



La inteligencia artificial y los retos de la nueva ética. Bases éticas necesarias para la regulación jurídica de las aplicaciones de inteligencia artificial en Ecuador

Palabras clave: Inteligencia Artificial, IA, Ética de la Inteligencia Artificial

Nombre de la institución: Cátedra UNESCO de Ética y Sociedad en la Educación Superior Universidad Técnica Particular de Loja.

Proponentes: Carlos Eduardo García Torres

Ciudad: Loja

País: Ecuador

Sitio web de la institución: <http://unescoeticayeduacion.utpl.edu.ec/>

Sitio Web del programa, proyecto y experiencia (si se tiene):

Datos de la persona responsable del programa, proyecto y experiencia

- **Nombre y apellidos:** Andrea Luna Acosta
- **Cargo:** Coordinador Cátedra Unesco de Ética y Sociedad en la Educación Superior
- **Unidad/Facultad/Escuela:** Departamento de Derecho y Ciencias Sociales
- **Breve reseña biográfica (CV):**

Doctor en Derecho y Ciencias Sociales “Cum Laude” UNED (España)
Profesor de Derecho Romano en la Universidad Técnica Particular de Loja y en la Universidad Privada del Norte del Perú.
Profesor de postgrado en la Universidad Técnica Particular del Loja y en la Universidad del Azuay.
Becario Fulbright en el curso “American Politics and Political Thought”, University of Massachusetts, Amherst, 2007.
Becario de la OEA en el curso el Poder Legislativo en la Democracia y la Integración (OEA-UPD) 2001.
Ponente en eventos en la Universidad Económica de Praga, la Universidad de Santiago de Compostela y la Universidad de San Petersburgo.
Coordinador actual de la Cátedra UNESCO de Ética y Sociedad en la Educación Superior.
5 libros de autoría única y 5 libros en colaboración.
- **Correo electrónico:** cegarcia@utpl.edu.ec



- **Nombre y apellidos: María Cristina Castellanos González.**
- **Cargo: Ayudante de Investigación.**
- **Unidad/Facultad/Escuela: Facultad de Derecho**
- **Breve reseña biográfica (CV):** Pasante en la Secretaría del Agua (SENAGUA).
Pasante en Municipio de Loja.
Ayudante de Cátedra en la materia de Derecho Romano.
Ayudante de Cátedra en la materia de Derecho Civil IV, Obligaciones y Contratos.
Ponente en la II Jornada de Derecho de los Estudiantes organizado por la Directiva de la Titulación de Derecho.
Ayudante de Investigación en el **proyecto: La inteligencia artificial y los retos de la nueva ética** Grupo Etis: Ética y Sociedad
Ponente en el Congreso Internacional de Ciencias Jurídicas, Sociales y Políticas organizado por la Universidad Técnica Particular de Loja.
Asistente Legal en LEXXUS Soluciones Legales
- **Correo electrónico: mccastellanos@utpl.edu.ec**

1. Resumen ejecutivo

BASES TEÓRICAS. LA INTELIGENCIA ARTIFICIAL Y LOS RETOS DE LA NUEVA ÉTICA.

Las limitaciones de la doctrina utilitarista se hacen patentes bajo la perspectiva de la sociedad global y plural del siglo XXI y las diversas nociones de felicidad que ella trae consigo. De forma que lo que yo pienso que puede traer felicidad a todos depende en gran parte de mis propias concepciones de placer y de dolor (más allá de las consideraciones meramente físicas) siendo que estas concepciones son influenciadas por los antecedentes culturales de cada individuo. La Ética de futuro debe tratar de responder a los desafíos que los problemas sociales, ambientales y tecnológicos plantean a:

- A) La supervivencia de la especie;
- B) La preservación del equilibrio ecológico;
- C) El respeto a la esencia de la especie humana; y,
- D) El cuidado de todos los seres vivos.

Para lograr estos objetivos la Ética de futuro debe tener dos cursos de acción:

- A) Establecer los fundamentos axiológicos que hacen determinadas conductas preferibles frente a otras con miras a la supervivencia y el bienestar de las generaciones futuras.
- B) Determinar las conductas humanas que hoy son deseables de acuerdo a esos fundamentos. El proceder de esta Ética de Futuro deberá tomar en todo momento en cuenta los fundamentales valores del pasado que constituyen el núcleo básico de la cultura humana de forma que los valores que la historia y la filosofía humanas nos han legado deben constituir el sustrato básico de la nueva Ética.

Algunos de los fundamentos axiológicos ya se encuentran determinados en parte



por varios instrumentos consensuados por la mayoría de los países, entre ellos:

- A) Declaración Universal de los Derechos Humanos y todo el sistema de tratados que de ella deriva.
- B) Objetivos de Desarrollo Sostenible.
- C) Documentos éticos de valor trascendental como las Encíclicas “Laudato Si” y “Fratelli Tutti”

Pero estos fundamentos axiológicos que pueden cimentar la nueva Ética sufren de una debilidad que los acontecimientos recientes han revelado, un sistema social, político y económico fundamentalmente injusto que excluye a grandes grupos poblacionales de todos los beneficios que el futuro puede traer y que es el mayor reto a los objetivos que la Ética de futuro busca.

