United Nations
Educational, Scientific and
Cultural Organization

Organisation
des Nations Unies
pour l'éducation,
la science et la culture

Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura

Организация
Объединенных Наций по
вопросам образования,
науки и культуры

منظمة الأمم المتحدة
للتربية والعلم والثقافة

联合国教育、
科学及文化组织

Conferencia General
40ª reunión - París, 2019

40 C

40 C/67
30 de julio de 2019
Original: inglés

Punto **5.24** del orden del día provisional

ESTUDIO PRELIMINAR SOBRE UN POSIBLE INSTRUMENTO NORMATIVO RELATIVO A LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

PRESENTACIÓN

Fuente: decisión 206 EX/42.

Antecedentes: en cumplimiento de la decisión 206 EX/42 y de conformidad con el Reglamento sobre las recomendaciones a los Estados Miembros y las convenciones internacionales previstas en el párrafo 4 del artículo IV de la Constitución, la Directora General presenta el estudio preliminar de los aspectos técnicos y jurídicos relativos a la conveniencia de disponer de un instrumento normativo sobre la ética de la inteligencia artificial.

Objeto: en el presente documento figuran el estudio preliminar de los aspectos técnicos y jurídicos relativos a la conveniencia de disponer de un instrumento normativo sobre la ética de la inteligencia artificial y los comentarios y observaciones del Consejo Ejecutivo al respecto.

Decisión requerida: párrafo 11.

Job: 1909935

I. ANTECEDENTES

1. El estudio preliminar de los aspectos técnicos y jurídicos relativos a la conveniencia de disponer de un nuevo instrumento normativo sobre la ética de la inteligencia artificial se presentó al Consejo Ejecutivo en su 206ª reunión (documento 206 EX/42), de conformidad con el artículo 3 del Reglamento sobre las recomendaciones a los Estados Miembros y las convenciones internacionales previstas en el párrafo 4 del artículo IV de la Constitución. El texto completo del estudio preliminar se reproduce en el anexo I del presente documento.

II. OBSERVACIONES Y DECISIONES DEL CONSEJO EJECUTIVO

2. Al examinar el estudio preliminar, el Consejo Ejecutivo expresó su reconocimiento a la Directora General por haber introducido en la UNESCO una cuestión tan de actualidad como la ética de la inteligencia artificial. Los miembros del Consejo Ejecutivo acogieron con beneplácito el estudio preliminar sobre la ética de la inteligencia artificial preparado por el grupo de trabajo ampliado de la COMEST sobre la ética de la inteligencia artificial y subrayaron que la inteligencia artificial plantea algunas cuestiones éticas importantes.

3. El Consejo Ejecutivo observó que las tecnologías de inteligencia artificial se están desarrollando muy rápidamente y tienen un inmenso potencial para redundar en beneficio de la humanidad, pero que, al mismo tiempo, existe incertidumbre y se plantea un gran número de nuevos desafíos y problemas éticos en relación con la forma de vida y de desarrollo de nuestras sociedades. El Consejo señaló que las tecnologías de inteligencia artificial no son neutrales, sino que pueden estar intrínsecamente sesgadas en función de los datos con los que hayan sido alimentadas. El Consejo subrayó que era importante asegurarse de que las tecnologías de inteligencia artificial se desarrollen con normas éticas claras para que respeten la dignidad y los derechos humanos. También señaló que no existe un marco ético internacionalmente aceptado en relación con todas las aplicaciones e innovaciones de la inteligencia artificial. Teniendo en cuenta que las tecnologías de inteligencia artificial pueden transformar el futuro de todas las esferas del mandato de la UNESCO, el Consejo Ejecutivo llegó a la conclusión de que era pertinente y oportuno que la Organización iniciara la elaboración de un instrumento normativo sobre la ética de la inteligencia artificial en forma de recomendación.

4. Muchos miembros del Consejo Ejecutivo destacaron el valor añadido de elaborar una recomendación de la UNESCO en este ámbito. Por una parte, además de enunciar unos principios básicos, una recomendación tendría repercusiones en la formulación de políticas, contribuyendo a potenciar y fortalecer la capacidad de los Estados para intervenir, cuando sea necesario, en ámbitos importantes que se ven afectados por el desarrollo de la inteligencia artificial, como la cultura, la educación, las ciencias y la información y la comunicación. Por otra parte, a través de su mecanismo de presentación de informes, una recomendación podría ser importante para el desarrollo de capacidades, el intercambio de conocimientos y la determinación de buenas prácticas.

5. Se subrayó la necesidad de determinar el alcance y la orientación adecuados de la labor de la UNESCO en relación con las cuestiones éticas en los ámbitos de la inteligencia artificial para la paz, el desarrollo sostenible y el bienestar humano, y de evitar la duplicación de los esfuerzos realizados, en particular, por otros órganos del sistema de las Naciones Unidas. Para ello, el Consejo Ejecutivo pidió a la Directora General que le presentara, en su 207ª reunión, un informe sobre la labor de otras organizaciones y convenciones internacionales en relación con diferentes aspectos de la inteligencia artificial.

6. Se hizo especial hincapié en la importancia de los vínculos entre la inteligencia artificial y los objetivos de desarrollo acordados internacionalmente en la Agenda 2030 para el Desarrollo Sostenible. Se subrayó que la reflexión actual y futura sobre la ética de la inteligencia artificial debería tener en cuenta el impacto de las tecnologías de inteligencia artificial en la perspectiva de alcanzar los 17 Objetivos de Desarrollo Sostenible (ODS). Asimismo, se señaló que se debe prestar especial atención a las preocupaciones de los países en desarrollo con respecto a las brechas

digital, tecnológica y de conectividad, que podrían verse exacerbadas por las tecnologías de inteligencia artificial, lo que daría lugar a un aumento de las desigualdades existentes entre los países; de ahí la necesidad de garantizar un acceso abierto y equitativo al conocimiento, los datos, los resultados de la investigación, las tecnologías innovadoras y el desarrollo de capacidades en el ámbito de la inteligencia artificial.

7. Además, se señaló que todos los países y pueblos, independientemente de su nivel de desarrollo, tienen la responsabilidad común de maximizar los beneficios y minimizar los riesgos de las tecnologías de inteligencia artificial. Se subrayó en particular la importancia de trabajar de manera inclusiva, multipartita, multidisciplinaria y claramente intergubernamental, de conformidad con los procedimientos establecidos. El Consejo Ejecutivo destacó la importancia de trabajar en cooperación con todos los Estados Miembros, las organizaciones intergubernamentales mundiales y regionales, las organizaciones internacionales no gubernamentales, las asociaciones profesionales, la sociedad civil, el sector privado y otras partes interesadas pertinentes para garantizar un carácter verdaderamente inclusivo y multidisciplinario del proceso relacionado con la elaboración de un nuevo instrumento normativo. Se hizo hincapié en la necesidad de convocar un número suficiente de reuniones de expertos intergubernamentales, con su presencia personal en la Sede de la UNESCO. En particular, se subrayó la necesidad de que participaran los países africanos, así como expertos de los países en desarrollo. Además, se recalcó la necesidad de tener en cuenta los contextos y jurisdicciones nacionales y las diversas capacidades de cada país para aplicar las tecnologías de inteligencia artificial mientras se da forma al debate sobre la ética de la inteligencia artificial. Se reiteró que las cuestiones relacionadas con la igualdad de género deben considerarse prioritarias.

8. Se subrayó en general que los métodos de trabajo y el calendario deben ajustarse plenamente al Reglamento sobre las recomendaciones a los Estados Miembros y las convenciones internacionales previstas en el párrafo 4 del artículo IV de la Constitución. Muchos miembros subrayaron el carácter estatutario de la labor normativa y destacaron la importancia de utilizar el presupuesto ordinario para la elaboración de un instrumento normativo sobre la ética de la inteligencia artificial, al tiempo que acogieron con beneplácito la aportación de contribuciones voluntarias. Varios miembros del Consejo Ejecutivo recordaron el apoyo que sus países ya habían prestado para promover la reflexión de la UNESCO sobre la ética de la inteligencia artificial y declararon que seguirían apoyando este ámbito de trabajo de la Organización.

9. Por último, el Consejo Ejecutivo aprobó la decisión 206 EX/42, cuyo texto completo se reproduce en el anexo II del presente documento.

III. HOJA DE RUTA PROPUESTA

10. En caso de que la Conferencia General decida proceder a la elaboración de una recomendación, la hoja de ruta propuesta para la preparación de un instrumento normativo sobre la ética de la inteligencia artificial figura en el anexo III del presente documento.

IV. PROYECTO DE RESOLUCIÓN

11. Habida cuenta de lo que antecede, la Conferencia General podría aprobar el siguiente proyecto de resolución:

> La Conferencia General,
>
> *Habiendo examinado* el documento 40 C/67,
>
> *Recordando* la decisión 206 EX/42,
>
> *Reconociendo* las preocupaciones por la creciente brecha digital y tecnológica entre los países, que podría verse exacerbada por la inteligencia artificial, y *reiterando* la importancia de atender las preocupaciones de los países en desarrollo respecto a la inteligencia artificial,

en particular mediante la transferencia de tecnología de inteligencia artificial, el desarrollo de capacidades y la educación en materia de inteligencia artificial, la divulgación de datos y el acceso a los datos,

Reconociendo también que, si bien la inteligencia artificial tiene potencial para transformar el futuro de la humanidad para mejor y en favor del desarrollo sostenible, también existe una conciencia generalizada de los riesgos y desafíos que conlleva, especialmente por lo que respecta a la agravación de las desigualdades y brechas existentes, así como las implicaciones para los derechos humanos,

Reiterando la necesidad de reforzar la colaboración internacional en el ámbito de la promoción de una inteligencia artificial con valores humanos en las esferas de la educación, las ciencias, la cultura, la información y la comunicación en beneficio de las generaciones presentes y futuras,

Reconociendo además que una recomendación podría constituir una herramienta fundamental para fomentar la elaboración de textos legislativos, políticas y estrategias nacionales e internacionales en el ámbito de la inteligencia artificial y reforzar su aplicación, así como para potenciar la cooperación internacional en torno al desarrollo y el uso éticos de la inteligencia artificial en apoyo de los Objetivos de Desarrollo Sostenible (ODS),

1. *Decide* que es oportuno y pertinente que la UNESCO prepare un instrumento normativo internacional sobre la ética de la inteligencia artificial en forma de recomendación;

2. *Pide* a la Directora General que vele por que se celebre un número suficiente de consultas intergubernamentales presenciales sobre el texto de la recomendación mencionada;

3. *Invita* a la Directora General a que le presente, en su 41ª reunión, el proyecto de recomendación sobre la ética de la inteligencia artificial de conformidad con el Reglamento sobre las recomendaciones a los Estados Miembros y las convenciones internacionales previstas en el párrafo 4 del artículo IV de la Constitución.

ANEXO I

ESTUDIO PRELIMINAR SOBRE LOS ASPECTOS TÉCNICOS Y JURÍDICOS RELATIVOS A LA CONVENIENCIA DE DISPONER DE UN INSTRUMENTO NORMATIVO SOBRE LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

(Documento 206 EX/42)

I. LA UNESCO, UN LABORATORIO DE IDEAS

1. La inteligencia artificial[1] (IA) es uno de los temas centrales de la era de las tecnologías convergentes, con profundas repercusiones para los seres humanos, las culturas, las sociedades y el medio ambiente. Es probable que la IA transforme el futuro de la educación, las ciencias, la cultura y la comunicación, que son todas esferas del mandato de la UNESCO.

2. Si bien la IA encierra el potencial de mejorar el futuro de la humanidad y promover el desarrollo sostenible, también existe una conciencia generalizada de los riesgos y desafíos asociados con la misma, especialmente en lo que se refiere a la agravación de las desigualdades y brechas existentes, así como las implicaciones para los derechos humanos. A fin de esbozar posibles escenarios y liberar el potencial de la IA, con miras a aprovechar las oportunidades en materia de desarrollo, a la vez que se gestionan los riesgos, es importante fomentar una comprensión más amplia de la forma en que las tecnologías disruptivas, como la IA, transforman las sociedades.

3. Este trabajo debe ir acompañado de una reflexión ética, ya que las tecnologías de IA no son neutrales, sino que están intrínsecamente sesgadas por los datos en los que se basan y las decisiones que se toman durante la integración de esos datos. De igual forma, las decisiones que toman las máquinas inteligentes no pueden predecirse totalmente. Por otra parte, como la IA es una tecnología distribuida, cuya gobernanza actual la ejercen al tiempo numerosas instituciones, organizaciones y empresas, la reflexión sobre su buena gobernanza requiere un enfoque pluralista, multidisciplinario, multicultural y multipartito, que plantee interrogantes sobre el tipo de futuro que queremos para la humanidad. Esta reflexión debe abordar los principales desafíos derivados del desarrollo de tecnologías de IA en cuanto a los sesgos incorporados en los algoritmos, incluidos los prejuicios de género, la protección de la privacidad de las personas y los datos personales, los riesgos de crear nuevas formas de exclusión y desigualdad, las cuestiones vinculadas a la distribución justa de los beneficios y los riesgos, la rendición de cuentas, la responsabilidad, las consecuencias para el empleo y el futuro del trabajo, la dignidad y los derechos humanos, la seguridad y los riesgos del doble uso.

4. Con el fin de aprovechar el potencial de la IA para el bien de la humanidad, la UNESCO acogió en París una serie de reuniones en 2018, como por ejemplo el seminario sobre inteligencia artificial organizado conjuntamente por las Delegaciones Permanentes del Brasil y Turquía ante la UNESCO (junio de 2018), un debate abierto sobre el tema Aprovechar la inteligencia artificial para fomentar las sociedades del conocimiento y la buena gobernanza, celebrado en la Fundación Mozilla (noviembre de 2018), y el debate sobre La IA en favor de los derechos humanos y los ODS: fomentar enfoques multipartitos, inclusivos y abiertos, llevado a cabo como parte del Foro para la Gobernanza de Internet (noviembre de 2018). En cooperación con los asociados de la UNESCO sobre el terreno, se celebraron varios actos en distintas regiones, como el Primer Foro sobre Inteligencia Artificial e Internet de las Cosas en Ciudades Inteligentes Sostenibles (Buenos Aires (Argentina), mayo de 2018). La reflexión en curso de la UNESCO sobre el diálogo mundial en torno a los aspectos éticos de la IA, centrada en las normas y directrices, también fue tratada en el Foro sobre IA en África llevado a cabo en diciembre de 2018 en Benguérir (Marruecos), en el que los participantes pidieron, por ejemplo, la formulación de una estrategia de IA para África, así como una IA centrada en la

[1] Aunque no existe una definición única de "inteligencia artificial" (IA), el debate tiende a centrarse en las máquinas capaces de imitar ciertas funcionalidades de la inteligencia humana, incluidas características como la percepción, el aprendizaje, el razonamiento, la resolución de problemas, la interacción lingüística e incluso la producción de un trabajo creativo.

dimensión humana. La acción global prevista de la UNESCO en materia de inteligencia artificial se presentó a los Estados Miembros en la reunión de información celebrada el 22 de enero de 2019, a la que siguió el debate internacional de expertos sobre El futuro de las tecnologías: ¿esperanza o temor? Esta reflexión se profundizó el 4 de marzo de 2019 en la Sede de la UNESCO, en el marco de una conferencia mundial titulada Principios de la inteligencia artificial: ¿hacia un enfoque humanista? y la Semana del aprendizaje móvil de la UNESCO 2019 (4-8 de marzo), que tuvo como tema central la IA y el desarrollo sostenible. En mayo de 2019, la Conferencia internacional sobre IA y educación se realizará en Beijing (China).

II. COMPETENCIAS DE LA UNESCO EN MATERIA DE ÉTICA DE LA CIENCIA Y TECNOLOGÍA, ASÍ COMO DE BIOÉTICA

5. En la Estrategia a Plazo Medio de la UNESCO para 2014-2021 (37 C/4) se pone de relieve que el establecimiento de normas es una de las funciones esenciales de la Organización. La UNESCO tiene una larga experiencia en la labor normativa mundial, y ha emprendido una reflexión sobre la bioética y la ética de la ciencia y la tecnología, por conducto de su Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST) y el Comité Internacional de Bioética (CIB). La UNESCO ha desempeñado un papel destacado en el establecimiento de normas y la cooperación en el plano internacional. Ciertamente, desde hace 25 años facilita un diálogo entre múltiples interesados y elabora recomendaciones normativas sobre la ética de la ciencia y la tecnología, como la Declaración Universal sobre el Genoma Humano y los Derechos Humanos (1997), la Declaración Internacional sobre los Datos Genéticos Humanos (2003), la Declaración Universal sobre Bioética y Derechos Humanos (2005), la Declaración de Principios Éticos en relación con el Cambio Climático (2017) y la Recomendación sobre la Ciencia y los Investigadores Científicos (2017). La COMEST y el CIB han reflexionado sobre cuestiones ligadas a la IA. En 2017, el CIB aprobó su informe sobre los macrodatos y la salud, mientras que la COMEST aprobó su informe sobre la ética de la robótica. Además, la COMEST comenzó a analizar las consecuencias éticas del Internet de las cosas a finales de 2017.

6. Habida cuenta de sus profundas implicaciones sociales, muchas organizaciones y gobiernos están preocupados por las repercusiones éticas de la IA. Se están elaborando estrategias y marcos nacionales, regionales y de otro tipo sobre la IA. Cada vez hay más informes y directrices sobre la IA y la ética, como los del Consejo de Europa, la Comisión Europea, el Instituto de Ingeniería Eléctrica y Electrónica (IEEE), la Organización de Cooperación y Desarrollo Económicos (OCDE), la Unión Internacional de Telecomunicaciones (UIT) y la Organización Mundial de la Salud (OMS), que han definido principios importantes para el diseño, desarrollo y despliegue de la IA. La UNESCO ha seguido de cerca estos debates, concretamente en calidad de observadora en el Grupo de expertos de alto nivel de la Comisión Europea sobre inteligencia artificial, como miembro del Grupo de expertos de la OCDE sobre IA (AIGO), como asociada de la Cumbre mundial de IA para el bienestar de la UIT, y como participante en otros foros intergubernamentales.

