control humano significativo para los Sistemas de Armas Letales Autónomos". Este principio establece que "los sistemas autónomos de armas que no requieren control humano significativo deben ser prohibidos" (Asamblea General de la ONU, 2016). Además, la ONU ha establecido un instituto de investigación en La Haya dedicado al estudio de la gobernanza de la robótica y la IA (UNICRI). En general, existen otras muchas iniciativas y ONG, que luchan por alcanzar IA y sistemas "autónomos" "para el bien", y para que las armas "autónomas" sean prohibidas. Un ejemplo es la Fundación para la Robótica Responsable.

En cambio, a nivel nacional las iniciativas son dispares. Mientras algunos países priorizan el desarrollo de normas para robots e IA y adoptan legislación pertinente (por ejemplo, para regular vehículos sin conductor en vías públicas), otros países aún tienen que comenzar a abordar estos temas. Evidencia de esta situación es la ausencia de un enfoque europeo armonizado, que ha llevado al Parlamento Europeo a dar los primeros pasos para establecer una regulación para la robótica avanzada.[10] Esto incluye el desarrollo de un marco ético rector para el diseño, producción y uso de robots.

[3] http://standards.ieee.org/news/2016/ethically_aligned_design.html
[4] https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx
[5] http://www.aies-conference.com/
[6] https://www.partnershiponai.org/
[7] https://openai.com/
[8] https://futureoflife.org/ai-principles/
[9] http://nouvelles.umontreal.ca/en/article/2017/11/03/montreal-declaration-for-a-responsible-development-of-artificial-intelligence/
[10] Parlamento Europeo, Comisión de Asuntos Jurídicos 2015/2103 (INL). Informe con recomendaciones a la Comisión sobre las normas de derecho civil sobre robótica, ponente Mady Delvaux.

Respecto a la regulación de la IA y las tecnologías "autónomas", el GEE enfatiza los riesgos que implica adoptar enfoques descoordinados y desequilibrados. Por ejemplo, los mosaicos normativos pueden dar paso a la "selección deliberada de marcos éticos", que resulta en el traslado de procesos de desarrollo y aplicación de la IA a regiones con estándares éticos más permisivos. Otro riesgo relevante se desprende de debates dominados por determinadas regiones, disciplinas, demografías o agentes de la industria. Debates basados en una participación limitada pueden llegar a excluir conjuntos más amplios de intereses y perspectivas sociales. Por último, respecto a las discusiones actuales, algunas de ellas se olvidan de establecer como punto de partida el grupo de tecnologías "autónomas" más factibles de ser estudiadas, desarrolladas e implementadas en la próxima década. Esta carencia provoca que marcos regulatorios basados en estas discusiones tengan una limitada capacidad previsora.

El GEE propugna una amplia y sistemática participación pública y debates sobre la ética de la IA, la robótica y la tecnología "autónoma". Asimismo, aboga por que se discuta sobre los valores elegidos por las diferentes sociedades para ser integrados en el desarrollo y la gobernanza de estas tecnologías. Este proceso, donde el GEE está listo para desempeñar su papel, debe proporcionar una plataforma para integrar las diversas y amplias iniciativas descritas anteriormente. Además, se debe realizar un debate social amplio, inclusivo y de gran alcance, que contenga las muy diversas perspectivas existentes. De esta forma, aquellos con diferentes conocimientos especializados y distintos valores, podrán ser escuchados. El GEE insta a la Unión Europea a situarse a la vanguardia de dicho proceso y pide a la Comisión Europea que apoye la puesta en marcha e implementación de tales actividades.

Como primer paso para alcanzar la formulación de un conjunto de directrices éticas que sirvan de base para el establecimiento de normas y medidas legislativas globales, **el GEE propone un conjunto de principios éticos fundamentales y prerrequisitos democráticos, basados en los valores establecidos en los Tratados de la UE y en la Carta de Derechos Fundamentales de la UE.**

## Principios éticos y prerrequisitos democráticos

*(a) Dignidad Humana*

El principio de la dignidad humana, el reconocimiento de la condición inherente del ser humano que lo hace digno de respeto, no debe ser violado por las tecnologías "autónomas". Esto implica, por ejemplo, que la toma de decisiones y la clasificación de individuos hechas por algoritmos y sistemas "autónomos" debe ser regulada, especialmente cuando los involucrados ignoran estas prácticas. También implica que tienen que existir límites (legales) para evitar que se le haga creer a las personas que están tratando con seres humanos, cuando en realidad están tratando con algoritmos y máquinas inteligentes. En este contexto es valioso adoptar una concepción relacional de la dignidad humana, que es aquella que se define según nuestras relaciones sociales. De acuerdo a esta concepción, es necesario que conozcamos si estamos interactuando con una máquina u otro ser humano y cuándo ocurre. Además, esta concepción de dignidad requiere que nos reservemos el derecho de decidir si asignamos determinadas tareas a humanos o máquinas.

