

# **Reporte sobre el primer borrador de la recomendación acerca de la ética en la Inteligencia Artificial**

Laboratorio de Inteligencia Artificial Microsoft  
Departamento de Computación  
División de Ingeniería Eléctrica  
Facultad de Ingeniería  
Universidad Nacional Autónoma de México

Octubre de 2020

## **Introducción**

En la sesión número 40 de la Conferencia General de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), celebrada en noviembre de 2019, se acordó la elaboración de un instrumento normativo sobre la ética en el campo de la Inteligencia Artificial (IA). Este instrumento será publicado en forma de recomendación para los estados miembros de la UNESCO.

UNESCO, a través de su Directora Audrey Azoulay, ha establecido a varios grupos de expertos, de las áreas correspondientes, para la elaboración del instrumento normativo sobre la ética en la IA. UNESCO también ha organizado una serie de eventos de consulta con este mismo propósito. Como resultado de estos esfuerzos, actualmente se cuenta con un primer borrador de este instrumento normativo.

De acuerdo a la Constitución de la UNESCO, las recomendaciones emitidas por la Conferencia General, deberán considerar observaciones y comentarios de los estados miembros. Como parte de este proceso, el presente documento contiene un resumen y comentarios sobre el primer borrador de la recomendación.

La estructura general del documento es la siguiente: primero se describen los objetivos generales de la recomendación; a continuación se presenta un resumen de cada una de las 10 áreas en que están agrupadas las políticas de acción, descritas en la recomendación. También se incluyen comentarios para cada área.

## **Objetivos**

Los objetivos generales de la recomendación son:

- Proveer un marco universal de valores, principios y acciones que sirvan de guía a los estados miembros, para la elaboración de leyes, políticas u otros instrumentos concernientes a la IA.
- Guiar las acciones de individuos, grupos, comunidades, instituciones y compañías del sector privado, para asegurar la implementación de la ética en todas las etapas del ciclo de vida de los sistemas de IA.
- Promover el respeto por la dignidad humana y la igualdad de género, para salvaguardar los intereses de la presente y futuras generaciones, y para proteger los derechos humanos, libertades fundamentales, el medio ambiente y los ecosistemas en todas las etapas del ciclo de vida de los sistemas de IA.
- Fomentar un diálogo plural y multidisciplinario sobre aspectos éticos en la IA.

- Promover acceso equitativo a los avances y conocimiento en el campo de la IA, y la distribución de beneficios, con especial atención a las necesidades y contribuciones de países con ingresos medios y bajos, países en desarrollo, países en desarrollo sin litorales y pequeños estados insulares.

En cuanto a los valores descritos por la recomendación, se mencionan el respeto, protección y promoción de la dignidad humana, de los derechos humanos y de las libertades fundamentales. También se describen los valores concernientes a la prosperidad del medio ambiente, la diversidad e inclusión, así como la vida en armonía y paz.

Los principios incluidos en el documento de recomendación son: proporcionalidad y no agresión, seguridad y protección, justicia y no discriminación, sustentabilidad, privacidad, transparencia y explicabilidad, sobre-vigilancia, responsabilidad y fiscabilidad, entre otros.

Los valores y principios están descritos en la generalidad pertinente a la diversidad de los estados pertenecientes a la UNESCO, también consideran la diferencias, fortalezas y limitaciones de los mismos.

## **Políticas de acción**

A continuación se resumen y comentan cada una de las diez áreas que agrupan las políticas de acción que se proponen en el documento de recomendación.

### **Área 1. Evaluación del impacto ético**

Se proponen cinco políticas de acción: i) identificar los beneficios y riesgos de los sistemas de IA; ii) tanto los estados como la industria privada deben prevenir y mitigar los impactos negativos de los sistemas de IA en los derechos humanos; iii) desarrollar e implementar las medidas pertinentes para monitorear todas las fases del ciclo de vida de los sistemas de IA; iv) adoptar marcos regulatorios para evaluar el impacto ético de los sistemas de IA; y v) establecer mecanismos de evaluación sobre políticas relevantes a la ética en la IA. La pertinencia de las políticas propuestas sobre la evaluación del impacto que tienen los sistemas de IA en aspectos éticos es indiscutible, sin embargo, no se describe la forma de implementarlas. Esto puede tener como objetivo no perder la generalidad requerida por este tipo de recomendación. Con el objeto de coadyuvar a la reducción de la brecha de capacidades de implementación de estas políticas por parte de los Estados, se pudieran describir formas de puesta en funcionamiento en particular para los estados más desfavorecidos y citados en el último objetivo general de la recomendación.