7. La UNESCO tiene una perspectiva única para alimentar este debate, dada su gran ventaja comparativa derivada de su composición universal y sus competencias multidisciplinarias. En este sentido, la Organización puede proporcionar realmente una plataforma mundial para el diálogo sobre la ética de la IA, reuniendo a países desarrollados y en desarrollo, diferentes puntos de vista culturales y morales, así como diversas partes interesadas de los sectores público y privado. Por lo tanto, además de las numerosas directrices y marcos éticos que están elaborando actualmente los gobiernos, las empresas y las organizaciones de la sociedad civil, la UNESCO puede contribuir al desarrollo de la IA en beneficio de toda la humanidad, el desarrollo sostenible y la paz. Con este fin, la UNESCO actúa como puente entre los Estados Miembros, que han destacado en repetidas ocasiones su apoyo a la labor prevista por la UNESCO en materia de IA, y la sociedad civil, la comunidad técnica, el mundo académico y el sector privado, incluidas las industrias culturales y creativas, aprovechando su experiencia en cuanto a consultas entre múltiples interesados y la búsqueda de un consenso imparcial.

8. Basándose en la labor de la COMEST, en agosto de 2018 la Directora General pidió a esta Comisión que preparara un estudio preliminar sobre la ética de la IA, a fin de ayudar a la UNESCO a fundamentar su reflexión en este ámbito. Se creó el Grupo de trabajo ampliado sobre la ética de la inteligencia artificial de la COMEST, integrado por tres expertos externos, encargado de elaborar un estudio preliminar sobre la ética de la inteligencia artificial, cuyo texto completo figura en el anexo del presente documento.

9. El estudio del Grupo de Trabajo ampliado sobre la ética de la inteligencia artificial de la COMEST profundiza las cuestiones éticas relacionadas con el mandato de la UNESCO. Desde la perspectiva de la Organización, la IA cuestiona el papel de la educación en las sociedades en muchos aspectos. Las ideas actuales sobre el "aprendizaje a lo largo de toda la vida" podrían tener que ampliarse para convertirse en un modelo de educación continua, que incluya la creación de otros tipos de programas de educación y diplomas. La IA requiere que la educación fomente la adquisición de conocimientos básicos sobre IA, el pensamiento crítico, la adaptación en el mercado laboral y la educación ética de los ingenieros. En las ciencias naturales y sociales, así como en las ciencias de la vida y las ciencias ambientales, la IA cuestiona fundamentalmente nuestros conceptos de comprensión y explicación científicas. Esto también repercute en cómo aplicamos el conocimiento científico en los contextos sociales. La IA requiere que se le introduzca de forma responsable en la práctica científica y en la toma de decisiones basada en sistemas de IA, teniendo en cuenta una evaluación y control humanos, y evitando la exacerbación de las desigualdades estructurales. Aunque la IA es una poderosa herramienta para la creatividad humana, plantea importantes interrogantes sobre el futuro del arte, los derechos y la remuneración de los artistas, y la integridad de la cadena de valor creativa. La IA debe fomentar la diversidad cultural, la inclusividad y el florecimiento de la experiencia humana, procurando no ampliar la brecha digital. También deberá promoverse una estrategia plurilingüe. La inteligencia artificial desempeña un papel cada vez más importante en el procesamiento, estructuración y suministro de información, y resulta indispensable prestar atención a las nuevas brechas digitales entre países y dentro de distintos grupos sociales. La IA deberá fortalecer la libertad de expresión, el acceso universal a la información y la calidad del periodismo, así como medios de comunicación libres, independientes y pluralistas, evitando la difusión de información falsa. Será importante el fomento de una gobernanza de Internet en la que participen múltiples partes interesadas.

10. En el estudio preliminar también se definen las dimensiones éticas y mundiales de la paz, la diversidad cultural, la igualdad de género y la sostenibilidad, en relación con la labor de la UNESCO. A fin de contribuir a la paz, la IA podría utilizarse para obtener información sobre los factores que impulsan los conflictos, y nunca deberá salirse del control humano. La nueva economía digital que está surgiendo trae consigo importantes desafíos y oportunidades para las sociedades de África y otros países en desarrollo. Desde el punto de vista ético, la IA deberá integrarse en las políticas y estrategias nacionales de desarrollo, dando cabida a las culturas, valores y conocimientos endógenos a fin de desarrollar las economías africanas. Deberá evitarse el sesgo de género en la elaboración de algoritmos, en los conjuntos de datos utilizados para su creación, y en su uso para la toma de decisiones. La IA deberá desarrollarse de manera sostenible, teniendo en cuenta el ciclo completo de su producción y de las tecnologías de la información. La IA puede emplearse para la vigilancia ambiental y la gestión de riesgos, y para prevenir y atenuar las crisis ambientales.

III. CONVENIENCIA O NECESIDAD DE UN INSTRUMENTO NORMATIVO

11. Los debates actuales reflejan que hoy en día, en el plano mundial, sería necesario contar con una orientación ética universal global sobre los valores fundamentales que deben sustentar la elaboración de los sistemas de IA. Debido a su carácter transnacional, las soluciones duraderas solo pueden encontrarse en dicho plano. Habida cuenta de su mandato normativo, la UNESCO deberá sensibilizar a las distintas partes interesadas acerca de las repercusiones éticas de la IA en los distintos aspectos sociales, culturales y científicos de la sociedad, trabajando en la creación de un instrumento normativo sobre la ética de la IA. Un instrumento de ese tipo deberá constituir un mecanismo mundial para documentar los cambios socioculturales provocados por el desarrollo

rápido y no lineal de la IA y las cuestiones éticas conexas. También deberá servir como un medio para integrar los valores universales en los sistemas de IA, que deben ser compatibles con las normas y los derechos humanos internacionalmente acordados, y ajustarse a una visión centrada en el ser humano.

12. Teniendo en cuenta el aumento reciente del número de declaraciones de principios éticos sobre la IA en 2018 en los planos nacional y regional, sería oportuno que la UNESCO estudiara la posibilidad de preparar un instrumento normativo mundial sobre la ética de la IA, aplicando un enfoque mundial, pluralista, multidisciplinario, multicultural y multipartito, basado en todas las esferas de competencia de la Organización y la diversidad de sus redes.

13. Además, un nuevo instrumento normativo sobre la ética de la IA podría inspirarse de la importante función de la UNESCO en la ejecución de la Cumbre Mundial sobre la Sociedad de la Información (CMSI). Desde 1997, la UNESCO ha emprendido una serie de iniciativas para abordar las dimensiones éticas de la sociedad de la información, que son una de las Líneas de Acción del Plan de Acción de la CMSI, cuya aplicación está bajo la responsabilidad de la UNESCO. Un nuevo instrumento normativo podría también sacar partido de la larga experiencia de la UNESCO en crear vínculos entre una visión de las sociedades del conocimiento centrada en el ser humano y las tecnologías digitales. El Programa Información para Todos (PIPT) de la UNESCO, de carácter intergubernamental, decidió que el trabajo sobre la ética sería una de sus prioridades. Los principios R.O.A.M. de la UNESCO relativos a la universalidad de Internet (que promueven el advenimiento de un Internet basado en los derechos humanos, abierto, accesible y gobernado por múltiples interesados) fueron aprobados por la Conferencia General en su 38ª reunión, y se recogen en las directrices de la UNESCO para la promoción de la diversidad de las expresiones culturales en el entorno digital. Junto con los indicadores conexos, los principios R.O.A.M. representan un marco acordado que puede ayudar aún más a la UNESCO y a las partes interesadas a decidir sobre sus enfoques relativos a las cuestiones de las aplicaciones, la gobernanza y la ética de la inteligencia artificial.

IV. FORMA DEL INSTRUMENTO

14. La naturaleza del posible documento normativo (declaración, recomendación o convención) deberá, de conformidad con el Reglamento sobre las recomendaciones a los Estados Miembros y las convenciones internacionales previstas en el párrafo 4 del Artículo IV de la Constitución, escogerla la Conferencia General. Los distintos informes mencionados en el estudio preliminar del Grupo de trabajo ampliado sobre la ética de la inteligencia artificial de la COMEST coinciden en la conclusión de que no solo es deseable, sino asimismo urgente, que se adopten medidas para establecer un instrumento mundial no vinculante en forma de recomendación. Una recomendación, teniendo en cuenta su carácter no vinculante y el hecho de que se centra en los principios y normas para la regulación internacional de una cuestión concreta, sería un método más flexible y adaptado a la complejidad de las cuestiones éticas planteadas por la IA.

V. MÉTODOS DE TRABAJO Y CALENDARIO PROPUESTOS

15. De conformidad con el *Reglamento sobre las recomendaciones a los Estados Miembros y las convenciones internacionales previstas en el párrafo 4 del Artículo IV de la Constitución*, se invita al Consejo Ejecutivo a examinar el presente estudio preliminar y a decidir si debe o no ser incluido en el orden del día provisional de la 40ª reunión de la Conferencia General. En el caso de que el Consejo Ejecutivo se muestre favorable a su inclusión, la Directora General, según lo dispuesto en el Reglamento mencionado, hará llegar a los Estados Miembros un ejemplar del estudio preliminar, junto con las observaciones que haya formulado y las decisiones que haya adoptado el Consejo Ejecutivo al respecto, por lo menos 70 días antes de la apertura de la 40ª reunión de la Conferencia General, es decir, antes de finales de agosto de 2019.

16. Con arreglo al artículo 6 de dicho Reglamento, se invitará a la Conferencia General a que, tras haber examinado el presente estudio y las observaciones del Consejo Ejecutivo al respecto, decida acerca del camino a seguir.

VI. ASPECTOS FINANCIEROS

17. La Secretaría ha evaluado cuidadosamente las implicaciones y los costos que tendría la preparación de un nuevo instrumento normativo si el Consejo Ejecutivo decide incluir esta cuestión en el orden del día provisional de la 40ª reunión de la Conferencia General. Dadas las limitaciones financieras, una parte de esos gastos podría sufragarse mediante contribuciones financieras voluntarias y en especie.

ANEXO

United Nations Educational, Scientific and Cultural Organization

World Commission on the Ethics of Scientific Knowledge and Technology (COMEST)

SHS/COMEST/EXTWG-ETHICS-AI/2019/1
París, 26 de febrero de 2019
Original: inglés

ESTUDIO PRELIMINAR SOBRE LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

Sobre la base de la labor de la COMEST relativa a la ética de la robótica (2017) y las consecuencias éticas de la Internet de los objetos, un Grupo de trabajo ampliado sobre la ética de la inteligencia artificial de la COMEST ha elaborado el presente estudio preliminar.

Este documento no pretende ser exhaustivo, ni representa necesariamente las opiniones de los Estados Miembros de la UNESCO.

ÍNDICE

ESTUDIO PRELIMINAR SOBRE LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

INTRODUCCIÓN

1. El mundo se enfrenta a un rápido ascenso de la "inteligencia artificial" (IA). Los avances en este campo dan lugar a la introducción de máquinas con la capacidad para aprender y realizar tareas cognitivas que solían estar al alcance únicamente de los seres humanos. Es probable que este desarrollo tecnológico tenga consecuencias sociales y culturales significativas. Dado que la IA es una tecnología cognitiva, sus implicaciones atañen intrínsecamente a los campos de actuación esenciales de la UNESCO: la educación, la ciencia, la cultura y la comunicación. Los algoritmos han pasado a desempeñar un papel crucial en la selección de la información y las noticias que se leen, la música que se escucha y las decisiones que se adoptan. Los sistemas de IA proporcionan asesoramiento cada vez más a médicos, científicos y jueces. También en el campo de la investigación científica, la IA interviene en el análisis y la interpretación de datos. Además, la sustitución en curso de trabajadores por tecnologías inteligentes exige nuevas formas de resistencia y flexibilidad en la mano de obra humana. Pensadores públicos como Stephen Hawking han llegado a manifestar su temor de que la IA pueda representar una amenaza existencial para la humanidad, debido a su potencial para asumir el control de numerosos aspectos de nuestra vida diaria y de la organización de la sociedad.

2. En la década de 1950 se introdujo el término "inteligencia artificial" para aludir a las máquinas capaces de realizar otras tareas al margen de las meramente rutinarias. Con el incremento de la potencia de computación, el término pasó a aplicarse a las máquinas que cuentan con la capacidad de aprender. Aunque no existe una única definición de IA, se conviene comúnmente en que las máquinas que se basan en la IA, o en la "informática cognitiva", son potencialmente capaces de imitar, o incluso exceder, las capacidades cognitivas humanas, incluida las sensoriales, la interacción lingüística, el razonamiento y el análisis, la resolución de problemas, e incluso la creatividad. Además, estas "máquinas inteligentes" pueden demostrar capacidades de aprendizaje similares a las humanas con mecanismos de autorrelación y autocorrección, sobre la base de algoritmos que encarnan el "aprendizaje automático" o incluso el "aprendizaje profundo", utilizando "redes neuronales" que imitan el funcionamiento del cerebro humano.

3. Recientemente, grandes compañías tecnológicas multinacionales en muchas regiones del mundo han comenzado a efectuar grandes inversiones en la utilización de la IA en sus productos. La potencia informática ha alcanzado el nivel suficiente para ejecutar algoritmos muy complicados y trabajar con los denominados "big data" o macrodatos, los enormes conjuntos de datos que pueden utilizarse en el aprendizaje automático. Estas empresas disponen de un acceso a una capacidad de computación casi ilimitada, así como a los datos recabados de miles de millones de personas con el fin de "alimentar" los sistemas de IA como factores de entrada para el aprendizaje. Por otro lado, a través de sus productos, la IA ejerce cada vez mayor influencia en la vida diaria de las personas y en ámbitos profesionales como la sanidad, la educación, la investigación científica, las comunicaciones, el transporte, la seguridad y el arte.

4. Esta profunda influencia de la IA plantea motivos de inquietud que podrían afectar a la confianza de las personas en estas tecnologías. Tales preocupaciones van desde la posibilidad de comisión de delitos, fraudes y robos de identidad, hasta el acoso y el abuso sexual; de la incitación al odio y la discriminación, a la propagación de la desinformación; y en términos más generales, de la transparencia de los algoritmos a la posibilidad de confiar en los sistemas de IA. Puesto que muchos de estos problemas no pueden abordarse únicamente mediante la regulación, la UNESCO ha venido proponiendo la gobernanza de múltiples partes interesadas como una modalidad óptima para procurar la participación de los diversos agentes en la formulación y la ejecución de normas, principios éticos y políticas, así como la capacitación de los usuarios.

5. Debido a sus profundas consecuencias sociales, numerosas organizaciones y gobiernos muestran su preocupación respecto a las implicaciones éticas de la IA. La Comisión Europea ha constituido un Grupo de Expertos de Alto Nivel sobre IA compuesto por representantes del ámbito académico, la sociedad civil, las empresas, así como una Alianza Europea de la IA, que es un foro de debate amplio y abierto sobre todos los aspectos que atañen al desarrollo de la IA y sus efectos. El Grupo Europeo de Ética de la Ciencia y las Nuevas Tecnologías ha publicado una *Declaración sobre la IA, la Robótica y los Sistemas Autónomos* (EGE, 2018)*.* La Comisión Europea ha publicado una *Comunicación sobre Inteligencia artificial para Europa* (CE, 2018), y el Consejo de Europa ha elaborado varios informes sobre IA y ha creado un Comité de Expertos para trabajar en las *Dimensiones en cuanto a los derechos humanos del tratamiento automatizado de datos y las diferentes formas de inteligencia artificial*. La organización IEEE ha puesto en marcha una Iniciativa Mundial sobre la Ética de los Sistemas Inteligentes y Autónomos. La OCDE ha emprendido el proyecto denominado "Hacia lo digital", cuyo objetivo es ayudar a los responsables de la formulación de políticas en todas las esferas pertinentes a comprender mejor la revolución digital que tiene lugar en los diferentes sectores de la economía y la sociedad en su conjunto. La OCDE también ha creado un grupo de expertos (AIGO) para que formulen directrices respecto a la especificación de los principios relativos a la inteligencia artificial en la sociedad. La UIT y la OMS han establecido un Grupo de Debate sobre la "inteligencia artificial para la salud". Además, muchos países han iniciado actividades de reflexión sobre su orientación ética y política respecto a la IA, como el informe Villani en Francia (Villani y cols., 2018); el informe de la Cámara de los Lores en el Reino Unido (Cámara de los Lores, 2017); y el informe de la Oficina Ejecutiva del Presidente de los Estados Unidos (2016).

6. La UNESCO cuenta con una perspectiva singular que añadir a este debate. La IA tiene consecuencias para los campos de actuación esenciales de la UNESCO. Por tanto, además de las numerosas directrices y marcos éticos que elaboran actualmente los gobiernos, las empresas y las organizaciones sociales, la UNESCO puede aportar un enfoque multidisciplinario, universal y global a la evolución de la IA al servicio de la humanidad, el desarrollo sostenible y la paz.