*(b) Autonomía*

El principio de autonomía implica la libertad del ser humano. Esto se traduce en responsabilidad humana. Para evitar que los sistemas "autónomos" menoscaben la libertad de los seres humanos de establecer sus propios estándares y normas, y de poder vivir de acuerdo con ellos, es necesario tener control y conocimiento sobre ellos. Por consiguiente, todas las tecnologías "autónomas" deben respetar la capacidad humana de elegir si delegarles determinadas decisiones o acciones, cuándo y cómo hacerlo. Esto requiere que los sistemas "autónomos" sean transparentes y previsibles, características sin las cuales sería imposible para los usuarios intervenir o detenerlos cuando lo así lo consideren moralmente necesario.

*(c) Responsabilidad*

El principio de responsabilidad debe ser fundamental en la investigación e implementación de la IA. Los sistemas "autónomos" sólo deberían desarrollarse y aplicarse si sirven al bienestar social y ambiental a nivel global. Establecer dicho bienestar requiere procesos democráticos deliberativos. En otras palabras, los sistemas "autónomos" deben ser diseñados de manera que sus impactos respeten la pluralidad de valores y derechos humanos fundamentales. En vista del gran desafío que supone el potencial abuso de tecnologías "autónomas", es crucial ser conscientes de los riesgos y adoptar el principio de precaución. Desde este punto de vista, las aplicaciones de la IA y la robótica no deben entrañar riesgos inaceptables para los seres humanos. Tampoco deben comprometer la libertad o la autonomía humana al reducir ilegítima y subrepticiamente las opciones o el conocimiento de los ciudadanos. Al contrario, el desarrollo y el uso de estas aplicaciones deberían dirigirse a incrementar el acceso al conocimiento y a las oportunidades para los individuos.

La investigación, el diseño y el desarrollo de la IA, la robótica y los sistemas "autónomos" deben ser guiados por un auténtico interés en la ética de la investigación, en la responsabilidad social de los programadores y en la cooperación académica mundial para proteger derechos y valores humanos fundamentales. Además, estas tecnologías deben ser diseñadas de forma que promuevan esos derechos y valores, evitando a toda costa que más bien los deterioren.

*(d) Justicia, equidad y solidaridad*

La IA debería contribuir a la justicia global y facilitar la igualdad de acceso a los beneficios y ventajas de la IA, la robótica y los sistemas "autónomos". Por ello, los sesgos discriminatorios en los conjuntos de datos utilizados para entrenar y ejecutar los sistemas

de IA, deben evitarse. De no ser posible, estos sesgos deben ser detectados, notificados y neutralizados en la etapa más temprana del proceso.

Necesitamos hacer un esfuerzo global coordinado para alcanzar la igualdad de acceso a las tecnologías "autónomas" y conseguir que la distribución de beneficios y oportunidades sean equitativas. Esta igualdad y equidad debe alcanzarse entre diferentes sociedades y en el seno de cada una de ellas. Para lograr esto, es esencial formular nuevos modelos justos de distribución equitativa y participación en los beneficios. Modelos que sean capaces de responder a las transformaciones del sistema económico ocasionadas por la automatización, la digitalización y la IA. Asimismo, se debe asegurar el acceso a tecnologías básicas de IA y facilitar la formación en ciencia, tecnología, ingeniería, matemática y disciplinas digitales. Todo lo anterior es de especial importancia cuando se trata de regiones o grupos sociales desfavorecidos. Además, es necesario estar atentos a los impactos negativos de la acumulación creciente y masiva de datos personales. Es importante mencionar la presión que puede ejercer sobre el concepto de solidaridad, por ejemplo, sobre los sistemas de asistencia mutua como el seguro social y la asistencia sanitaria. Puede inclusive llegar a socavar la cohesión social y dar lugar a un individualismo radical.

*(e) Democracia*

Las decisiones clave sobre a la regulación de la IA, específicamente sobre su desarrollo y aplicaciones, deben ser el resultado de procesos de debate democrático y participación ciudadana. Al respecto, un espíritu de cooperación global y procesos de diálogo público asegurarán que estas decisiones sean inclusivas, informadas y con visión de futuro. Garantizar el derecho a la educación y a la información sobre las nuevas tecnologías y sus implicaciones éticas, facilitará que todos comprendan los riesgos y oportunidades en juego. Asimismo, facultará al público para participar en los procesos de toma de decisiones que son cruciales para construir nuestro futuro.