### **Área 2. Administración y gobernanza ética**

Esta área incluye 17 políticas de administración pública, entre las que se encuentran: los mecanismos de gobernanza deben ser transparentes, inclusivos, multidisciplinarios; los Estados deben investigar, sancionar y asegurar el resarcimiento de los daños causados por los sistemas de IA; fomentar el desarrollo de ecosistemas digitales para el diseño e implementación éticos de sistemas de IA; entre otras. Todas estas políticas relacionadas con la gobernanza para el desarrollo ético de sistemas de IA, son oportunas y convenientes. Más aún, podrían extenderse a sistemas de tecnologías de la información de manera general y no solamente a sistemas de IA.

### **Área 3. Política de datos**

Las políticas de gobernanza de datos deben incluir la evaluación continua de la calidad de los datos usados para entrenar sistemas basados en Inteligencia Artificial. En estas políticas se deben integrar todos los procesos dentro del ciclo de vida de la IA. También deben tomarse en cuenta la seguridad de estos datos, así como su protección especial cuando se trata de datos personales y sensibles. Se incluye el derecho a la privacidad, los derechos humanos,

transparencia de procesos y su consecuente deslinde de responsabilidades. De la misma manera, se hace énfasis en la promoción de políticas de datos abiertos y acceso a la información y conocimiento, esto cuidando que dicha información sea de calidad, robusta y confiable. Finalmente, se propone que la información sea anotada de forma estándar para incrementar la interoperabilidad de herramientas, interfaces y datos, esto tanto para el sector público como para el privado.

#### **Área 4. Desarrollo y cooperación internacional**

Se debe dar prioridad a la ética dentro de la Inteligencia Artificial, incluyendo discusiones abiertas sobre este tema dentro de los sectores público, privado y académico. De esta manera se debe asegurar que el uso de la IA se adhiera a los valores y principios de esta recomendación. Esto aplica en la cooperación internacional entre los países miembros incluyendo apoyo en el desarrollo de esta tecnología y promoviendo colaboraciones en investigación y desarrollo, tanto de Inteligencia Artificial como de la ética aplicada a ésta. Finalmente, se requiere del desarrollo y la implementación de un marco legal para que la descrita cooperación internacional sea una realidad.

#### **Área 5. Medio ambiente y ecosistemas**

Se debe evaluar el impacto ambiental directo e indirecto del desarrollo y uso de Inteligencia Artificial, incluyendo el ciclo de vida completo de éste. De esta manera, se garantizará el cumplimiento de las leyes relacionadas con el medio ambiente por parte de todos los actores. Para esto se propone introducir incentivos que apoyen a garantizar el desarrollo y adopción de soluciones basadas en inteligencia artificial que sean éticas y consideren el cuidado del medio ambiente. Estas soluciones incluyen protección del medio ambiente, resistencia al riesgo de desastres y preservación del planeta teniendo en cuenta a las poblaciones indígenas y buscando soportar la economía circular.

#### **Área 6. Género**

Se debe asegurar que las tecnologías digitales y la Inteligencia Artificial contribuyan a lograr la equidad de género. De tal manera, los derechos y libertades fundamentales de las niñas y las mujeres, incluyendo su seguridad e integridad, no deben ser violadas en ninguna etapa del desarrollo de sistemas basados en IA. Se propone dedicar fondos especiales a financiar esquemas de género para asegurarse que las niñas y las mujeres no queden fuera del desarrollo generado por la economía digital soportada por la IA. Los sistemas desarrollados deben mejorar la situación actual de la equidad de género apoyando a disminuir la brecha, el acceso a tecnologías, la distribución del trabajo no pagado y las responsabilidades en cuidados. Para esto se debe garantizar que los sistemas creados no integren estereotipos de género o actitudes discriminatorias, así como cualquier tipo de violencia. Finalmente, el emprendimiento femenino debe ser propiciado generando políticas que permitan una participación balanceada de ambos géneros en las áreas de investigación y académicas, así como en posiciones de toma de decisiones tanto en ambientes públicos como privados. Esto incluye políticas de libre acoso en todos los sectores y niveles.