7. En este sentido, existen varios marcos e iniciativas en los que basarse. En primer lugar, nos encontramos con el marco de los *derechos humanos*, que fundamentó la Declaración de Principios de Ginebra de la Cumbre Mundial sobre la Sociedad de la Información (CMSI) de 2003, en el que se afirma que "el uso de las TIC y la creación de contenidos debería respetar los derechos humanos y las libertades fundamentales de otros, lo que incluye la privacidad personal y el derecho a la libertad de opinión, conciencia y religión de conformidad con los instrumentos internacionales relevantes." (CMSI, 2003). La CMSI (2005) propone un enfoque de múltiples partes interesadas que exige la cooperación efectiva de todas ellas, incluidos los gobiernos, el sector privado, la sociedad civil, las organizaciones internacionales y las comunidades técnica y académica. En el proceso de seguimiento de la CMSI, la UNESCO ha adoptado este enfoque de múltiples partes interesadas y ha asumido la responsabilidad de la ejecución de las líneas de acción relativas al acceso (C3), el aprendizaje electrónico (C7), la diversidad cultural (C8), los medios de comunicación (C9) y la dimensión ética de la sociedad de la información (C10).

8. En segundo lugar figura el marco de la *Universalidad de Internet* y los *Principios ROAM* conexos aprobados por la Conferencia General, en su 38ª reunión, en 2015 (UNESCO, 2015b). Estos principios tratan de los derechos humanos, la apertura, la accesibilidad y la participación de múltiples partes interesadas, y son el fruto del estudio de la UNESCO denominado "Keystones" (Piedras angulares) para la 38ª reunión de la Conferencia General (UNESCO, 2015a). En el documento final "Connecting the Dots" (Conectando los puntos) de esta conferencia, la UNESCO se compromete a promover la reflexión ética basada en los derechos humanos, la investigación y el diálogo público sobre las implicaciones de las tecnologías nuevas y emergentes y sus posibles impactos sociales. Además, en la 18ª reunión del Consejo Intergubernamental del Programa Información para Todos (PIPT) de la UNESCO se examinó y aprobó el Código de ética para la sociedad de la información, elaborado por el Grupo de Trabajo del PIPT sobre ética de la información.

9. Con el fin de investigar las implicaciones éticas de la IA, en este estudio se explicará en primer lugar qué es la Inteligencia Artificial, cómo funciona y en qué se diferencia de otras

tecnologías. En el segundo apartado se abordarán los aspectos éticos de la IA, tomando como punto de partida los campos de actuación de la educación, la ciencia, la cultura y la comunicación de la UNESCO, así como las dimensiones éticas globales de la paz, la diversidad cultural, la igualdad de género y la sostenibilidad. Esta investigación debe percibirse como una exploración, más que como un análisis exhaustivo, en la que se aborda desde la diversidad cultural hasta la confianza en la ciencia, de la creatividad artística, al pensamiento crítico, y de la toma de decisiones basada en la IA, al papel de esta en los países en desarrollo. En el tercer apartado del presente estudio preliminar se esbozarán las dimensiones esenciales que una reflexión ética adecuada sobre la IA debe tener desde la perspectiva de la UNESCO.

I. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?

I.1. Definición

10. La idea de "inteligencia artificial" (IA) (en referencia a los seres, máquinas o herramientas "creados artificialmente" e "inteligentes") se ha planteado a lo largo de la historia humana. Sus diversas formas pueden encontrarse en las distintas religiones, mitologías, literaturas y tradiciones filosóficas, tanto occidentales, como no occidentales. En este sentido, todas ellas atestiguan la perenne curiosidad de la humanidad respecto a tales entidades, y a pesar de la expresión de esta curiosidad a través de apariencias culturalmente diversas, se trata de un fenómeno compartido o intercultural. En la actualidad, la fascinación por la IA –incluidas sus dimensiones éticas– se ve amplificada por su desarrollo y sus aplicaciones en el mundo real.

11. Todo examen de las implicaciones éticas de la IA requiere una aclaración de sus posibles significados. El término lo acuñaron en 1955 John McCarthy, Marvin L. Minsky, Nathaniel Rochester y Claude E. Shannon. El "estudio de la inteligencia artificial" se planificó para "proceder sobre la base de la conjetura de que todos los aspectos del aprendizaje o cualquier otro rasgo de la inteligencia pueden, en principio, ser descritos de una forma tan precisa que se puede crear una máquina que los simule" (McCarthy y cols., 2006 [1955], p. 12). A medida que el campo se desarrolló y diversificó en las décadas posteriores, aumentó el número de significados de la "IA", y no existe una definición consensuada universalmente en la actualidad. Varias definiciones de IA están relacionadas con diversos enfoques disciplinarios como la informática, la ingeniería eléctrica, la robótica, la psicología o la filosofía.

12. A pesar de la multitud y diversidad de definiciones de IA, existe cierto consenso, al nivel más general, respecto a que sus dos aspectos pueden distinguirse: uno etiquetado habitualmente como "teórico" o "científico", y el otro como "pragmático" o "tecnológico".

13. Hablar de IA "teórica" o "científica" es aludir la "uso de conceptos y modelos de IA para ayudar a responder preguntas sobre los seres humanos y otras cosas vivas" (Boden, 2016, p.2). Así, la IA "teórica" o "científica" se interconecta de manera natural con disciplinas como la filosofía, la lógica, la lingüística, la psicología y la ciencia cognitiva. Se ocupa de cuestiones como: ¿qué se entiende por "inteligencia" y cómo distinguir la inteligencia "natural" de la "artificial"? ¿Es necesario el lenguaje simbólico para los procesos de pensamiento? ¿Es posible crear una "IA fuerte" (una inteligencia genuina del mismo tipo y nivel de generalidad que la inteligencia humana), frente a una "IA débil" (aquélla que únicamente *imita* la inteligencia humana y es capaz de realizar un número limitado de tareas definidas de forma restringida)? Aunque cuestiones como estas son teóricas o científicas, atañen a diversos aspectos de índole metafísica o espiritual (p. ej., sobre la singularidad humana o el libre albedrío) que poseen por sí mismos implicaciones éticas indirectas, pero graves en cualquier caso.

14. La IA "pragmática" o "tecnológica" se orienta a la ingeniería. Se basa en diversas ramas de la IA (son ejemplos propios de libros de texto el procesamiento del lenguaje natural, la representación del conocimiento, el razonamiento automatizado, el aprendizaje automático, el aprendizaje profundo, la visión informática y la robótica (Russell y Norvig, 2016, p. 2-3)) con el fin de crear máquinas o programas capaces de realizar tareas de manera independiente que de otro modo requerirían de la inteligencia y la intervención humanas. La IA "pragmática" o "tecnológica" cosechó un notable éxito al combinarse con las TIC (tecnología de la información y las comunicaciones). Las innovaciones en IA se utilizan actualmente en numerosas áreas de la vida moderna, como el transporte, la medicina, la comunicación, la educación, la ciencia, las finanzas, el derecho, las fuerzas armadas, el marketing, los servicios al cliente o el entretenimiento. Tales innovaciones suscitan preocupaciones éticas directas, que van desde la desaparición de los empleos tradicionales, pasando por la responsabilidad por posibles daños físicos o psicológicos a los seres humanos, hasta la deshumanización general de las relaciones humanas y la sociedad en general. Por el momento, ningún sistema de IA puede considerarse como un instrumento inteligente de uso general capaz de funcionar adecuadamente en una amplia gama de entornos, lo que constituye una capacidad propia de la inteligencia humana.

15. Una de las particularidades de la IA atañe a su "falta de familiaridad" respecto a nosotros, los humanos, en el sentido de que el modo en que funciona su inteligencia nos parece extraño y misterioso. La esencia de esta "falta de familiaridad" es lo que podría denominarse "rendimiento sin sensibilización". Una IA de alta funcionalidad, como AlphaGo o Watson, puede actuar de manera impresionante sin reconocer lo que está haciendo. AlphaGo derrotó a varios maestros de "go" sin ni siquiera saber que estaba practicando un juego humano denominado go. Watson respondió con enorme rapidez preguntas complicadísimas que a la mayoría de los seres humanos les resultan difíciles incluso de comprender en un tiempo dado. Sin embargo, Watson no "responde" en el sentido humano, sino que "computa" las probabilidades de varias respuestas candidatas con arreglo a su análisis automatizado de una base de datos disponible. AlphaGo y Watson presentan un desempeño brillante sin ser conscientes de lo que hacen.

16. No cabe duda de que se plantean cuestiones filosóficas importantes sobre si el "juego" de AlphaGo y la "respuesta" de Watson son "genuinas" o no. Sin embargo, un hecho más significativo desde el punto de vista ético es que los seres humanos no estamos acostumbrados a este tipo de inteligencia. Siempre que nos encontramos con obras de arte, literatura o ciencia impresionantes, consideramos de manera natural la inteligencia "consciente" que subyace a las mismas. Reconocemos el carácter único de Beethoven detrás de su 9ª sinfonía, y la abrumadora mentalidad investigadora que se esconde tras el teorema de incompletitud de Goedel. El mero hecho de que no debamos aplicar esta regla general que nos resulta tan familiar con respecto a los desempeños actuaciones brillantes cuando interactuamos con la IA de alto rendimiento plantea serios desafíos sociales y éticos. Puesto que estamos acostumbrados a interactuar emocional y socialmente con agentes conductualmente inteligentes, interactuamos de forma natural del mismo modo con la "IA de alto rendimiento sin conciencia", como en el caso de los denominados "robots emocionales o sociales", por ejemplo, los "asistentes domésticos inteligentes" (Alexa, Siri, o el asistente de Google). En el nivel actual de desarrollo tecnológico, una IA de alto rendimiento y sin conciencia no puede responder adecuadamente a las expectativas emocionales y sociales complejas de los interlocutores humanos, mientras que su comportamiento externo, combinado con la imaginación humana, puede generar una esperanza "poco realista" de que se establezcan interacciones genuinas con los seres humanos. Es importante que recordemos que la mente aparentemente "emocional" de la IA es mucho más un producto de nuestra imaginación que de la realidad. Existe un consenso general respecto a que los sistemas artificialmente inteligentes carecen de conciencia en el sentido humano experiencial, aunque puedan responder preguntas acerca del contexto de sus acciones. Es importante no hacer equivaler la experiencia con la inteligencia, aun cuando algunos expertos hayan sugerido que los recientes avances en la IA podrían constituir asimismo un motivo para reexaminar la importancia de esta experiencia o conciencia para ser humano. Si la experiencia constituye un elemento esencial del ser humano, las consideraciones éticas deben garantizar que esta se proteja y potencie mediante el uso de la IA en lugar de que se margine o se deshabilite. No obstante, puede suceder que nuestra experiencia con la IA de alto rendimiento y sin conciencia

pueda influir en cualquier caso en nuestras interacciones con los seres humanos comunes y corrientes con conciencia.

I.2. ¿Cómo funciona la IA?

17. Para poder realizar las tareas de la mente humana, una máquina de IA ha de ser capaz de percibir el entorno y recabar datos de manera dinámica, con el fin de procesarlos con prontitud y de responder, basándose en su "experiencia" pasada, en sus principios preestablecidos para la toma de decisiones y en su anticipación respecto al futuro. Sin embargo, la tecnología que subyace a la IA es una TIC clásica: se basa en la recopilación/adquisición de datos, y en su almacenamiento, tratamiento y comunicación. Los rasgos singulares de las máquinas cognitivas se derivan de las cantidades, que se transforman en cualidades. La tecnología de la IA se basa en los siguientes componentes:

    a) *Datos dinámicos*. El sistema debe exponerse a entornos cambiantes y a todos los datos relevantes adquiridos a través de diversos sensores, para clasificarlos y almacenarlos, y para poder procesarlos con rapidez.

    b) *Procesamiento rápido*. Las máquinas cognitivas deben reaccionar rápidamente. Por lo tanto, la IA ha de disponer de recursos informáticos y de comunicación fiables, veloces y sólidos.

    c) *Principios de adopción de decisiones*. La adopción de decisiones de la IA se basa en algoritmos de aprendizaje automático. Por consiguiente, su respuesta a una tarea específica depende de su "experiencia", es decir, de los datos a los que se ha expuesto. Los algoritmos que subyacen a las decisiones adoptadas por las máquinas cognitivas se basan en ciertos principios generales a los que el algoritmo se atiene y que intenta optimizar, teniendo en cuenta los datos que se le proporcionan.

La capacidad actual de integrar eficientemente algoritmos de adquisición de datos dinámicos y aprendizaje automático para una rápida adopción de decisiones permite la creación de "máquinas cognitivas".

I.3. ¿En qué se diferencia la IA de otras tecnologías?

18. La mayoría de las tecnologías del siglo XX se atienen a modelos. Es decir, los científicos estudian la naturaleza y proponen un modelo científico para describirla, y la tecnología avanza sobre la base de tales modelos. Por ejemplo, entender la propagación de las ondas electromagnéticas constituye la base de la tecnología de la telecomunicación inalámbrica. Sin embargo, la modelización del cerebro humano es una tarea que, al parecer, dista mucho de encontrarse en una etapa en la que una máquina cognitiva puede basarse en un modelo. Por lo tanto, la IA se desarrolla con arreglo a un planteamiento diferente: un enfoque basado en los datos.

19. Tal enfoque constituye el núcleo del *aprendizaje automático*, que se basa comúnmente en "redes neuronales artificiales" (RNA). Las RNA las forman diversos nodos similares conceptualmente a las neuronas cerebrales interconectadas a través de una serie de capas. Los nodos de la capa de entrada reciben información del entorno y, en cada nodo, se aplica una transformación no lineal. Estos sistemas "aprenden" a realizar tareas mediante la consideración de ejemplos (datos etiquetados), generalmente sin estar programados con reglas o modelos específicos de cada tarea. El aprendizaje profundo, para concluir, se basa en las RNA de varias capas, lo que habilita a la máquina para reconocer conceptos complejos como las caras y los cuerpos humanos, la comprensión del lenguaje y la clasificación de todo tipo de imágenes.

20. La cuestión clave en cuanto a la capacidad de la IA para mostrar capacidades similares a las humanas es su escalabilidad. El rendimiento de las máquinas de IA depende de los datos a los que se exponen, y para lograr el mejor rendimiento, el acceso a los datos pertinentes debe ser ilimitado.

Puede que existan limitaciones técnicas en el acceso a los datos, pero el modo en que estos se seleccionan y clasifican constituye asimismo una cuestión sociocultural (Crawford, 2017). La clasificación es específica de cada cultura y un producto de la historia, y puede generar un sesgo en las decisiones adoptadas por el algoritmo. Si la misma máquina está expuesta a diversos conjuntos de datos, su sesgo puede reducirse, pero no suprimirse por completo (Oficina Ejecutiva del Presidente, 2016). Es importante señalar que, con el fin de cumplir lo dispuesto en el artículo 27 de la Declaración Universal de los Derechos Humanos (en la que se afirma que cada ser humano tiene derecho a los beneficios del progreso científico) y garantizar la diversidad de los conjuntos de datos a disposición de la IA, es pertinente promover el refuerzo de las capacidades de los Estados, tanto en lo que se refiere a las capacidades humanas como a las infraestructuras.

21. La tecnología de la IA ha madurado bajo el impulso de empresas multinacionales que no se encuentran limitadas por las restricciones locales y nacionales. Además, para garantizar el procesamiento rápido y la fiabilidad de los sistemas, la ubicación real de los procesos informáticos se distribuye, y la localización de una máquina de IA no queda definida por el lugar en el que opera. En la práctica, la IA se basa en la tecnología de computación en la nube, en la que la ubicación de las unidades de almacenamiento y procesamiento puede ser cualquier lugar. La tecnología de la IA se caracteriza por lo siguiente:

a) aunque muchas de sus aplicaciones pertenecen al ámbito público, la tecnología de la IA la desarrollan y lideran empresas multinacionales, que operan en su mayoría en el sector privado y se encuentran menos obligadas respecto al bien público.

b) La IA no se limita a un emplazamiento tangible. Esto plantea un reto en cuanto a cómo regular la tecnología de la IA a escala nacional e internacional.

c) La tecnología se basa en la accesibilidad a los datos personales y públicos.

d) Las tecnologías de la IA no son neutrales, sino sesgadas de manera inherente debido a los datos sobre cuya base se capacitan, y a las opciones adoptadas durante la capacitación respecto a los datos.

e) Las decisiones de la inteligencia artificial y las máquinas cognitivas no pueden ser plenamente predecibles o explicables. En lugar de funcionar de manera mecánica o determinista, el software de la IA aprende de los datos dinámicos a medida que se desarrolla e incorpora la experiencia del mundo real en su toma de decisiones.

II. CONSIDERACIONES ÉTICAS

22. La inteligencia artificial tiene importantes implicaciones sociales y culturales. Al igual que otras muchas tecnologías de la información, la IA plantea cuestiones relacionadas con la libertad de expresión, la privacidad y la vigilancia, la propiedad de los datos, el sesgo y la discriminación, manipulación de la información y la confianza, las relaciones de poder, y el impacto medioambiental en lo que se refiere a su consumo de energía. Además, la IA plantea específicamente nuevos retos relacionados con su interacción con las capacidades cognitivas humanas. Los sistemas basados en la IA tienen consecuencias para la comprensión y la experiencia humanas. Los algoritmos de las redes sociales y los sitios de noticias pueden facilitar la propagación de desinformación y repercutir en el significado percibido de los "hechos" y la "verdad", así como en la interacción y la participación en el ámbito político. El aprendizaje automático puede integrar y exacerbar el sesgo, lo que puede dar lugar a la desigualdad, a la exclusión, y a una amenaza a la diversidad cultural. La escala y el poder generados por la tecnología de la IA acentúa la asimetría entre individuos, grupos y países incluida la denominada "brecha digital" en cada nación, y entre naciones. Tal brecha puede verse agravada por la falta de acceso a elementos fundamentales como los algoritmos para el aprendizaje y la clasificación, los datos para capacitar y evaluar los algoritmos, los recursos humanos para codificar y configurar el software y preparar los datos, así como los recursos computacionales para el almacenamiento y el procesamiento de los datos.