El derecho de los seres humanos a la autodeterminación a través de medios democráticos es central a la dignidad humana y a la autonomía. Además, para nuestros sistemas políticos democráticos son de vital importancia el pluralismo como valor, la diversidad y la incorporación de las diferentes concepciones de lo que es tener una vida buena. Las nuevas tecnologías no deben poner en peligro a los ciudadanos, despojarlos de sus derechos o de su individualidad. Tampoco deben inhibir o influir en la toma de decisiones políticas, infringir la libertad de expresión y el derecho a recibir y difundir información sin interferencia. Al contrario, estas tecnologías deberían ser herramientas para beneficiarnos de la inteligencia colectiva, y para apoyar y mejorar los procesos cívicos de los que dependen nuestras sociedades democráticas.

*(f) Estado de derecho y rendición de cuentas*

El estado de derecho, el acceso a la justicia y el derecho de recibir una compensación y un juicio justo, proporcionan el marco necesario para garantizar la observancia de las normas de derechos humanos. Asimismo, estos proveen los mecanismos para el desarrollo de eventuales regulaciones específicas para la IA. Esto incluye la protección contra la violación de los derechos humanos por parte de los sistemas "autónomos", por ejemplo la seguridad o la privacidad.

Los desafíos legales prácticos deben abordarse con un esfuerzo oportuno para desarrollar soluciones sólidas que asignen responsabilidades de manera clara y justa y para establecer una legislación vinculante eficiente.

En este sentido, los gobiernos y las organizaciones internacionales deben incrementar sus esfuerzos para establecer en quién recae la responsabilidad de los daños causados por el desempeño no deseado de los sistemas "autónomos". Asimismo, deben instituirse sistemas efectivos de mitigación de daños.

*(g) Seguridad, protección, e integridad física y mental*

La seguridad y la protección de los sistemas "autónomos" se concretan en: (1) la seguridad externa, que se ofrece al entorno y a los usuarios, (2) la confiabilidad y la robustez interna, por ejemplo contra la piratería y (3) la seguridad emocional, que se refiere a la interacción humano-máquina. Estas tres dimensiones de la seguridad y la protección deben ser tomadas en cuenta por los desarrolladores de IA y deben ser estrictamente evaluadas antes del lanzamiento de cualquier sistema "autónomo". Esto con el fin de garantizar que estos sistemas no infrinjan el derecho de los seres humanos a la integridad física y mental, y a un entorno seguro. Se debe prestar especial atención a aquellas personas en posiciones vulnerables, así como al posible doble uso y a la militarización de la IA. Por ejemplo, en los campos de ciberseguridad, finanzas, infraestructura y conflictos armados.

*(h) Protección de datos y privacidad*

En una era de recopilación generalizada y masiva de datos a través de tecnologías digitales de la comunicación, el derecho a la protección de la información personal y el derecho a la privacidad están siendo decisivamente cuestionados. En efecto, la IA debe respetar las regulaciones de protección de datos. Esto tanto si la IA se concreta en la forma de robots físicos, aquellos que forman parte del internet de las cosas, o en la forma de *softbots*, aquellos que operan a través de la red informática mundial. Ni robots ni *bots* deben recopilar o difundir datos, ni ser ejecutados en conjuntos de datos para los que estas actividades no han sido consentidas.

Los sistemas "autónomos" no deben interferir en el derecho a la vida privada. Esto incluye el derecho a estar libres de tecnologías que influyan en las opiniones y el desarrollo personal, el derecho a establecer y desarrollar relaciones con otros seres humanos, y el derecho a estar libres de vigilancia. También se deben definir criterios precisos y establecer los mecanismos apropiados para asegurar que el desarrollo y la aplicación de los sistemas "autónomos" sean éticamente correctos.

Los posibles efectos de los sistemas "autónomos" en la vida privada y la privacidad generan gran preocupación. A la luz de tales inquietudes, es valioso considerar el debate actual sobre la introducción de dos nuevos derechos: el derecho al contacto humano significativo y el derecho a no ser perfilado, medido, analizado, aconsejado (en inglés *coached*) o provocado (en inglés *nudged*).