A partir de aquí conviene subrayar que la inteligencia artificial no es, en ningún caso, una copia exacta de la inteligencia humana. Definida en su nivel más básico puede decirse que la inteligencia artificial es la capacidad de un algoritmo para incorporar y utilizar nueva información a partir de árboles de decisión.

El Diccionario Oxford del inglés vivo dice: “La teoría y el desarrollo de sistemas computacionales capaces de realizar tareas que normalmente requieren inteligencia humana, como la percepción visual, reconocimiento de habla, toma de decisiones, y traducción entre lenguas.”

Un vagón sin control se aproxima a una división en las vías. Si continúa su curso matará a cinco trabajadores. Si el conductor se va por la otra vía el vagón matará a un solo trabajador. Si usted fuera el conductor del vagón ¿qué haría? ¿Qué haría un robot o una computadora?” (Allen et. Al., 2003: 12). Este dilema se ha vuelto relevante con la llegada de sistemas de trenes sin conductor que aparecen en aeropuertos y metros. (Allen et. Al., 2003: 12)

“Ya tenemos robots semiautónomos y agentes de software que violan rutinariamente los estándares éticos. Por ejemplo, un motor de búsqueda que puede



recolectar información que, legalmente, se considera privada. (Allen et. Al., 2003: 12).

Dos acontecimientos recientes refuerzan estas preocupaciones. En primer lugar, la intervención decisiva de tecnologías de inteligencia artificial en las elecciones norteamericanas de hace cuatro años y, en segundo lugar, las multas de que ha sido objeto Google por sus algoritmos de búsqueda que extraen información delicada.

“Las nuevas tecnologías en los campos de IA, genómica y nanotecnología se combinarán en una miríada de vías inimaginables para ofrecer promesas en todo, desde incrementar la productividad hasta curar enfermedades. Por tanto, necesitamos integrar agentes morales artificiales dentro de estas nuevas tecnologías para manejar su complejidad. Estos AMA deben ser capaces de tomar decisiones que honren la privacidad, sostengan estándares éticos compartidos, protejan los derechos civiles y la libertad individual” (Allen et. Al., 2003: 13)

Bostrom y Yudkowski han escrito algunas recomendaciones para el Manual Oxford de Inteligencia Artificial (2011).

Ellos recomiendan:

1. Algoritmos transparentes a la inspección. Es decir que se pueda saber los modos como toman sus decisiones.
2. Los algoritmos deben resultar predecibles por quienes sufren sus consecuencias, tanto como son predecibles los sistemas jurídicos.
3. Los algoritmos de IA deben ser robustos en contra de la manipulación.
4. Debe existir a fin de cuentas un ser humano responsable por las decisiones que tome la máquina.

“Construir una IA que actúe de forma segura mientras actúa en varios dominios con

muchas consecuencias (...) se puede especificar la buena conducta en estos términos: “Sea tal X que las consecuencias de X no sean dañinas para los humanos” (Bostrom y Yudkowski, 2011: 5).

“Responsabilidad, transparencia, susceptibilidad de ser auditado, incorruptibilidad, posibilidad de predicción y una tendencia a no hacer que víctimas inocentes griten con frustración indefensa: todos estos criterios se aplican a humanos desempeñando funciones sociales; todos estos criterios deben ser considerados en un algoritmo que pretenda reemplazar la capacidad de juicio humana”. (Bostrom y Yudkowski, 2011: 3).

HITOS

A partir de estas bases teóricas el proyecto, partiendo de una fase general ha presentado diversos resultados en varios eventos académicos, siendo el último el Congreso internacional de Ciencias Sociales, Jurídicas y Políticas.

2. Fundamento y relación con las encíclicas Laudato Si’ o Fratelli Tutti

El presente proyecto relacionado con la ética de la Inteligencia Artificial plantea lineamientos básicos para la regulación jurídica de la IA; esto es:

1. **Respeto de la autonomía humana:** Las personas que interactúen con sistemas de IA deben poder mantener una autonomía plena y efectiva sobre sí mismas y ser capaces de participar en el proceso democrático. Los sistemas de IA no deberían subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos de manera injustificada.
2. **Prevención del daño:** Los sistemas de IA no deberían provocar daños (o agravar los existentes), ni perjudicar de cualquier otro modo a los seres humanos. Es decir, deben ser robustos desde el punto de vista técnico



3. **Equidad:** Distribución justa e igualitaria de los beneficios y costes; y, igualdad de oportunidades
4. **Explicabilidad:** Los procesos deben ser transparentes (Unión Europea, 2019)

Sentando la base de los prenombrados principios básicos, se desarrollaría normativa interna en pro de salvaguardar la privacidad, transparencia y demás derechos humanos. Pues ya lo dijo el Beato papa Pablo VI y se cita en el párrafo 4, el Sumo Pontífice también habló a la FAO sobre la posibilidad de una «catástrofe ecológica bajo el efecto de la explosión de la civilización industrial», subrayando la «urgencia y la necesidad de un cambio radical en el comportamiento de la humanidad», porque «los progresos científicos más extraordinarios, las proezas técnicas más sorprendentes, el crecimiento económico más prodigioso, si no van acompañados por un auténtico progreso social y moral, se vuelven en definitiva contra el hombre».