23. En este sentido, la inteligencia artificial exige un análisis minucioso. Desde la perspectiva de la UNESCO, las cuestiones éticas más importantes en lo que concierne a la inteligencia artificial atañen a sus implicaciones para la cultura y la diversidad cultural, la educación, el conocimiento científico, y la comunicación y la información. Además, dada la orientación mundial de la UNESCO, los temas éticos globales de la paz, la sostenibilidad, la igualdad de género, y los retos específicos para África también merecen una atención específica.

II.1. Educación

24. La inteligencia artificial pone en cuestión el papel de la educación en las sociedades en muchos aspectos. En primer lugar, la IA exige un replanteamiento del papel social de la educación. El desplazamiento laboral causado por algunas formas de IA requiere, entre otras medidas, la reconversión profesional de los empleados y un nuevo enfoque para formular las cualificaciones finales de los programas educativos. Además, en un mundo de IA, la educación debe habilitar a los ciudadanos para que desarrollen nuevas formas de pensamiento crítico, incluida la "concienciación respecto a los algoritmos", y la capacidad de reflexionar sobre el impacto de la IA en la información, el conocimiento y la toma de decisiones. Un segundo ámbito de las cuestiones éticas relativas a la IA y la educación se refiere al papel de la primera en el proceso educativo en sí, como elemento de los entornos de aprendizaje digital, la robótica educativa y los sistemas de "análisis del aprendizaje", todos los cuales requieren un desarrollo y una implementación responsables. Por último, los ingenieros y desarrolladores de software deben recibir la formación adecuada para garantizar el diseño y la implantación responsables de la IA.

II.1.1. El papel social de la educación

25. Una de las principales preocupaciones sociales respecto a la IA consiste en el desplazamiento laboral. El ritmo de las transformaciones que propicia la IA plantea retos sin precedentes (Illanes y cols., 2018). Conllevará, en un futuro próximo, la necesidad de reconvertir laboralmente a un gran número de trabajadores y tendrá consecuencias radicales para las trayectorias profesionales que los estudiantes tendrán que seguir. Según una encuesta entre expertos de McKinsey de 2017, "los ejecutivos ven cada vez más la inversión en la reconversión profesional y la "mejora de las aptitudes" de los trabajadores existentes como una prioridad empresarial urgente" (Illanes y cols., 2018).

26. La IA, por tanto, instará a las sociedades a reconsiderar la educación y sus funciones sociales. Puede que la educación tradicional reglada impartida por las universidades deje de ser suficiente ante el auge de las economías digitalizadas y las aplicaciones de IA. Hasta la fecha, el modelo educativo estándar ha consistido habitualmente en proporcionar "conocimientos básicos" (Oppenheimer, 2018), y se ha centrado en competencias formales como la lectura, la escritura y las matemáticas. En el siglo XXI, la información y el conocimiento son omnipresentes, y exigen no solo una "capacitación en materia de datos" que permita a los alumnos leer, analizar y gestionar eficazmente esta información, sino también una "capacitación para la IA" que permita la reflexión crítica sobre el modo en que los sistemas informáticos inteligentes han participado en el reconocimiento de las necesidades de información, y en la selección, la interpretación, el almacenamiento y la representación de los datos.

27. Por otra parte, en un mercado de trabajo en continuo desarrollo, el sistema educativo ya no puede aspirar a instruir a las personas para una determinada profesión. La educación debe habilitar a las personas para que sean versátiles y resilientes, preparadas para un mundo en el que las tecnologías creen un mercado laboral dinámico y en el que los empleados deben volver a formarse periódicamente. Es posible que las ideas actuales sobre el "aprendizaje permanente" deban reorientarse al alza hacia un modelo de educación continua, incluido el desarrollo de otros tipos de titulaciones y diplomas.

II.1.2. La IA en la enseñanza y el aprendizaje

28. Los recursos educativos abiertos (REA) han constituido una incorporación importante al panorama del aprendizaje, con la libre disponibilidad de clases de alta calidad y otros recursos docentes a través de Internet. El potencial de los REA para influir en la educación de la población de todo el mundo es incomparable, pero aún no se ha materializado plenamente, como demuestran los limitados índices de finalización de cursos en línea masivos y abiertos (CLMA). La gran variedad y profundidad de los recursos disponibles han dado lugar a dos problemas. En primer lugar, el problema de encontrar el recurso adecuado para un determinado alumno o profesor que deseen reutilizar un recurso en sus propios materiales docentes. Esto ha dado lugar al segundo problema de reducción de la diversidad, debido a que algunos recursos adquieren una gran popularidad a expensas de otros contenidos potencialmente más relevantes, pero menos accesibles.

29. Un ejemplo a este respecto es el proyecto "X5GON" de Horizon 2020 (destinado a crear una red mundial de REA intermodal, intercultural, interlingüística, para múltiples dominios y sitios). Este proyecto, financiado por la Unión Europea, desarrolla métodos de inteligencia artificial que permitan tanto a los alumnos como a los profesores identificar los recursos que se ajustan a sus objetivos de aprendizaje, teniendo en cuenta los aspectos específicos de su situación. Por ejemplo, un profesor en África podría ser dirigido a conferencias en las que se presente un tema basado en el conocimiento local e indígena que resulte apropiado para el contexto cultural y local específico, pero, del mismo modo, el sistema también permitiría a un alumno de cualquier otra procedencia interesado en entender determinados desafíos africanos encontrar contenido relevante de este continente posiblemente traducido de una lengua local.

30. De este modo, la IA podría abordar los dos problemas identificados anteriormente. El primero se afronta ayudando a identificar los recursos que se adaptan mejor a las necesidades del alumno o del profesor mediante la modelización de sus intereses y objetivos, aprovechando al mismo tiempo una representación extendida de los enormes repositorios de REA disponibles en todo el mundo. Al ajustar las recomendaciones al alumno o el profesor concreto, se aborda asimismo el segundo problema, ya que las recomendaciones ya no dirigirán por defecto al recurso más popular sobre un tema determinado. Existe la opción adicional de vincular a los alumnos de diferentes culturas con el fin de potenciar el intercambio transcultural de ideas y, de este modo, promover el entendimiento y el respeto mutuos.

II.1.3. La formación de los ingenieros especializados en IA

31. El desarrollo de las tecnologías futuras está en manos de expertos técnicos. Tradicionalmente, los ingenieros han recibido formación para desarrollar productos que optimicen el rendimiento utilizando los mínimos recursos (potencia, espectro, espacio, peso, etc.), con unas restricciones externas dadas. En las últimas décadas, la ética de la tecnología ha desarrollado diversos métodos para incorporar la reflexión ética, la responsabilidad y el razonamiento al proceso de diseño. En el contexto de la IA, se ha acuñado el término "diseño éticamente alineado" (DEA) para aludir a los procesos de diseño en los que se consideran explícitamente valores humanos (IEEE, 2018).

32. Es fundamental aplicar el diseño éticamente alineado en la IA y otros sistemas autónomos e inteligentes (SAI), ya que tal aplicación hace posible abordar cuestiones éticas en una fase en la que la tecnología todavía puede adaptarse. Un buen ejemplo es el de la "privacidad desde el diseño". La privacidad podrá infringirse menos si no se almacenan todos los datos, sino únicamente los que requiere cada tarea. Un ejemplo de esta práctica es el del recuento de multitudes, es decir, contar el número de personas en una multitud basándose en las fotos disponibles. En este caso, si la foto se pretrata para extraer únicamente los contornos (bordes) de las figuras, las personas seguirán siendo irreconocibles y el algoritmo de conteo funcionará bien sin violar la privacidad. Del mismo modo, los encargados de la concepción de la IA pueden considerar otras cuestiones éticas como la prevención del sesgo algorítmico y la trazabilidad, reduciendo al mínimo la capacidad para hacer un mal uso de la tecnología, y la "explicabilidad" de las decisiones algorítmicas.

33. En la actualidad, la formación en ingeniería a escala global se centra fundamentalmente en cursos científicos y tecnológicos que no están intrínsecamente relacionados con los análisis de los valores humanos diseñados abiertamente para reforzar positivamente el bienestar humano y medioambiental. Resulta esencial cambiar esta situación e instruir a los futuros ingenieros y científicos informáticos para que adopten el diseño éticamente alineado de los sistemas de IA. Esto requiere una concienciación explícita de las posibles implicaciones sociales y éticas, y de las consecuencias de la tecnología en el diseño, así como de su posible uso indebido. La IEEE (una organización de ámbito mundial de más de 400.000 ingenieros eléctricos) promueve ya esta cuestión a través de su iniciativa global sobre la ética de los sistemas autónomos e inteligentes (https://ethicsinaction.ieee.org/). Abordar este asunto comprende asimismo la tarea de garantizar un esfuerzo activo encaminado a la inclusión de género, así como al fomento de la diversidad social y cultural de los ingenieros, y a una consideración global de las implicaciones sociales y éticas del diseño de los sistemas de IA. Deben promoverse las ocasiones para el diálogo entre los ingenieros y el público en general, con el fin de facilitar la comunicación sobre las necesidades y las visiones de la sociedad, y sobre el modo en que los ingenieros trabajan en la práctica y conducen sus investigaciones en sus actividades cotidianas.

II.2. La IA y el conocimiento científico

34. En el ámbito de la práctica científica, es probable que la IA tenga profundas implicaciones. En las ciencias naturales y sociales, así como en las ciencias de la vida y medioambientales, la IA pone en cuestión nuestros conceptos de comprensión y explicación científicas de una manera fundamental. Esto también tiene consecuencias respecto al modo en que aplicamos el conocimiento científico en diversos contextos sociales.

II.2.1. La IA y la explicación científica

35. Debido a las formas cada vez más potentes de aprendizaje automático y profundo, la IA pone en cuestión las concepciones existentes de una explicación científica satisfactoria, así como lo que podemos esperar de manera natural de teorías científicas predeciblemente exitosas. En la visión convencional de la ciencia, conforme al denominado modelo deductivo-nomológico, una explicación científica adecuada es capaz de dar lugar a predicciones correctas de fenómenos específicos basados en leyes, teorías y observaciones científicas. Por ejemplo, podemos afirmar legítimamente que explicamos cómo se mueve la luna alrededor de la Tierra en términos de la mecánica newtoniana solamente cuando podemos emplear dicha mecánica de una manera deductiva para predecir la órbita lunar. Tales predicciones suelen basarse en la comprensión causal, o en una comprensión unificadora de fenómenos aparentemente dispares.

36. Por el contrario, la IA puede generar predicciones impresionantemente precisas basadas en conjuntos de datos sin proporcionarnos ninguna explicación causal o unificadora de sus predicciones. Sus algoritmos no funcionan con los mismos conceptos semánticos que emplean los seres humanos para alcanzar la comprensión científica de un fenómeno. Esta brecha entre las predicciones de éxito, por un lado, y un entendimiento científico satisfactorio, por el otro, desempeñará probablemente un papel clave en la práctica científica, así como en la toma de decisiones basada en la IA.

37. Este hecho podría repercutir en la confianza en la ciencia, que se basa habitualmente en el método científico que explica los diferentes fenómenos de una manera sistemática y transparente, realizando predicciones racionales y basadas en datos contrastados. El éxito aparente de los algoritmos de aprendizaje automático para ofrecer resultados comparables sin tal modelo científicamente justificado podría tener implicaciones para la percepción y la evaluación públicas de la ciencia y la investigación científica.

38. Además, las investigaciones ponen de relieve que la calidad del aprendizaje automático depende en gran medida de los datos disponibles utilizados para capacitar a los algoritmos. En cualquier caso, dado que la mayoría de las aplicaciones de IA las desarrollan empresas privadas,

no siempre existe suficiente transparencia respecto a tales datos, en contraste con el método científico tradicional que garantiza la validez de los resultados al exigir la replicabilidad, es decir, la posibilidad de reproducirlos mediante la repetición de los mismos experimentos.

II.2.2. La IA, las ciencias de la vida y la salud

39. En el ámbito de las ciencias de la vida y la medicina en particular, el desarrollo de tecnologías de IA ha transformado significativamente el panorama de la atención sanitaria y la bioética a lo largo de los años. Tales tecnologías pueden producir efectos positivos, como una mayor precisión en la cirugía robótica, y un mejor cuidado de los niños autistas, pero, al mismo tiempo, generan preocupaciones éticas, como el coste que suponen en el contexto de la escasez de recursos en el sistema sanitario, y la transparencia que deberían aportar para respetar la autonomía de los pacientes.

40. Desde una perspectiva individual, la IA proporciona una nueva forma de tratar los asuntos médicos y de salud a la población lega. El uso de sitios de Internet y la multiplicación de aplicaciones de software para teléfonos móviles para el autodiagnóstico han brindado a las personas la oportunidad de generar diagnósticos de salud sin la participación de un profesional sanitario. Esta tendencia podría repercutir en la autoridad médica y la aceptación de la automedicación, incluidos los peligros que conlleva. También modifica la relación entre médicos y pacientes, y exige algún tipo de regulación sin obstaculizar la innovación y la autonomía.

41. Las tecnologías de la IA pueden propiciar que los proveedores de servicios sanitarios dispongan de más tiempo para dedicárselo a sus pacientes, por ejemplo, al facilitar la entrada de datos y el trabajo administrativo, pero, al mismo tiempo podrían sustituir los elementos holísticos y humanos de la asistencia. La conocida tecnología Watson for Oncology de IBM constituye un avance en el tratamiento del cáncer, pero también plantea cuestiones importantes respecto al carácter y las expectativas de la competencia técnica y la formación médicas, y las responsabilidades de los médicos que trabajan con el sistema. Se plantean preocupaciones similares con el desarrollo de "chatbots" dirigidos a personas que buscan ayuda y asesoramiento psicológicos, de aplicaciones para la detección temprana de episodios de enfermedades psiquiátricas, o de sistemas de IA para la elaboración de diagnósticos psiquiátricos sobre la base de la información recabada de la actividad de las personas en las redes sociales e Internet, lo que obviamente también tiene implicaciones importantes en lo que se refiere a la privacidad. Además, en el caso de las personas de edad avanzada, se introducen tecnologías basadas en la IA, como los robots sociales asistenciales, que pueden resultar útiles por motivos médicos para los pacientes con demencia, por ejemplo, pero que también plantean dudas respecto a la reducción de la asistencia prestada por humanos y el consiguiente aislamiento social.

42. La IA introduce asimismo una nueva dimensión en el debate en curso sobre el "perfeccionamiento humano" frente a la "terapia". Existen iniciativas encaminadas a integrar la IA en el cerebro humano utilizando una "interfaz neuronal": una malla que crece con el cerebro, que serviría como interfaz fluida entre este y el ordenador, y circularía por las venas y arterias del huésped (Hinchliffe, 2018). Este desarrollo tecnológico tiene implicaciones importantes para la cuestión de lo que significa ser humano, y de lo que es el funcionamiento humano "normal".

II.2.3. La IA y las ciencias medioambientales

43. La IA puede resultar beneficiosa para la ciencia medioambiental a través de diversas aplicaciones. Puede utilizarse para procesar e interpretar datos en el ámbito de la ecología, la biología de sistemas, la bioinformática, la investigación espacial y del clima, mejorando así la comprensión científica de procesos y mecanismos. La mejora del reciclaje, y de la vigilancia y la rehabilitación del medio ambiente, así como un consumo de energía más eficiente, pueden generar beneficios medioambientales directos. La IA en la agricultura y la ganadería puede dar lugar a una mejora de la producción de los cultivos (p. ej., mediante la fertilización y el riego automatizados) y del bienestar de los animales, y a atenuar los riesgos de enfermedades, plagas o amenazas

meteorológicas. Por otra parte, la IA puede propiciar cambios en la percepción humana de la naturaleza, ya sea de manera positiva mediante el fomento de la conciencia humana de la belleza o la independencia, o de manera negativa mediante una mayor "instrumentalización" de la naturaleza o la separación entre los seres humanos y los animales y el medio ambiente.

44. En todas las aplicaciones, los beneficios potenciales deben equilibrarse con el impacto medioambiental del ciclo de producción completo de la IA y las tecnologías de la información (TI). Se incluye a este respecto la minería de tierras raras y otras materias primas, la energía necesaria para producir y suministrar energía a las máquinas, y los residuos generados durante la producción y al final de los ciclos de vida útil. Es probable que el refuerzo de la IA contribuya a la creciente preocupación respecto al aumento del volumen de residuos electrónicos y la presión sobre las tierras raras generados por la industria informática. Además de los impactos sobre el medio ambiente y la salud, los residuos electrónicos tienen importantes implicaciones sociopolíticas, sobre todo en lo que atañe a la exportación a los países en desarrollo y a las poblaciones vulnerables (Heacock y cols., 2015).

45. La gestión del riesgo de catástrofes es un área en la que la IA puede facilitar las tareas de predicción y respuesta a peligros medioambientales como los tsunamis, terremotos, tornados y huracanes. Un ejemplo concreto es el de la aplicación Geoserver de la G-WADI (Información sobre los Recursos Hídricos y el Desarrollo) de la UNESCO, que se utiliza para fundamentar la planificación de emergencias y la gestión de riesgos hidrológicos, como inundaciones, sequías y fenómenos meteorológicos extremos. Su sistema auxiliar PERSIANN (*Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks*) es un algoritmo de recuperación de precipitaciones basado en satélites que proporciona información casi en tiempo real. El algoritmo del sistema de clasificación de nubes de PERSIANN (accesible en: http://hydis.eng.uci.edu/) se ha optimizado para observar precipitaciones extremas, en particular con una resolución espacial muy elevada, y se utiliza ampliamente en todo el mundo para el seguimiento de tormentas. También ofrece la aplicación móvil iRain (http://en.unesco.org/news/irain-new-mobile-app-promote-citizen-science-and-support-water-management), en la que la producción colectiva brinda oportunidades para recabar la participación de ciudadanos científicos en la recopilación de datos.