#### **Área 7. Cultura**

Con respecto a la Política de Cultura, es positivo que se aliente la incorporación de sistemas de Inteligencia Artificial, siempre y cuando se tenga el compromiso de la preservación, promoción y accesibilidad del patrimonio cultural, tangible o intangible. Además, el que se busque la preservación de los idiomas, lenguas y dialectos en peligro de extinción es otro de puntos clave, donde la Inteligencia Artificial actual puede aportar de manera significativa, ya sea a través de aplicaciones de procesamiento del lenguaje natural, traducción automática y asistentes de voz. Resalta también la promoción de la educación en Inteligencia Artificial y la formación digital para artistas y profesionales que crean, producen, distribuyen, preservan y difunden diversos

bienes y servicios culturales. Además, es importante que se aliente a los museos, galerías y bibliotecas a utilizar sistemas de Inteligencia Artificial para destacar sus colecciones, fortalecer sus bases de datos y otorgarles acceso a sus usuarios.

### **Área 8. Educación e investigación**

Sobre la política de educación e investigación, se destaca el que se busque fomentar la adopción de sistemas de Inteligencia Artificial para la alfabetización de las personas y la reducción de las desigualdades al acceso digital. Esto podría propiciar en la población un pensamiento crítico, el que se trabaje en equipo y se adquieran nuevas habilidades emocionales y éticas. Se puede lograr también la concientización sobre nuevos desarrollos, oportunidades y desafíos que plantean las nuevas tecnologías de Inteligencia Artificial. Por otro lado, fomentar las iniciativas de investigación sobre el uso responsable de la IA a través de la enseñanza, formación docente y el aprendizaje es otro de los puntos necesarios para la mejora de oportunidades y la mitigación de los desafíos y riesgos involucrados en este creciente campo de conocimiento. Además, promover una investigación ética, científica y rigurosa es otro de los aportes necesarios para tener productos y publicaciones de calidad. Se rescata también el que se promueva la diversidad de género en la investigación, esto con el propósito de luchar contra los estereotipos de género y el acoso dentro de la comunidad de investigación de Inteligencia Artificial.

### **Área 9. Economía y trabajo**

Sin lugar a duda, esta iniciativa que invita a evaluar el impacto de los sistemas de Inteligencia Artificial en los mercados laborales y sus implicaciones para definir nuevos requisitos de educación, puede abrir una amplia gama de habilidades interdisciplinarias que se requiere para comunicar, tener un pensamiento crítico, trabajar en equipo y transferir el conocimiento en torno a nuevos desafíos de la Inteligencia Artificial. Para abordar estos requisitos de habilidades será necesario crear o adecuar programas y estrategias de capacitación para nuevos puestos de trabajo que demanda la industria. Además, es importante que se promueva la enseñanza y aprendizaje del desarrollo de proyectos de Inteligencia Artificial a través de la articulación entre empresas del sector privado, universidades y centros de investigación. Por otra parte, el que se aliente y apoye prácticas de readaptación y reasignación laboral, ante la automatización a través de la Inteligencia Artificial, es otro punto importante que se necesita para contrarrestar las consecuencias de posibles pérdidas de empleos.

### **Área 10. Salud y bienestar social**

La promoción del empleo de sistemas de Inteligencia Artificial para el mejoramiento de la salud y protección del derecho a la vida es sumamente importante, sobre todo para hacer frente a riesgos en la salud local, regional y mundial, como está ocurriendo ahora con COVID-19. Además, el que se busque considerar las relaciones del paciente con su familia y con el personal de atención médica, será clave para garantizar una atención de salud de calidad. Por otra parte, el que se anime a prestar especial atención a la prevención, detección y tratamiento oportuno de la salud será clave para garantizar una supervisión acertada, prestar la debida atención al paciente a través de un seguimiento constante, asegurar que los datos estén siendo utilizados de manera correcta, y asegurar que la decisión final de diagnóstico y tratamiento sean tomados por humanos, reconociendo que los sistemas de Inteligencia Artificial también pueden ayudar en este trabajo.