*(i) Sostenibilidad*

La tecnología de IA debe responder a la responsabilidad humana de garantizar los prerrequisitos fundamentales para la vida en nuestro planeta, la continua prosperidad de la humanidad y la conservación del medioambiente para las generaciones futuras. Las estrategias para evitar que las futuras tecnologías afecten negativamente la vida humana y la naturaleza necesitan fundamentarse en políticas que prioricen la protección del medio ambiente y la sostenibilidad.

La inteligencia artificial, la robótica y los sistemas "autónomos" son capaces de traer prosperidad, contribuir a la calidad de vida, y ayudar a alcanzar los ideales morales y los objetivos socioeconómicos europeos. Lo anterior es posible, solo si estas tecnologías se diseñan y se implementan de forma sensata. Para construir un mundo futuro que haga realidad estos propósitos, es necesario tomar en cuenta las consideraciones éticas y los valores morales compartidos. Estos deben ser interpretados como estímulos y oportunidades para la innovación, no como obstáculos y barreras.

El GEE hace un llamado a la Comisión Europea para que investigue los instrumentos jurídicos existentes disponibles para enfrentarse efectivamente a los problemas discutidos en esta declaración y si nuevos instrumentos normativos y de gobernanza son requeridos.

Por último, el GEE insta al lanzamiento de un proceso que allane el camino hacia un marco ético y legal común e internacional para el diseño, la producción, el uso y la gobernanza de la inteligencia artificial, la robótica y los sistemas "autónomos".

## ¿Cómo contactar a la UE?

**EN PERSONA**

A lo largo de toda la Unión Europea hay cientos de centros de información *Europe Direct* a su disposición. La dirección de la oficina más cercana se puede encontrar en la siguiente dirección: **http://europa.eu/contact.**

**CONTACTO TELEFÓNICO O POR CORREO ELECTRÓNICO**

*Europe Direct* es un servicio que brinda respuestas a sus preguntas sobre la Unión Europea.

Usted puede contactar este servicio:

– llamando al siguiente teléfono gratuito: **00 800 6 7 8 9 10 11** (es posible que algunos operadores cobren una tarifa por este servicio);

– llamando al siguiente número estándar: **+32 22999696**;

– por correo electrónico disponible en la dirección: **http://europa.eu/contact.**

## ¿Cómo encontrar más información sobre la UE?

**EN LÍNEA**

Información sobre la Unión Europea está disponible en todas sus lenguas oficiales, en la página web Europa: **http://europa.eu.**

**PUBLICACIONES DE LA UE**

Usted puede descargar u hacer pedidos de publicaciones gratuitas o publicaciones la venta a través del sitio del EU *Bookshop*: **http://bookshop.europa.eu.** Además, es posible obtener múltiples copias de las publicaciones gratuitas contactando a las oficinas de *Europe Direct* o a su centro de información más cercano (http://europa.eu/contact).

**NORMATIVA DE LA UE Y DOCUMENTOS RELACIONADOS**

Para obtener información sobre la normativa de la UE, incluyendo la legislación completa desde 1951 en todas las lenguas oficiales, favor visitar el sitio EUR-Lex: **http://eur-lex.europa.eu.**

**INFORMACIÓN DE LIBRE ACCESO DE LA UE**

El portal de información de libre acceso de la UE (**http://data.europa.eu/euodp/en/data**) facilita el acceso a bases de datos de la UE. La información puede ser descargada y reutilizada de manera gratuita, tanto para fines comerciales como no comerciales.

*Los avances en inteligencia artificial, robótica y las llamadas tecnologías "autónomas" han originado una serie de dilemas morales cada vez más urgentes y complejos. Actualmente se están haciendo esfuerzos para orientar estas tecnologías hacia el bien común y para encontrar soluciones a los desafíos éticos, sociales y legales que generan. Sin embargo, estos esfuerzos han resultado ser un mosaico de iniciativas dispares. Esta situación genera la necesidad de implementar un proceso colectivo, amplio e inclusivo de reflexión y diálogo. Este diálogo debe estar basado en los valores en los que queremos fundamentar nuestra sociedad y en el papel que queremos que la tecnología desempeñe.*

*Esta declaración hace un llamado para iniciar la construcción de un marco ético y legal común e internacionalmente reconocido para el diseño, producción, uso y gobernanza de la inteligencia artificial, la robótica y los sistemas "autónomos". Además, esta declaración propone un conjunto de principios éticos fundamentales que pueden servir de guía para el desarrollo de este marco ético y legal. Estos principios están basados en los valores establecidos en los Tratados de la UE y en la Carta de Derechos Fundamentales de la UE.*

*Política de Investigación e Innovación*