46. Curiosamente, incluso las empresas privadas han contribuido recientemente a la gestión de catástrofes. Un ejemplo de este tipo de colaboración es el del proyecto de previsión de inundaciones de Google basado en la IA (https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/). En este sentido, debe fomentarse el desarrollo de tecnologías de IA capaces de aportar beneficios potenciales a la gestión de catástrofes.

II.2.4. La IA y las ciencias sociales

47. En términos generales, la investigación en ciencias sociales tiene como objetivo determinar la estructura causal de las interacciones personales y sociales. Puesto que la mayoría de los fenómenos sociales se ven influidos por múltiples factores causales, los científicos sociales suelen basarse en el análisis estadístico de los datos empíricos pertinentes para determinar los factores causales destacados y la intensidad de sus efectos. En este sentido, resulta crucial distinguir las meras correlaciones estadísticas de las verdaderas conexiones causales. Ciertamente, la IA posee un potencial inequívoco para ayudar a los científicos sociales a navegar por enormes conjuntos de datos para encontrar mecanismos causales plausibles, así como para verificar la validez de los propuestos. Por otro lado, la IA puede "sobreajustar" los datos y plantear "pseudorelaciones" causales cuando no existe ninguna en realidad. Esta posibilidad puede dar lugar a controversias sociales, en especial cuando las relaciones causales propuestas sean éticamente sensibles, como las sugerencias de que existen diferencias raciales en cuanto a inteligencia. De nuevo en este caso no debemos aceptar automáticamente las "conclusiones" de la IA sin una evaluación humana.

II.2.5. La toma de decisiones basada en la IA

48. Los métodos basados en la IA pueden ejercer un enorme impacto en una amplia gama de campos, desde las profesiones jurídicas y el poder judicial, a la facilitación de la toma de decisiones de los organismos públicos legislativos y administrativos. Por ejemplo, pueden aumentar la eficiencia y la precisión de los juristas tanto en el asesoramiento como en los litigios, con beneficios para los abogados, sus clientes y la sociedad en su conjunto. Los sistemas de software existentes para jueces pueden complementarse y reforzarse mediante herramientas de IA que les ayuden en la redacción de nuevas resoluciones (CEPEJ, 2018).

49. Una cuestión clave en tales usos es la de la naturaleza y la interpretación de los resultados de los algoritmos, que no siempre son inteligibles para los seres humanos[2]. Esta cuestión puede extenderse al ámbito más amplio de la toma de decisiones basada en los datos disponibles. Se considera que un motor de IA, al poder analizar, procesar y clasificar cantidades muy grandes de datos potencialmente en rápida evolución y de muy diversa índole, es capaz de proponer –y si se permite, de adoptar– decisiones en situaciones complejas. Entre los ejemplos de tales usos analizados en este informe figuran los de la vigilancia medioambiental, la predicción y la respuesta en caso de catástrofe, la anticipación de disturbios sociales, y la planificación de la acción militar en el campo de batalla.

50. En cualquier caso, la validez de una decisión basada en la IA debe tratarse con precaución. Tal decisión no es necesariamente equitativa, justa, precisa o apropiada. Es susceptible de inexactitudes, resultados discriminatorios, sesgos implícitos o incorporados, y limitaciones del proceso de aprendizaje. Un ser humano no solo posee una "visión del mundo" mucho más amplia, sino que también cuenta con un conocimiento tácito que superará a la IA en situaciones críticas y complejas, como las que corresponden a las decisiones que se adoptan en el campo de batalla. Idealmente, una decisión sería la que adoptaría un ser humano si hubiera podido procesar la montaña de datos en cuestión en un plazo razonable. Sin embargo, los seres humanos poseen capacidades diferentes y toman decisiones basadas en arquitecturas para tal tarea fundamentalmente diferentes, incluida la sensibilidad respecto a posibles sesgos.

51. Es muy cuestionable que la IA –al menos en un futuro próximo– tenga la capacidad de hacer frente a datos ambiguos y en rápida evolución, o de interpretar y ejecutar cuáles habrían sido las intenciones humanas si los humanos hubieran podido hacerse cargo de los a datos complejos y multifacéticos en cuestión. Incluso contar con un humano "en el bucle" para moderar la decisión de una máquina puede no resultar suficiente para generar una decisión "buena": como la IA cognitiva no adopta decisiones del mismo modo que los humanos, el humano no dispondría del conocimiento y la información que necesitaría para decidir si la acción basada en los datos se atiene a sus intenciones humanas. Además, el comportamiento estocástico de la IA cognitiva, junto con la consiguiente incapacidad del ser humano para saber por qué el sistema ha efectuado una elección concreta, dan lugar a que sea menos probable que se confíe en tal elección.

52. Un relato admonitorio que ilustra algunos de los problemas asociados a utilizar la IA para facilitar la toma de decisiones en contextos sociales es el de la Allegheny Family Screening Tool (AFST), un modelo predictivo empleado para predecir el abandono y el abuso de menores en Allegheny, Pensilvania (véase https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/07/AFST-Frequently-Asked-Questions.pdf). La herramienta comenzó a aplicarse en la creencia de que las decisiones basadas en datos darían lugar a decisiones objetivas e imparciales que resolverían los problemas de la administración pública con recursos escasos. Es posible que la Autoridad que implementó dicha herramienta tuviera buenas intenciones. Sin embargo, en estudios

[2] Como señala K.D. Ashley: "puesto que un algoritmo de aprendizaje automático (AA) aprende reglas basadas en regularidades estadísticas que pueden sorprender a los seres humanos, es posible que las reglas del algoritmo no le parezcan necesariamente razonables a los humanos. […] Aunque las reglas inducidas por la máquina pueden dar lugar a predicciones precisas, no se refieren a la experiencia humana y puede que no sean tan inteligibles para los seres humanos como las reglas elaboradas manualmente por un experto. Dado que las normas que infiere el algoritmo [...] no reflejan necesariamente conocimientos o experiencias jurídicos explícitos, puede que no correspondan a los criterios de razonabilidad de un experto humano." (Ashley, 2017, p. 111)

recientes se ha argumentado que la herramienta AFST tiene consecuencias negativas para la población a la que confiaba servir (Eubanks, 2018b, p.190; Eubanks, 2018a). Efectúa un sobremuestreo de los pobres y utiliza valores sustitutivos para entender y predecir el abuso de menores de una manera que perjudica intrínsecamente a las familias trabajadoras desfavorecidas. De este modo, exacerba la discriminación estructural existente contra los pobres y ejerce un impacto desproporcionadamente adverso en las comunidades vulnerables.

53. En algunos contextos, el empleo de la IA como instrumento encargado de la toma de decisiones (ya sea con asistencia humana o de manera plenamente autónoma) puede incluso considerarse como un pacto con el diablo: para aprovechar la velocidad y las grandes capacidades de integración y clasificación de datos de un motor de IA, tendremos que renunciar a la capacidad de influir en la decisión de que se trate. Además, los efectos de estas decisiones pueden ser profundos, especialmente en situaciones de conflicto.

II.3. Cultura y diversidad cultural

54. Es probable que la IA tenga consecuencias sustanciales para la cultura y la expresión artística. Aunque todavía se encuentra en sus inicios, comenzamos a asistir a los primeros casos de colaboración artística entre los algoritmos inteligentes y la creatividad humana, lo que podría plantear en última instancia importantes desafíos respecto a los derechos de los artistas, los sectores de la cultura y la creación (SCC), y el futuro del patrimonio. Al mismo tiempo, es probable que el papel de los algoritmos en los medios de retransmisión (streaming) en línea y en la traducción automática repercuta en la diversidad cultural y el lenguaje.

II.3.1. Creatividad

55. La inteligencia artificial está cada vez más conectada con la creatividad humana y la práctica artística: desde el software de "autoajuste" que corrige automáticamente el tono de voz de los cantantes, hasta los algoritmos que ayudan a crear arte visual, componer música, o escribir novelas y poesía. La creatividad, entendida como la capacidad de producir contenidos nuevos y originales mediante la imaginación o la invención, desempeña un papel fundamental en las sociedades abiertas, inclusivas y plurales. Por esta razón, el impacto de la IA en la creatividad humana merece atención. Aunque la IA constituye una poderosa herramienta para la creación, plantea cuestiones importantes acerca del futuro del arte, los derechos y la remuneración de los artistas y la integridad de la cadena de valor creativa.

56. El caso del "nuevo Rembrandt" – en el que se elaboró un nuevo cuadro de Rembrandt utilizando la IA y una impresora 3D – es un buen ejemplo de esta situación (Microsoft Europe, 2016). Obras de arte como esta requieren una nueva definición de lo que significa ser un "autor", para hacer justicia a la obra creativa tanto del autor "original" como de los algoritmos y tecnologías que produjeron la obra de arte en sí. Esto plantea otra cuestión: ¿qué sucede cuando la IA posee la capacidad de crear obras de arte por sí misma? Si a un autor humano lo reemplazan máquinas y algoritmos, ¿en qué medida pueden atribuirse los derechos de autor? ¿Puede y debe reconocerse a un algoritmo como autor y que este goce de los mismos derechos que un artista?

57. Aunque la IA es claramente capaz de producir obras creativas "originales", siempre participan personas en el desarrollo de tecnologías y algoritmos de IA, y a menudo en la creación de obras de arte que sirven de inspiración para el arte generado mediante IA. Desde esta perspectiva, la IA puede percibirse como una nueva técnica artística que da lugar a un nuevo tipo de arte. Si queremos preservar la idea de autoría en las creaciones de la IA, es necesario realizar un análisis de los distintos autores que intervienen en cada obra de arte y de las relaciones entre ellos. En consecuencia, hemos de desarrollar nuevos marcos para diferenciar la piratería y el plagio de la originalidad y la creatividad, y reconocer el valor del trabajo creativo humano en nuestras interacciones con la IA. Estos marcos son necesarios para evitar la explotación deliberada del trabajo y la creatividad de los seres humanos, y para garantizar una remuneración y un

reconocimiento adecuados para los artistas, la integridad de la cadena de valor cultural y la capacidad del sector cultural para ofrecer empleos dignos.

II.3.2. Diversidad cultural

58. La IA también mantiene una estrecha relación con la diversidad cultural. Si bien puede impactar positivamente en los sectores de la cultura y la creación, no todos los artistas y emprendedores poseen las habilidades y los recursos para utilizar tecnologías basadas en la IA en la creación y la distribución de su trabajo. La lógica comercial de las grandes plataformas puede conducir a una mayor concentración de la oferta cultural, de los datos y los ingresos en manos de unos pocos agentes, con posibles consecuencias negativas para la diversidad de expresiones culturales en términos más generales, incluido el riesgo de generar una nueva brecha creativa y una creciente marginación de los países en desarrollo.

59. A medida que estas plataformas se convierten en el medio dominante para disfrutar de las obras de arte, resulta fundamental garantizar la diversidad y el acceso equitativo a dichas plataformas para los artistas de todos los géneros y procedencias. En este contexto, los artistas de los países en desarrollo requieren una consideración especial. Los artistas y los empresarios culturales deberían disponer de acceso a la formación, las oportunidades de financiación, las infraestructuras y los equipos necesarios para participar en estos nuevos ámbitos y mercados culturales.

60. Además, los algoritmos utilizados por empresas de retransmisión de medios en tiempo real como Spotify y Netflix ejercen una influencia significativa en la selección de música y películas que consume el público. Puesto que estas plataformas no solo ofrecen obras de arte, sino que también las *sugieren* para que sus usuarios las disfruten, es importante que sus algoritmos se diseñen de un modo que no privilegie obras de arte específicas por encima de otras mediante la limitación de sus sugerencias a las obras más dominantes de un determinado género, o a las opciones más populares de los usuarios y otros consumidores. Otras instituciones han expresado preocupaciones similares (ARCEP, 2018*).* La transparencia de estos algoritmos, y la asunción de responsabilidades respecto a los mismos, resultan esenciales para garantizar el acceso a expresiones culturales diversas y la participación activa en la vida cultural.

61. También en su relación con el patrimonio cultural, la IA puede desempeñar un papel importante. La IA puede utilizarse, por ejemplo, para supervisar y analizar los cambios en los emplazamientos del patrimonio, en lo que atañe a las presiones del desarrollo, el cambio climático, las catástrofes naturales y los conflictos armados. También puede utilizarse para realizar un seguimiento del tráfico ilícito de objetos culturales y la destrucción de los bienes culturales, así como para facilitar la recogida de datos destinados a tareas de recuperación y reconstrucción.

II.3.3. Lenguaje

62. En un mundo en proceso de rápida globalización, es probable que la traducción automática de idiomas desempeñe un papel cada vez más importante. Por este motivo, la IA ejercerá un impacto sustancial en el lenguaje y la expresión humana, en todas las dimensiones de la vida. Tal situación conlleva que deba asumirse la responsabilidad de tratar con cuidado las lenguas "naturales" (frente a los lenguajes artificiales y los códigos informáticos) y su diversidad. Después de todo, el lenguaje es la base de la identidad humana, la cohesión social, la educación y el desarrollo humano. Desde su fundación, la UNESCO ha reconocido la importancia del lenguaje para promover el acceso a una educación de calidad, construir sociedades del conocimiento integradoras, y transmitir el patrimonio y las expresiones culturales (UNESCO, 2002).

63. Un elemento central de la compleja relación entre la IA y el lenguaje es el papel intermediario de los "lenguajes formales" (aquéllos con términos derivados de un alfabeto). Las tecnologías de IA requieren a menudo que las palabras y frases expresadas en cualquiera de las numerosas lenguas naturales utilizadas en todo el mundo se traduzcan a lenguajes formales que puedan ser procesadas

por los ordenadores. La traducción de muchas lenguas naturales a lenguajes formales no es un proceso neutral, porque cada traducción del lenguaje natural al formal da lugar a una "pérdida" de significado, dado que no todas las especificidades e idiosincrasias de las lenguas pueden formalizarse por completo.

64. Un segundo elemento es la traducción entre lenguas naturales, que tiene lugar a través de estos lenguajes oficiales. Existen varios problemas intrínsecos con las traducciones automáticas: las palabras pueden tener significados diferentes en distintas lenguas, y puede haber una falta de correspondencia lingüística o conceptual entre las lenguas. En estos casos, la traducción resulta muy difícil, si no técnicamente imposible. Además, las connotaciones contextuales y culturales de las palabras y expresiones no siempre son plenamente traducibles. Aunque ha mejorado enormemente en los últimos años, al menos para las lenguas más comunes, la traducción automática resulta a menudo demasiado poco fiable para su uso, por ejemplo, en ámbitos técnicos en los que la precisión léxica y conceptual es crucial, o en la expresión cultural y la literatura.

65. Estos dos aspectos de la traducción automática tienen consecuencias importantes, no solo para la calidad de la traducción y el riesgo del mal entendimiento entre lenguas, sino también para la diversidad lingüística. Es muy probable que la traducción automática, al menos a corto plazo, se desarrolle fundamentalmente para las principales lenguas mundiales, en especial el inglés. La tecnología requiere grandes conjuntos de datos compilados a partir de las traducciones realizadas por seres humanos. Estos conjuntos de datos no suelen estar disponibles en cifras significativas en el caso de las lenguas menos habladas. Al mismo tiempo, esta tecnología también puede desempeñar un papel positivo, al permitir que las personas se expresen en lenguas menos habladas.

66. Un proceso análogo ya ha tenido lugar en la práctica con la radio. Mientras que la radio comercial produce en gran medida contenidos en lenguas de uso más generalizado, reforzando así las culturas incorporadas en las lenguas dominantes, las emisoras de las distintas comunidades generan a menudo contenidos en lenguas locales, potenciando de este modo el pluralismo y la diversidad en los medios de comunicación. Como se señala en el manual de la UNESCO sobre los medios de comunicación comunitarios: "[Los medios de comunicación comunitarios están] presentes en todas las regiones del mundo, ya que los movimientos sociales y las organizaciones de base comunitaria han buscado una vía de expresión de sus problemas, preocupaciones, culturas y lenguas" (UNESCO, 2013, p.7). Por lo tanto, los medios de comunicación de masas pueden ayudar en la práctica a preservar las lenguas y la diversidad cultural.

67. Del mismo modo, la traducción automática ya se ha utilizado como herramienta para promover la diversidad y proteger las lenguas indígenas. Por ejemplo, en Australia, un investigador del *Centro de Excelencia para la Dinámica del Lenguaje* del ARC ha registrado casi 50 000 horas de lenguaje hablado. Para procesar estas grabaciones, los lingüistas tuvieron que seleccionar segmentos cortos que incluyeran secciones clave de gramática y vocabulario, mediante la escucha de las grabaciones y su transcripción. Sin IA, esta labor habría llevado unos 2 millones de horas. Hasta la fecha, este uso de la IA ha facilitado la formulación de modelos de 12 lenguas indígenas habladas en Australia, entre las que figuran el *kunwok, kriol, mangarayi, nakkara, pitjantjatjara, warlpiri, y wubuy*, entre otras (O'Brien, 2018).

68. Estos ejemplos ponen de relieve que la IA, como cualquier tecnología, debe desarrollarse y utilizarse de maneras que no amenacen la diversidad cultural, sino que la preserven. Si queremos proteger el plurilingüismo y la interoperabilidad entre diferentes idiomas, deberán facilitarse los recursos técnicos y económicos adecuados para ello (Palfrey y Gasser, 2012; Santosuosso y Malerba, 2015).

II.4. Comunicación e información

69. La inteligencia artificial desempeña un papel cada vez más importante en el tratamiento, la estructuración y la provisión de información. El periodismo automatizado y el suministro algorítmico

de noticias en las redes sociales son solo algunos ejemplos de esta evolución, lo que plantea problemas relacionados con el acceso a la información, la desinformación, la discriminación, la libertad de expresión, la privacidad, la alfabetización mediática, y la cultura en materia de información. Al mismo tiempo, debe prestarse atención a las nuevas brechas digitales entre países y en cada uno de los diferentes grupos sociales.

II.4.1. Desinformación

70. La inteligencia artificial puede fortalecer el libre flujo de información y la actividad periodística, pero también puede utilizarse para propagar desinformación, a la que en ocasiones se denomina "noticias falsas". Ejemplos recientes, como el caso de Cambridge Analytica, han puesto de relieve que los algoritmos diseñados para evitar el sesgo político humano a la hora de decidir qué contenido aparecerá de manera prominente en las redes sociales pueden aprovecharse para promover deliberadamente la difusión de contenido inventado, manipulador y generador de división dirigido a grupos objetivo específicos. En algunos casos, este contenido puede incluir información formateada de manera fraudulenta como noticias, así como elementos que sirvan como propaganda emocional.

71. Esta práctica puede tener efectos negativos en las normas del debate civil e informado, en la confianza social y el debate público, o incluso en los procesos democráticos. La existencia de opiniones diferentes, en ocasiones polarizadas, constituye una característica habitual de cualquier sociedad abierta y democrática que ofrezca un espacio público libre y abierto. Sin embargo, los algoritmos de las redes sociales pueden exacerbar la polarización de las opiniones intensificando y amplificando el contenido emotivo mediante las opciones de "me gusta", "compartir", "retuits", el autocompletado en consultas de búsqueda y otras formas de recomendación y participación en línea, dando lugar a los denominados "filtros burbuja" y las "cámaras de eco", en lugar de proporcionar una infraestructura para el comentario y el debate. Las personas que comparten la misma "burbuja" pueden verse expuestas al filtro de los contenidos informativos y, a cambio, el espacio público abierto puede llegar a caracterizarse por grupos de opinión cada vez más homogeneizados, y al mismo tiempo cada vez más polarizados entre sí.

72. Aunque algunas grandes empresas de redes sociales comienzan a reconocer el problema y la necesidad de abordarlo con la intervención de las distintas partes interesadas, entre las que figura la sociedad civil y los reguladores estatales, parece que las soluciones todavía no están claras. Una forma de examinar distintas soluciones consiste en utilizar el marco D.A.A.M de la UNESCO (derechos, apertura, accesibilidad para todos, participación de múltiples partes interesadas) para identificar sistemáticamente dónde pueden realizarse las mejoras y cómo éstas se interrelacionan con la totalidad de los principios considerados.

73. En ocasiones, la moderación del contenido puede justificarse precisamente como medio para evitar la propagación de desinformación y los contenidos que inciten a la violencia, el odio y la discriminación, así como una vía para prevenir la comunicación personal agresiva. El filtrado lo pueden efectuar humanos, pero a menudo se facilita o incluso se automatiza mediante algoritmos de IA. El reto concreto en este caso no consiste únicamente en identificar el contenido ofensivo, sino también en evitar que el filtro resulte demasiado inclusivo y, en consecuencia, recibir acusaciones de censura automatizada y restricción de la expresión legítima. La respuesta a la desinformación y a la "incitación al odio" debe basarse en las normas internacionales sobre libertad de expresión y ha de ser conforme con las convenciones y declaraciones de las Naciones Unidas sobre la cuestión (artículo 19, 2018a).

II.4.2. Periodismo de datos y periodismo automatizado

74. La reciente aparición de una IA funcionalmente potente trae consecuencias para el periodismo de varias maneras diferentes. Por un lado, las crecientes posibilidades de utilizar datos y herramientas informáticas en la investigación periodística pueden reforzar la labor en este sector. Por otro lado, la AI también podría asumir ciertas tareas periodísticas. Las tareas rutinarias para las que se dispone de numerosos "datos de práctica" son las que más se prestan a ser imitadas por la

IA, y una parte sustancial del trabajo periodístico es, en realidad, rutinario: recabar y seleccionar datos pertinentes, resumir los resultados y describirlos de un modo claro. La IA ya realiza trabajos relativamente sencillos y de formato fijo en la redacción de artículos, en áreas donde se requieren actualizaciones continuas, como los análisis de mercados o los artículos deportivos. Este desarrollo es ambivalente: también puede liberar a los periodistas para que realicen tareas finales superiores en el terreno de la interpretación, el análisis, la verificación y la presentación de noticias.

75. La redacción automatizada de noticias sin intervención ni supervisión humanas es una realidad que a menudo se oculta al lector. Ya en 2006, algunos servicios de noticias (por ejemplo, Thomson Financial) anunciaron el uso de ordenadores para generar historias basadas en datos, con el fin de transmitir la información a sus usuarios de manera rápida. En 2014, Wibbitz (Israel) obtuvo el Gran Premio Netexplo en el Foro UNESCO/Netexplo, al proponer una aplicación que permite a los canales de noticias crear fácilmente vídeos utilizando contenidos de texto de Internet, proporcionando un resumen de las principales ideas del texto. En los últimos tiempos, varios de los principales medios de comunicación tradicionales utilizan el "periodismo robotizado": Le Monde, Asociación de Prensa, Xinhua, por nombrar algunos, han declarado que emplean algoritmos de generación de lenguaje natural para cubrir diferentes temas periodísticos.

76. La producción y difusión de contenidos mediáticos delegan cada vez más la autoridad analítica y de toma de decisiones atribuida a algoritmos complejos. Las organizaciones de los medios de comunicación dependen cada vez más de algoritmos que analizan las preferencias de los usuarios y los patrones de consumo de los medios (personalización). Aplicados al periodismo, se recurre a los algoritmos para analizar comunidades geográficas específicas en lo que atañe a sus variables demográficas, sociales y políticas, con el fin de producir la información más relevante para estas comunidades, incluidas las previsiones meteorológicas y los artículos sobre deporte. Esta práctica posee el potencial necesario para sostener el periodismo y los periódicos locales. De esta manera, la IA puede ayudar a fortalecer determinados modelos de negocio para el periodismo.

77. Al mismo tiempo, el periodismo basado en la IA plantea cuestiones de responsabilidad, transparencia y derechos de autor. La responsabilidad puede representar un problema cuando resulta complicado determinar la culpa en las informaciones elaboradas mediante algoritmos, por ejemplo, en casos de difamación. La transparencia y la credibilidad plantean problemas en los casos en los que los consumidores no saben o no pueden saber cuando el contenido lo generan máquinas, de qué fuentes procede, y en qué medida se ha verificado o es incluso falsa la información, con debates actuales sobre los "deep fakes" o "falsedades profundas" como casos extremos. Los derechos de autor constituyen un problema inminente, ya que los contenidos generados por la IA dependen cada vez menos de la intervención humana, razón por la que algunos argumentan que cierta forma de responsabilidad asociada a los derechos de autor debe atribuirse a los propios algoritmos.

78. Para abordar estos desafíos, muchos sostienen que los periodistas y editores deben colaborar con los tecnólogos que construyen los algoritmos. Un ejemplo de esta cooperación es el reciente lanzamiento de una plataforma de código abierto por Quartz AI Studio, un proyecto con sede en Estados Unidos para ayudar a los periodistas a utilizar el aprendizaje automático, asistiéndoles en diversas tareas.

II.5. La IA en la consolidación de la paz y la seguridad

79. De conformidad con la misión y el mandato de la UNESCO de promover y construir la paz, el presente estudio también pretende investigar el papel de la inteligencia artificial en los asuntos de la consolidación de la paz y la seguridad. El hecho de que se considere el posible uso militar de la IA no debilita en modo alguno su compromiso con la paz.

80. Se argumenta que la IA es capaz de analizar, procesar y clasificar cantidades muy grandes de datos en rápida evolución y de muy diversa índole (Payne, 2018; Roff, 2018; Gupta, 2018).

Entre los datos "objetivos" figurarían las imágenes por satélite y otras imágenes de vigilancia, y la inteligencia de señales y electrónica, mientras que los datos "subjetivos" podrían incluir informes, documentos, boletines noticias, aportaciones a las redes sociales y datos políticos y sociológicos. Se anuncia que la IA es capaz de clasificar esta enorme cantidad de datos para identificar amenazas externas e internas, descubrir los objetivos y estrategias de los distintos agentes, e interpretar las intenciones complejas y multifacéticas que subyacen a sus actividades y las estrategias respecto al modo de adelantarse a las acciones previstas o contrarrestar estas.

81. Esta herramienta de conocimiento de la situación podría constituir un poderoso instrumento para la prevención y la resolución de conflictos (Spiegeleire y cols., 2017). Podría proporcionar una visión de los factores que impulsan los empeños humanos y sus resultados, con una posible aplicación en el terreno de la desradicalización. La "inteligencia anticipativa" basada en el aprendizaje podría prever el desarrollo de disturbios sociales y la inestabilidad social, y proponer vías de prevención. Un conocimiento más profundo de los motores del conflicto podría facilitar la tarea de disuadir a los posibles generadores del mismo de materializar sus malas intenciones. Podríamos detectar patologías sociales en una etapa temprana, averiguar qué acciones podrían desacelerar una situación amenazante, o descubrir rutas no inflamatorias efectivas para contrarrestar los intentos de avivar los conflictos sectarios. A escala social, al facilitar el seguimiento y la comprensión de las dinámicas que fortalecen o debilitan la resiliencia social, la IA puede conducirnos a una sociedad más resistente, y ayudarnos a avanzar hacia un mundo más pacífico y libre de conflictos.

82. En el lado negativo, *la IA transformará la naturaleza y la práctica del conflicto,* con el consiguiente impacto en la sociedad que se extenderá mucho más allá de los asuntos estrictamente militares (Payne, 2018; Spiegeleire y cols., 2017). No sólo cambiará cómo se utiliza la fuerza explosiva al aumentar la efectividad del despliegue de sistemas de armas, sino que también promete mejorar drásticamente la velocidad y la precisión en todos los terrenos, desde la logística militar, la inteligencia y el conocimiento de las situaciones, hasta la planificación y la ejecución y las operaciones en el campo de batalla. El sistema de IA en sí podría utilizarse para que formule sus propias sugerencias respecto a las acciones que deben emprenderse: podría crear un conjunto de órdenes que exploten las debilidades del enemigo que ha identificado conforme a su propio análisis o tras encontrar patrones en las acciones del enemigo/insurgentes, idear contramedidas a las acciones de agresión predichas. También podría llevar a cabo sus propios "juegos de guerra" para comprobar las respuestas probables a acciones concretas.

83. La velocidad a la que podrían funcionar estas herramientas de planificación elevaría la capacidad de actuar en situaciones que cambian rápidamente. Cabe prever, por ejemplo, el desarrollo de una respuesta algorítmica a los ataques coordinados a cargo de, por ejemplo, enjambres de drones y otros activos sin presencia humana, como los misiles entrantes. La velocidad de la respuesta basada en la IA puede percibirse como un incentivo para su empleo y, por tanto, resultar potencialmente desestabilizadora. O, de hecho, desastrosa, como lo han demostrado ejemplos del pasado de alertas generadas por máquinas que, afortunadamente, no han sido atendidas por el mando humano interviniente. En cualquier caso, un Estado que no transite por esta vía de respuesta de la IA se encontraría en una situación de enorme desventaja, lo que alienta la proliferación de este tipo de capacidad.

84. Existe la posibilidad de que una máquina encargada de la toma de decisiones y asistida por la IA implemente sus propias decisiones de ataque y muerte sin intervención humana, como, por ejemplo, en el caso de un arma totalmente autónoma. La idea de tal entidad *no humana* con un mandato específico podría cambiar radicalmente nuestra interpretación de la política a todos los niveles. Además, la cercanía de los posibles usos militares de la IA a su desarrollo civil ("facilidad de la proliferación de armas") significa que no es una categoría específicamente delimitada, una característica que complica tanto la ética como la regulación de su desarrollo y aplicación.

85. Si bien podría considerarse que la IA representa simplemente otra revolución más en el ámbito militar que permite a las fuerzas armadas realizar tareas similares con herramientas similares, tal

vez su verdadero potencial "revolucionario" (Payne, 2018; Spiegeleire y cols., 2017) radica en la *transformación del concepto de "fuerza armada"* en una entidad cuyas armas son más sutiles que los dispositivos explosivos. El poder de la IA en los conflictos reside no solo en mejorar las tecnologías físicas, sino también en redefinir lo que podría ser una "fuerza armada".

86. Ya estamos asistiendo a este proceso en el contexto cibernético, en el que la IA proporciona capacidades tanto de defensa como de ataque. Mediante el cotejo de patrones, el aprendizaje profundo y la observación de las desviaciones respecto a la actividad normal, pueden detectarse vulnerabilidades de software y, a continuación, convertirse en un arma para eludir las defensas. Las redes neuronales profundas pueden detectar y prevenir las intrusiones. Para ser eficaces, las defensas cibernéticas tendrán que funcionar a gran velocidad y, por consiguiente, poseer un alto grado de autonomía.

87. La propaganda constituye otra arma que la IA ha fortalecido. La facilidad de simular voces, y falsificar imágenes y noticias, así como de propagarlas entre determinadas audiencias, amenaza la ingeniería social y la conformación (deformación) de la opinión pública. En esencia, la IA facilita la mentira persuasiva y el perfeccionamiento de la falsificación. La consiguiente amenaza para la confianza en la integridad de la información aumenta la posibilidad de estimar erróneamente la intención de un adversario percibido, tanto táctica como estratégicamente.

88. La IA también propicia el sabotaje económico y la alteración de infraestructuras críticas. Al trasladar la guerra radio-electrónica al modo cognitivo, la IA pueden resultar esencial en la tarea de interferir en el acceso al espectro electromagnético. Ya se comercializan sistemas que utilizan el aprendizaje automático, los algoritmos "inteligentes" y el procesamiento adaptativo de señales.

89. Por último, en lo que se refiere a la seguridad *interna* de los Estados, el uso del análisis de conjuntos de datos y el reconocimiento facial dan lugar a una nueva relación entre la sociedad y las instituciones encargadas de protegerla. Obviamente este proceso tiene implicaciones éticas significativas.

II.6. La IA y la igualdad de género

90. Los sistemas de IA tienen implicaciones significativas para la igualdad de género, ya que pueden reflejar los sesgos sociales existentes, con la posibilidad de exacerbarlos. La mayoría de los sistemas de IA se construyen utilizando conjuntos de datos que reflejan el mundo real, que puede ser imperfecto, injusto y discriminatorio (Marda, 2018). Recientemente, se descubrió que una herramienta de contratación utilizada por Amazon era sexista, ya que daba prioridad a los candidatos masculinos a puestos técnicos (Reuters, 2018). Estos sistemas pueden resultar peligrosos, no solo porque perpetúan las desigualdades de género en la sociedad, sino también porque integran tales desigualdades de manera opaca, al tiempo que se las aclama como "objetivos" y "precisos" (O'Neil, 2018).

91. Estas desigualdades son principalmente el resultado del modo en que aprenden las máquinas. De hecho, dado que el aprendizaje automático se basa en los datos que se introducen, debe prestarse especial atención para promover la recogida de datos que tienen en cuenta las cuestiones de género, y desglosados por sexo. En el caso de la herramienta de contratación de Amazon, el sesgo surgió porque la herramienta aprendía de los candidatos anteriores de la empresa –que eran predominantemente varones– y había "aprendido" que debía preferirse a los solicitantes varones respecto a las candidatas mujeres (Short, 2018). Por tanto, prestar atención a los datos sesgados ayudaría a atenuar el punto flaco relativo a la mejor manera de adaptar y diseñar los sistemas de IA tanto para hombres como para mujeres. Además, la aplicación de datos desglosados por sexo a los análisis de la IA representa una oportunidad para comprender mejor las cuestiones de género a las que nos enfrentamos actualmente.

92. Es importante señalar que las desigualdades de género comienzan en las primeras etapas de conceptualización y diseño de los sistemas de IA. La disparidad de género en los ámbitos técnicos es bien conocida y evidente (Hicks, 2018), desde las brechas salariales, a los ascensos (Brinded, 2017). A este fenómeno se le conoce en general como un proceso de "fugas en el camino", en el que la participación femenina en el campo de la tecnología y la ingeniería cae en un 40% entre el momento en que las alumnas obtienen su titulación, y en el que se convierten en ejecutivas en el sector (Wheeler, 2018). La baja proporción de mujeres entre el personal que se dedica a la IA –y en el desarrollo de competencias digitales en general– significa que la voz de las mujeres no se encuentra representada por igual en los procesos de toma de decisiones que acompañan al diseño y el desarrollo de los sistemas de IA. Como resultado, nos arriesgamos a construir estas tecnologías únicamente para algunos grupos demográficos (Crawford, 2016).

93. Además, los sesgos que las personas aplican en su vida diaria pueden reflejarse e incluso amplificarse a través del desarrollo y el uso de sistemas de IA. La "atribución de género" a los asistentes digitales, por ejemplo, puede reforzar la conceptualización de las mujeres como serviles y obedientes. De hecho, las voces femeninas se eligen de manera habitual como bots de asistencia personal, atendiendo principalmente tareas de servicio al cliente, mientras que la mayoría de los bots de servicios profesionales como los de los sectores jurídico y financiero, por ejemplo, se codifican como voces masculinas. Estas opciones tienen consecuencias educativas respecto al modo en que entendemos las competencias masculinas frente a las femeninas, y a la manera en que definimos las posiciones de autoridad frente a las subordinadas. Además, el concepto de "género" en los sistemas de IA suele consistir en una opción sencilla: hombre o mujer. De este modo se ignora y excluye activamente a las personas transgénero, y es posible que se las discrimine de maneras humillantes (Costanza-Chock, 2018).

II.7. África y los retos de la IA

94. África, al igual que otras regiones en desarrollo, se enfrenta a la aceleración en el uso de las tecnologías de la información y la IA. La nueva economía digital emergente plantea importantes desafíos sociales y oportunidades para las creativas sociedades africanas.

95. Concretamente, en términos de conectividad de las infraestructuras, África adolece de un déficit enorme y se sitúa muy por detrás de otras regiones en desarrollo; así, las conexiones domésticas, los vínculos regionales y el acceso continuo a la electricidad constituyen una gran desventaja. Los servicios de infraestructura se pagan a un precio elevado, aun cuando son cada vez más los africanos -incluso en los barrios de chabolas de las ciudades- que poseen teléfono móvil propio.

96. Los problemas de desarrollo a los que se enfrentan los países africanos son numerosos. El marco de los derechos humanos y los objetivos de desarrollo sostenible (ODS) proporcionan una forma coherente de orientar el desarrollo de la IA. En este sentido, ¿cómo pueden compartirse y orientarse la tecnología y el conocimiento de la IA a través de las prioridades definidas por los propios países en desarrollo? Se incluyen aquí retos como los que atañen a las infraestructuras, las destrezas, las lagunas de conocimiento, las capacidades de investigación y la disponibilidad de datos locales, como se expresó en el Foro de la UNESCO sobre Inteligencia Artificial en África que tuvo lugar en la Universidad Politécnica Mohammed VI, en Benguérir (Marruecos), los días 12 y 13 de diciembre de 2018.

97. El papel de las mujeres es crucial. Como agentes económicos de gran dinamismo en África, las mujeres realizan la mayoría de las actividades agrícolas, poseen un tercio del total de las empresas y pueden representar, en algunos países, hasta el 70% de los empleados. Constituyen las principales palancas de la economía doméstica y el bienestar familiar, y desempeñan un papel de liderazgo absolutamente indispensable en sus respectivas comunidades y naciones. Al situar la igualdad de género en el centro de su estrategia para promover el desarrollo en África, el Banco Africano de Desarrollo reconoce el papel fundamental de la paridad de género en la consecución de un crecimiento integrador y en el surgimiento de sociedades resilientes. El acceso a la educación, a la capacitación para la IA y, más globalmente, a las tecnologías de la información y la comunicación (TIC) son elementos clave del empoderamiento de las mujeres para evitar su marginación.

98. Prestando especial atención a la investigación científica, la ciencia, la tecnología, la ingeniería y las matemáticas, junto con la educación para la ciudadanía basada en valores, derechos y obligaciones, la IA debe integrarse en las políticas y estrategias de desarrollo nacional sobre la base de las culturas, los valores y los conocimientos endógenos para desarrollar las economías africanas.

III. INSTRUMENTO NORMATIVO

III.1. Declaración frente a Recomendación

99. El Grupo de Trabajo examinó cuidadosamente dos de las herramientas normativas de la UNESCO –la Declaración y la Recomendación–, vinculadas a los análisis de los dos primeros apartados de este estudio preliminar sobre la ética de la IA. El Grupo de Trabajo también se sirvió de la experiencia previa de la COMEST, que puso en marcha la Declaración de Principios Éticos en relación con el Cambio Climático de 2017 y participó en la revisión de la Recomendación sobre la Ciencia y los Investigadores Científicos de ese mismo año. El Grupo de Trabajo sopesó las ventajas e inconvenientes de cada una de estas dos herramientas normativas.

100. En lo que atañe a la propuesta de Declaración sobre la ética de la inteligencia artificial, el Grupo de Trabajo observó el reciente aumento del número de declaraciones de principios éticos sobre IA en 2018. El *Declaración de Montreal para el desarrollo responsable de la IA* (Universidad de Montreal, 2018), la *Declaración de Toronto sobre: protección del derecho a la igualdad y a la no discriminación en los sistemas de aprendizaje automático* (Amnistía Internacional y Access Now, 2018), y la Declaración del Future of Life Institute relativa a los *Principios de Asilomar sobre la IA* (Future of Life Institute, 2017) tienen su origen en diferentes iniciativas y cuentan con el apoyo de diversas organizaciones (universidades, administraciones públicas, asociaciones profesionales, empresas, ONG). Debemos añadir a este conjunto de declaraciones varias propuestas éticas como las *Directrices éticas sobre una IA confiable* del Grupo de Expertos de Alto Nivel sobre IA de la

Comisión Europea, que se basa en los derechos humanos, y el segundo documento de la IEEE (actualmente objeto de consulta), en el que se propugna una concepción conforme a la ética (*Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems),* y que se dirige a los ingenieros y apunta a integrar diversos valores en los sistemas inteligentes autónomos. Todas estas iniciativas son positivas, ya que ponen en marcha debates sobre la ética de la IA a diferentes escalas.

101. No obstante, el Grupo de trabajo concluyó que existía una gran heteronomía en los principios y en la implantación de los valores promovidos por unos u otros. Esta heteronomía es tanto la consecuencia de la definición que se ha elegido para la IA, como de los objetivos que se persiguen: gobernanza, formación de los ingenieros, y política pública. La cuestión es la siguiente: ¿permitiría una Declaración de la UNESCO sobre la ética de la IA que esta heteronomía se atenuase en torno a unos pocos principios rectores que respondieran de manera exhaustiva a las cuestiones éticas de la IA, así como a las preocupaciones específicas de la UNESCO en los ámbitos de la educación, la cultura, la ciencia y la comunicación? El Grupo de Trabajo cree que esto resultaría posible, pero con el riesgo de que, durante el proceso que conduce a la Declaración, los Estados miembros convengan esencialmente algunos principios generales, abstractos y no vinculantes, ya que se trata de una Declaración. Desde esta perspectiva, ¿aportaría la Declaración de la UNESCO sobre la ética de la IA valor añadido frente a otras declaraciones e iniciativas en curso? Es cuestionable que este instrumento se establezca inmediatamente como una referencia internacional, en un contexto de competencia entre marcos éticos, en un período en el que emergen distintas tecnologías y sus usos no se han estabilizado aún.

102. Por tanto, el Grupo de trabajo consideró si una Recomendación constituiría entonces una herramienta más adecuada en la situación actual. A escala internacional y europea y en el contexto político nacional de varios países, existe una tendencia hacia formas similares de regulación con respecto a la economía digital, pero teniendo en cuenta además las relaciones entre las dos principales potencias digitales (Estados Unidos y China). El aumento de las críticas sobre la falta de transparencia, los sesgos o las formas de actuar de las grandes empresas, y el incremento de la desconfianza de la población frente a los ciberataques, crean un nuevo clima político que repercute en el desarrollo de la IA. En este sentido, el movimiento a favor de la regulación digital, iniciado por la Unión Europea respecto a la protección de datos personales, podría ampliarse a la escala internacional en campos emergentes como el de la IA. Sin embargo, a esta escala, las herramientas se encuentran aún en sus primeras etapas de desarrollo, si bien la estrategia de la OCDE formulada a través de su Grupo de Expertos en Inteligencia Artificial (AIGO) hace hincapié en la responsabilidad, la seguridad, la transparencia, la protección y la rendición de cuentas:

> la OCDE asiste a los gobiernos mediante el análisis de políticas, el diálogo y la participación, y la identificación de buenas prácticas. Dedicamos un esfuerzo significativo a las tareas de determinación de los impactos económicos y sociales de las tecnologías y aplicaciones de la IA y sus implicaciones para la formulación de políticas. Esta labor comprende mejorar la medición de la IA y sus impactos, así como arrojar luz sobre cuestiones políticas importantes como la evolución del mercado de trabajo y las destrezas para la era digital, la privacidad, la asunción de responsabilidades respecto a las decisiones basadas en la IA, y las cuestiones de responsabilidad, seguridad y protección que genera la IA. (OCDE, 2019)

103. Las prioridades en materia de política pública de la OCDE responden más a cuestiones de gobernanza y buenas prácticas en relación con la IA. Parece en este caso que el enfoque de la UNESCO podría ser complementario a escala internacional al de la OCDE, pero haciendo hincapié en aspectos que suelen pasarse por alto, como la cultura, la educación, la ciencia y la comunicación. Estas dimensiones afectan directamente a las personas y a las poblaciones en su vida diaria y en sus aspiraciones individuales y colectivas. El enfoque de la UNESCO respecto a una Recomendación sobre la ética de la IA se presentaría como una alternativa complementaria a una visión de la gobernanza económica. El Grupo de Trabajo cree por tanto que, al poner en marcha una Recomendación, aunque requiera más tiempo y energía que una Declaración, la UNESCO podría distinguirse no solo en lo que se refiere al contenido ético, sino también a través de

propuestas específicas dirigidas a los Estados miembros. Uno de los objetivos es potenciar y reforzar la capacidad de los Estados para intervenir en áreas esenciales que se ven afectadas por el desarrollo de la IA, como la cultura, la educación, la ciencia y la comunicación.

104. La Recomendación debe contener dos dimensiones. La primera alude a la afirmación de una serie de principios básicos para una ética de la IA. La segunda consiste en el esbozo de propuestas específicas para ayudar a los Estados a supervisar y regular los usos de la IA en las áreas bajo el mandato de la UNESCO a través del mecanismo de presentación de informes de la Recomendación, así como a identificar herramientas de evaluación ética para revisar periódicamente sus políticas y orientar el desarrollo de la IA. En este sentido, la UNESCO se encontraría en una posición única para ofrecer una perspectiva multidisciplinar, y para establecerse como una plataforma universal para el desarrollo de una Recomendación sobre la ética de la IA. En concreto, la UNESCO podría reunir a países tanto desarrollados como en desarrollo, diferentes perspectivas culturales y morales, así como diversas partes interesadas en los ámbitos público y privado, en un proceso verdaderamente internacional para elaborar un conjunto integral de principios y propuestas para la ética de la IA.

105. En el siguiente apartado se presentan algunas de estas propuestas.

III.2. Propuestas respecto a un instrumento normativo

106. Sobre la base de su análisis de las posibles implicaciones de la inteligencia artificial para la sociedad, el Grupo de Trabajo desea proponer diversos elementos que podrían incluirse en una eventual Recomendación sobre la ética de la IA. Estas propuestas incorporan la perspectiva global de la UNESCO, así como los ámbitos específicos de competencia de la Organización.

107. En primer lugar, el Grupo de Trabajo desea proponer una serie de principios genéricos para el desarrollo, la implantación y el uso de la IA. Estos principios son los que siguen:

a) **Derechos humanos**: la IA debe desarrollarse e implementarse de acuerdo con las normas internacionales de los derechos humanos.

b) **Integración**: la IA debe ser inclusiva, con el objetivo de evitar sesgos, propiciar la diversidad y prevenir una nueva brecha digital.

c) **Prosperidad**: la IA debe desarrollarse para mejorar la calidad de vida.

d) **Autonomía**: la IA debe respetar la autonomía humana mediante la exigencia del control humano en todo momento.

e) **Explicabilidad**: la IA debe ser explicable, capaz de proporcionar una idea de su funcionamiento.

f) **Transparencia:** los datos utilizados para capacitar los sistemas de IA deben ser transparentes.

g) **Conocimiento y capacitación**: el conocimiento de los algoritmos y una comprensión básica del funcionamiento de la IA son necesarios para capacitar a los ciudadanos.

h) **Responsabilidad**: los desarrolladores y las empresas deben tener en cuenta la ética al desarrollar los sistemas inteligentes autónomos.

i) **Asunción de responsabilidades**: Deben desarrollarse mecanismos que permitan atribuir responsabilidades respecto a las decisiones basadas en la IA y la conducta de los sistemas de IA.

j) **Democracia**: la IA debe desarrollarse, implantarse y utilizarse con arreglo a principios democráticos.

k) **Buena gobernanza**: los gobiernos deben presentar informes periódicos sobre su utilización de la IA en los ámbitos de la actividad policial, la inteligencia y la seguridad.

l) **Sostenibilidad**: En todas las aplicaciones de la AI, los beneficios potenciales deben equilibrarse con el impacto medioambiental del ciclo de producción completo de la IA y las TI.

108. En particular, al Grupo de Trabajo le gustaría señalar algunas cuestiones éticas fundamentales en relación con el enfoque específico de la UNESCO:

a) **Educación**: la IA requiere que la educación fomente la adquisición de competencias en materia de IA, el pensamiento crítico, la resiliencia en el mercado laboral y la instrucción ética de los ingenieros.

b) **Ciencia**: la IA requiere una introducción responsable en la práctica científica y en la toma de decisiones basada en sistemas de IA, exigiendo la evaluación y el control humanos y que se evite la exacerbación de las desigualdades estructurales.

c) **Cultura**: la inteligencia artificial debe fomentar la diversidad cultural, la inclusión y el fomento de la experiencia humana, evitando una profundización de la brecha digital. Debe promoverse un enfoque multilingüe.

d) **Comunicación e información**: la inteligencia artificial debe consolidar la libertad de expresión, el acceso universal a la información, la calidad del periodismo y los medios libres, independientes y pluralistas, evitando al mismo tiempo la propagación de la desinformación. Debe promoverse la gobernanza de múltiples partes interesadas.

e) **Paz**: para contribuir a la paz, la IA podría utilizarse en la obtención de información sobre los factores que impulsan los conflictos, y nunca debería funcionar al margen del control humano.

f) **África**: la IA debe integrarse en las políticas y estrategias de desarrollo nacional sobre la base de las culturas, los valores y los conocimientos endógenos para desarrollar las economías africanas.

g) **Género**: debe evitarse el sesgo de género en el desarrollo de algoritmos, en los conjuntos de datos utilizados para su formación, y en su uso en la toma de decisiones.

h) **Medio ambiente**: la IA debe desarrollarse de manera sostenible, teniendo en cuenta todo el ciclo de producción de la IA y las TI. La IA puede utilizarse en tareas de vigilancia medioambiental y gestión de riesgos, así como para prevenir y mitigar las crisis medioambientales.

BIBLIOGRAFÍA

AI Now. 2016. *The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*. New York, The White House and the New York University's Information Law Institute. Available at: https://ainowinstitute.org/AI_Now_2016_Report.pdf

Ajunwa, I., Crawford, K., and Schultz, J. 2017. Limitless Worker Surveillance. *California Law Review.* No. 735, pp. 101-142.

Allen, G. and Chan, T. 2017. Artificial Intelligence and National Security. *Harvard Kennedy School, Belfer Center for Science and International Affairs.* Online. Available at: https://www.belfercenter.org/publication/artificial-intelligence-and-national-security

Amnesty International and Access Now. 2018. *The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems.* Toronto, RightsCon 2018. Available at: https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf

ARCEP (Autorité de régulation des communications électroniques et des postes). 2018. *Smartphones, tablets, voice assistants... Devices, the weak link in achieving an open Internet*. Paris, ARCEP. Available at: https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018-ENG.pdf

Article 19. 2018a. *Free speech concerns amid the "fake news" fad*. Online. Available at: https://www.article19.org/resources/free-speech-concerns-amid-fake-news-fad/

Article 19. 2018b. *Privacy and Freedom of Expression in the Age of Artificial Intelligence*. Online. Online. Available at: https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf

Ashley, K.D. 2017. Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge, Cambridge University Press.

Boden, M.A. 2016. *AI: Its Nature and Future*. Oxford, Oxford University Press.

Brinded, L. 2017. "Robots are going to turbo charge one of society's biggest problems", *QUARTZ* (28 December 2017). Online. Available: https://qz.com/1167017/robots-automation-and-ai-in-the-workplace-will-widen-pay-gap-for-women-and-minorities/

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B. and Anderson, H. 2018. *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Available at: https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Bunnin, N. and Yu, J. 2008. *The Blackwell dictionary of western philosophy*. John Wiley & Sons.

Butterfield, A., Ngondi, G.E. and Kerr, A. eds. 2016. *A dictionary of Computer Science*. Oxford, Oxford University Press.

Costanza-Chock, S. 2018. "Design justice, AI, and escape from the matrix of domination", *Journal of Design and Science.* Online. Available at: https://jods.mitpress.mit.edu/pub/costanza-chock

Crawford, K. 2016. "Artificial Intelligence's White Guy Problem", *The New York Times* (Opinion, 25 June 2016). Online. Available at: https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

Crawford, K. 2017. 'The Trouble with Bias', NIPS 2017 Keynote. Available at: https://www.youtube.com/watch?v=fMym_BKWQzk

Cummings, M. L., Roff, H. M., Cukier, K., Patakilas, J. and Bryce, H. 2018. *Artificial Intelligence and International Affairs: Disruption Anticipated.* Chatham House Report. Available at: https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf

Brookfield Institute and Policy Innovation Hub (Ontario). 2018. *Policymakers: Understanding the Shift*. Online. Available at: https://brookfieldinstitute.ca/wp-content/uploads/Brookfield-Institute_-The-AI-Shift.pdf

Eubanks, V. 2018a. "A Child Abuse Prediction Model Fails Poor Families", *WIRED*. Online. Available at: https://www.wired.com/story/excerpt-from-automating-inequality/

Eubanks, V. 2018b. *Automating Inequality: How high tech tools profile, police, and punish the poor*. New York, St. Martin's Press.

European Commission (EC). 2018. *Artificial Intelligence for Europe.* Communication from the Commission to the European Parliament, the European council, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels, European Commission. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN

European Commission for the Efficiency of Justice (CEPEJ). 2018. *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment.* Strasbourg, CEPEJ. Available at: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

European Group on Ethics in Science and New Technologies (EGE). 2018. *Statement on AI, Robotics, and Autonomous System*. Brussels, European Commission. Available at: https://publications.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382

Executive Office of the President (USA). 2016. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*. Washington, D.C., Executive Office of the President. Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

Frankish, K. and Ramsey, W.M. eds. 2014. *The Cambridge handbook of artificial intelligence*. Cambridge, Cambridge University Press.

Future of Life Institute. 2017. *Asilomar AI Principles.* Cambridge, Future of Life Institute. Available at: https://futureoflife.org/ai-principles/?cn-reloaded=1

Gupta, D.K. 2018. "Military Applications of Artificial Intelligence", *Indian Defence Review* (22 March 2019). Online. Available at: http://www.indiandefencereview.com/military-applications-of-artificial-intelligence/

Heacock, M., Kelly, C.B., Asante, K.A., Birnbaum, L.S., Bergman, Å.L., Bruné, M.N., Buka, I., Carpenter, D.O., Chen, A., Huo, X. and Kamel, M. 2015. "E-waste and harm to vulnerable populations: a growing global problem", *Environmental health perspectives*, Vol. 124, No. 5, pp. 550-555.

Hicks, M. 2018. "Why tech's gender problem is nothing new", *The Guardian* (12 October 2018). Online. Available at: https://amp.theguardian.com/technology/2018/oct/11/tech-gender-problem-amazon-facebook-bias-women?__twitter_impression=true

Hinchliffe, T. 2018. “Medicine or poison? On the ethics of AI implants in humans”, *The Sociable*. Online. Available at: https://sociable.co/technology/ethics-ai-implants-humans/

House of Lords. 2017. *AI in the UK: ready, willing and able?* London, House of Lords Select Committee on Artificial Intelligence. Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Illanes, P., Lund, S., Mourshed, M., Rutherford, S. and Tyreman, M. 2018. *Retraining and reskilling workers in the age of automation*. Online, McKinsey Global Institute. Available at: https://www.mckinsey.com/featured-insights/future-of-work/retraining-and-reskilling-workers-in-the-age-of-automation

Institute of Electrical and Electronic Engineers (IEEE). 2018. *Ethically Aligned Design – Version 2 for Public Discussion*. New Jersey, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Available at: https://ethicsinaction.ieee.org/

Laplante, P.A. 2005. *Comprehensive dictionary of electrical engineering*. Boca Raton, CRC Press.

Latonero, M. 2018. *Governing Artificial Intelligence: Upholding Human Rights & Dignity*. Data & Society. Available at: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

Marda, V. 2018. “Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making”, *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*. Online. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3240384

Matias, Y. 2018. Keeping people safe with AI-enabled flood forecasting. *The Keyword* (24 September 2018). Online. Available at: https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/

Matsumoto, D.E. 2009. *The Cambridge dictionary of psychology.* Cambridge, Cambridge University Press.

McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C. E. 2006 [1955]. “A proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, *AI Magazine*, vol. 27, no. 4, pp.12-14.

Microsoft Europe. 2016. “The Next Rembrandt”, *Microsoft News Centre Europe*. Online. Available at: https://news.microsoft.com/europe/features/next-rembrandt/

National Science and Technology Council (USA). 2016. *The National Artificial Intelligence Research and Development Strategic Plan*. Washington, D.C., National Science and Technology Council. Available at: https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf

O’Brien, A. 2018. “How AI is helping preserve Indigenous languages”, *SBS News*. Online. Available at: https://www.sbs.com.au/news/how-ai-is-helping-preserve-indigenous-languages

O’Neil, C. 2018. “Amazon’s Gender-Biased Algorithm Is Not Alone”, *Bloomberg Opinion* (16 October 2018). Online. Available at: https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone

OECD. 2019. *Going Digital.* Paris, OECD. Available at: http://www.oecd.org/going-digital/ai/

Oppenheimer, A. 2018. *¡Sálvese quien pueda!: El futuro del trabajo en la era de la automatización*. New York, Vintage Español.

Palfrey, J.G. and Gasser, U. 2012. *Interop: The Promise and Perils of Highly Interconnected Systems*. New York, Basic Books.

Payne, K. 2018. "Artificial Intelligence: A Revolution in Strategic Affairs?", *Survival*, Vol. 60, No. 5, pp. 7-32.

Peiser, J. 2019. "The Rise of the Robot Reporter", *The New York Times* (5 February 2019). Online. Available at: https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html

Reuters. 2018. "Amazon ditched AI recruiting tool that favored men for technical jobs", *The Guardian* (11 October 2018). Online. Available at: https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

Roff, H.M. 2018. 2018. "COMPASS: a new AI-driven situational awareness tool for the Pentagon?", *Bulletin of the Atomic Scientists* (10 May 2018). Online. Available at: https://thebulletin.org/2018/05/compass-a-new-ai-driven-situational-awareness-tool-for-the-pentagon/

Rosenberg, J.M. 1986. *Dictionary of artificial intelligence and robotics*. New York, John Wiley & Sons.

Russell, S.J. and Norvig, P. 2016. *Artificial Intelligence: A Modern Approach*, 3rd ed. Harlow, Pearson.

Santosuosso, A. and Malerba, A. 2015. "Legal Interoperability As a Comprehensive Concept in Transnational Law", *Law, Innovation and Technology*, Vol. 5, No. 1, pp. 51-73.

Short, E. 2018. "It turns out Amazon's AI hiring tool discriminated against women", *Siliconrepublic* (11 October 2018). Online. Available at: https://www.siliconrepublic.com/careers/amazon-ai-hiring-tool-women-discrimination

Spiegeleire, S. De, Maas, M. and Sweijs, T. 2017. *Artificial Intelligence and the Future of Defence*. The Hague, The Hague Centre for Strategic Studies.

UNI Global Union. 2016. Top 10 principles for ethical artificial intelligence. Switzerland, UNI Global Union. Available at: http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf

UNICEF. 2017. *Children in a Digital World*. New York UNICEF. Available at: https://www.unicef.org/publications/files/SOWC_2017_ENG_WEB.pdf

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2002. *UNESCO Universal Declaration on Cultural Diversity: a vision, a conceptual platform, a pool of ideas for implementation, a new paradigm*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000127162

UNESCO. 2013. *Community Media: A Good Practice Handbook*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000215097

UNESCO. 2015a. *Keystones to foster inclusive knowledge societies: access to information and knowledge, freedom of expression, privacy and ethics on a global internet.* Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000232563

UNESCO. 2015b. *Outcome document of the "CONNECTing the Dots: Options for Future Action" Conference*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000234090

University of Montreal. 2018. *Montreal Declaration for a Responsible Development of AI*. Montreal, University of Montreal. Available at: https://www.montrealdeclaration-responsibleai.com/

Vernon, D. 2014. *Artificial cognitive systems: A primer*. Cambridge, MIT Press.

Villani, C., Schoenauer, M., Bonnet, Y., Berthet, C., Cornut, A.-C., Levin, F. and Rondepierre, B. 2018. *For A Meaningful Artificial Intelligence: Towards a French and European Strategy*. Paris. Available at:

https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

Wheeler, T. 2018. ”Leaving at Lightspeed : the number of senior women in tech is decreasing”, *OECD Forum* (23 March 2018). Online. Available: https://www.oecd-forum.org/users/91062-tarah-wheeler/posts/31567-leaving-at-lightspeed-the-number-of-senior-women-in-tech-is-decreasing

World Summit on the Information Society (WSIS). 2003. *Declaration of Principles. Building the Information Society: A global challenge in the new Millenium*. Geneva, WSIS. Available at: http://www.itu.int/net/wsis/docs/geneva/official/dop.html

WSIS. 2005. *Tunis Agenda for the Information Society*. Tunis, WSIS. Available at: http://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html

ANEXO: COMPOSICIÓN DEL GRUPO DE TRABAJO AMPLIADO SOBRE ÉTICA E INTELIGENCIA ARTIFICIAL

1. **Prof. (Sr.) Peter-Paul VERBEEK (Coordinador Conjunto)**
   Profesor de Filosofía de la Tecnología en la Universidad de Twente (Países Bajos)
   Miembro de la COMEST (2016-2019)

2. **Prof. (Sra.) Marie-Hélène PARIZEAU (Coordinadora Conjunta)**
   Profesora, Facultad de Filosofía, Université Laval, Quebec (Canadá).
   Miembro de la COMEST (2012-2019).
   Presidenta (2016-2019) y Vicepresidenta (2014-2015) de la COMEST.

3. **Prof. (D.) Tomislav BRACANOVIĆ**
   Investigador Adjunto, Instituto de Filosofía, Zagreb (Croacia).
   Miembro de la COMEST (2014-2021).
   Ponente de la COMEST (2018-2019).

4. **Sr. John FINNEY**
   Profesor emérito de física, Departamento de Física y Astronomía, Londres (Reino Unido).
   Coordinador del Grupo de Trabajo sobre Ética Científica, Conferencia Pugwash sobre Ciencia y Asuntos Mundiales.
   Miembro por derecho propio de la COMEST.

5. **Sr. Javier JUAREZ MOJICA**
   Comisario, Consejo del Instituto Federal de Telecomunicaciones de México, Ciudad de México
   Miembro del Grupo de Expertos sobre IA (AIGO) de la OCDE.
   Miembro de la COMEST (2018-2021).

6. **Sr. Mark LATONERO**
   Jefe de investigación sobre datos y derechos humanos, Data & Society (Estados Unidos de América).

7. **Sra. Vidushi MARDA**
   Oficial de Programas Digitales en ARTICLE 19. (La Sra. Marda trabaja en la India)

8. **Prof. (Sra.) Hagit MESSER-YARON**
   Profesora de Ingeniería Eléctrica y ex Vicepresidenta de Investigación y Desarrollo de la Universidad de Tel Aviv, Tel Aviv (Israel).
   Miembro del Comité Ejecutivo, Iniciativa Global de la IEEE sobre la Ética de los Sistemas Inteligentes y Autónomos.
   Miembro de la COMEST (2016-2019).

9. **Dr. (Sr.) Luka OMLADIC**
   Profesor, Universidad de Liubliana, Liubliana (Eslovenia).
   Miembro de COMEST (2012-2019).

10. **Prof. (Sra.) Deborah OUGHTON**
    Profesora y Directora de Investigación, Centro de Radioactividad Ambiental, Universidad de Ciencias de la Vida de Noruega.
    Miembro de la COMEST (2014-2021).

11. Prof. (Sr.) Amedeo SANTOSUOSSO

Fundador y Director Científico del Centro Europeo de Derecho, Ciencia y Nuevas Tecnologías (ECLT), Universidad de Pavía, Pavía (Italia).
Presidente, Sala Primera, Tribunal de Apelación de Milán (Italia).
Miembro de la COMEST (2018-2021).

12. Prof. (Sr.) Abdoulaye SENE

Sociólogo ambiental, Coordinador de "Ética, Gobernanza, Responsabilidad Medioambiental y Social", Instituto de Ciencias Ambientales, Universidad Cheikh Anta Diop, Dakar (Senegal).
Miembro de la COMEST (2012-2019).
Vicepresidente de la COMEST (2016-2019).

13. Prof. (Sr.) John SHAWE-TAYLOR

Cátedra UNESCO de Inteligencia Artificial, University College of London y Presidente de la Knowledge 4 All Foundation (Reino Unido).

14. D. Davide STORTI

Especialista del programa, Sección de Aplicaciones de las TIC a la Educación, la Ciencia y la Cultura , Sector de Comunicación e Información, UNESCO.

15. Prof. (Sr.) Sang Wook YI

Profesor de Filosofía, Universidad de Hanyang, Seúl (República de Corea).
Miembro de la COMEST (2018-2021).

ANEXO II

Punto 42 Estudio preliminar sobre un posible instrumento normativo relativo a la ética de la inteligencia artificial (206 EX/42)

El Consejo Ejecutivo,

1. Habiendo examinado el documento 206 EX/42,

2. Reconociendo las preocupaciones por la creciente brecha digital y tecnológica entre los países, que podría verse exacerbada por la inteligencia artificial, y reiterando la importancia de atender las preocupaciones de los países en desarrollo respecto a la inteligencia artificial, en particular mediante la transferencia de tecnología de inteligencia artificial, el desarrollo de capacidades y la educación en materia de inteligencia artificial, la divulgación de datos y el acceso a los datos,

3. Reconociendo también que, si bien la inteligencia artificial tiene potencial para transformar el futuro de la humanidad para mejor y en favor del desarrollo sostenible, también existe una conciencia generalizada de los riesgos y desafíos que conlleva, especialmente por lo que respecta a la agravación de las desigualdades y brechas existentes, así como las implicaciones para los derechos humanos,

4. Tomando nota del estudio preliminar sobre los aspectos técnicos y jurídicos relativos a la conveniencia de disponer de un instrumento normativo sobre la ética de la inteligencia artificial,

5. Recomienda que la Conferencia General, en su 40ª reunión, en caso de que apruebe la propuesta de elaborar un instrumento normativo o un documento normativo sobre la inteligencia artificial, pida que se celebre un número suficiente de consultas intergubernamentales presenciales sobre el texto del instrumento normativo o el documento normativo mencionados;

6. Pide a la Directora General que le presente, en su 207ª reunión, un informe sobre la labor de otras organizaciones y convenciones internacionales en relación con diferentes aspectos de la inteligencia artificial;

7. Decide incluir este punto en el orden del día de la 40ª reunión de la Conferencia General;

8. Invita a la Directora General a que presente a la Conferencia General, en su 40ª reunión, el estudio preliminar sobre los aspectos técnicos y jurídicos relativos a la conveniencia de disponer de un instrumento normativo sobre la ética de la inteligencia artificial, que figura en el documento 206 EX/42, junto con las observaciones y decisiones pertinentes del Consejo Ejecutivo al respecto;

9. Recomienda también que la Conferencia General, en su 40ª reunión, invite a la Directora General a presentar un proyecto de nuevo instrumento normativo sobre la ética de la inteligencia artificial, en forma de recomendación, para que se someta al examen de la Conferencia General en su 41ª reunión.

ANEXO III

POSIBLE INSTRUMENTO NORMATIVO SOBRE LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL – HOJA DE RUTA

CALENDARIO	ACTIVIDAD
Diciembre de 2019	La Directora General envía una invitación a los Estados Miembros para que propongan expertos, antes de mediados de enero de 2020, para su participación en un grupo especial de expertos (categoría VI) encargado de preparar un proyecto de texto del instrumento normativo.
Febrero de 2020	• Se completa el proceso de selección y se envían cartas de invitación a los expertos para que formen parte del grupo especial de expertos. • Se establece el grupo y se anuncia su composición. • Se envía una carta a los miembros del grupo especial de expertos para indicar la tarea encomendada, la hoja de ruta y los documentos de referencia, con una invitación a la primera reunión.
Abril de 2020	Convocatoria de la primera reunión del grupo especial de expertos (cinco días, dos idiomas) para preparar el primer proyecto de texto del instrumento normativo.
Mayo-julio de 2020	• Se celebran al menos cuatro reuniones de consulta abierta y de múltiples interesados sobre el primer proyecto del instrumento normativo, organizadas en colaboración con todos los sectores del programa en los planos regional e internacional, centradas en temas y grupos de interesados específicos. En las consultas participarán los Estados Miembros, la sociedad civil, el sector privado y otros interesados pertinentes. Las consultas podrían requerir también la presencia de determinados miembros del grupo especial de expertos dependiendo de los temas, las regiones y los grupos de interesados. o Objetivo de las cuatro reuniones de consulta. o Financiación inicial para coorganizar consultas adicionales con asociados regionales, nacionales o locales. • Consulta pública en línea sobre el primer proyecto de texto del instrumento normativo.
Agosto de 2020	Se convoca la segunda reunión del grupo especial de expertos (cinco días, dos idiomas) para preparar el segundo proyecto de texto del instrumento normativo. La Secretaría, en consulta con el grupo especial de expertos, incluye en el segundo proyecto de texto del instrumento normativo las observaciones recibidas durante el proceso de consulta abierta y de múltiples interesados.

CALENDARIO	ACTIVIDAD
Septiembre de 2020	Se transmite a los Estados Miembros el segundo proyecto de texto del instrumento normativo para que formulen sus comentarios a más tardar el 31 de diciembre de 2020, junto con el informe preliminar preparado por la Secretaría (artículo 10.2 del Reglamento sobre las recomendaciones a los Estados Miembros y las convenciones internacionales previstas en el párrafo 4 del artículo IV de la Constitución).
Diciembre de 2020	La Secretaría recibe los comentarios y observaciones de los Estados Miembros sobre el segundo proyecto de texto del instrumento normativo.
Enero-marzo de 2021	La Secretaría prepara un informe final en el que se incluirán uno o varios proyectos de texto del instrumento normativo sobre la base de los comentarios y observaciones de los Estados Miembros (artículo 10.3).
Abril de 2021	• Se transmite a los Estados Miembros, por lo menos siete meses antes de la Conferencia General, el informe final que incluirá uno o varios proyectos de texto del instrumento normativo (artículo 10.3). • Se convoca la primera reunión del Comité Especial de expertos intergubernamentales (reunión de categoría II) encargado de preparar un proyecto definitivo de instrumento normativo (cinco días, seis idiomas) (artículo 10.4).
Junio de 2021	Se convoca la segunda y última reunión del Comité Especial de expertos intergubernamentales (reunión de categoría II) a fin de ultimar el proyecto (cinco días, seis idiomas) (artículo 10.4).
Julio de 2021	Se finaliza y transmite el proyecto definitivo a la Secretaría de los Órganos Rectores (GBS) para que prepare el documento para la Conferencia General.
Mediados de agosto de 2021	El Comité Especial de expertos intergubernamentales transmite a los Estados Miembros el proyecto definitivo del instrumento normativo (artículo 10.5).
Otoño de 2021	**41ª reunión de la Conferencia General:** examen y posible aprobación del instrumento normativo por la Conferencia General.