La sociedad debe reconocerse como parte del todo, en el que se debe mantener un equilibrio social, económico, ambiental y tecnológico, pues si se desestabiliza una de estas bases, sería cuestión de tiempo para que el resto dejara de funcionar. Como lo dijo San Juan Pablo II y se citan en el párrafo 102 «la ciencia y la tecnología son un maravilloso producto de la creatividad humana donada por Dios»; en tal sentido, se debe desarrollar con responsabilidad desde el momento en que se decide el tipo de algoritmo de IA que se va a utilizar (Redes Bayesianas, redes neuronales u otras), los datos que se utilizaran y el fin último de los mismos; y el campo en el que se utilizaran dichos agentes.

Como se cita en el párrafo 105 de la carta encíclica, cada época tiende a desarrollar una escasa autoconciencia de sus propios límites. Por eso es posible que hoy la humanidad no advierta la seriedad de los desafíos que se presentan, y «la posibilidad de que el hombre utilice mal el poder crece constantemente» cuando no está «sometido a norma alguna reguladora de la libertad, sino únicamente a los supuestos imperativos de la utilidad y de la seguridad». El ser humano no es plenamente autónomo. Su libertad se enferma cuando se entrega a las fuerzas ciegas del inconsciente, de las necesidades inmediatas, del egoísmo, de la violencia. En ese sentido, está desnudo y expuesto frente a su propio poder, que sigue creciendo, sin tener los elementos para controlarlo. Puede disponer de mecanismos superficiales, pero podemos sostener que le falta una ética sólida, una cultura y una



espiritualidad que realmente lo limiten y lo contengan en una lúcida abnegación.

Evidentemente, las leyes de la robótica planteadas por Isaac Asimov en 1942 resultan una guía frente a el plano indeterminable de la IA:

1. Un robot no puede dañar a un ser humano o por inacción permitir que un ser humano sufra daños.

Dado el caso de un agente de inteligencia artificial, el agente de inteligencia artificial no puede dañar a un ser humano; o por inacción permitir que un ser humano sufra daños.

2. Un robot debe obedecer las órdenes dadas por el ser humano; excepto cuando entren en conflicto con la primera ley.

En el caso de la inteligencia artificial, un agente debe obedecer las órdenes dadas por el ser humano excepto cuando le genere un daño.

3. un robot debe proteger su propia existencia siempre que dicha protección no entre en conflicto con la primera o segunda ley.

El agente de inteligencia artificial debe proteger su existencia siempre y cuando no dañe a un ser humano ni desobedezca las órdenes dadas por él.

En relación con la carta encíclica **Fratelli Tutti'** en su párrafo 42 se refiere que paradójicamente, mientras se desarrollan actitudes cerradas e intolerantes que nos clausuran ante los otros, se acortan o desaparecen las distancias hasta el punto de que deja de existir el derecho a la intimidad. Todo se convierte en una especie de espectáculo que puede ser espiado, vigilado, y la vida se expone a un control constante. En la comunicación digital se quiere mostrar todo y cada individuo se convierte en objeto de miradas que hurgan, desnudan y divulgan, frecuentemente de manera anónima. El respeto al otro se hace pedazos y, de esa manera, al mismo tiempo que lo desplazo, lo ignoro y lo mantengo lejos, sin poder alguno puedo invadir su vida hasta el extremo

El presente proyecto aporta una visión preponderante en relación al derecho a la privacidad y transparencia del uso de datos en el desarrollo de la inteligencia artificial, desde un uso responsable y un diseño responsable.

En relación a la privacidad, es necesario mencionar que la recopilación y almacenamiento



de datos es digital, por lo que una posibilidad sería la privacidad diferencial, que encripta la salida de las consultas, dejando a discreción la posibilidad de que en un futuro se identifique a la entidad responsable en caso de filtración de datos y se materialice la responsabilidad sea civil o penal. Por otra parte, dada la opacidad y el sesgo, se plantea la supervisión y decisión humana como primer punto, pues un sistema de inteligencia artificial nunca podrá reemplazar la responsabilidad y la rendición de cuentas final por parte de un ser humano.

3. Contexto

La Cátedra UNESCO de Ética y Sociedad en la Educación Superior es Un espacio de investigación y diálogo que tiene como núcleo la reflexión sobre la Ética en sí misma y sobre su papel en la sociedad del siglo XXI, así como la difusión de los frutos de su labor en todos los ámbitos educativos. La Cátedra, desde un paradigma de tolerancia y amplitud de pensamiento acoge todas las visiones de la Ética en tanto conocimiento abarcador del pensamiento axiológico de la humanidad. Para lograr su objetivo busca integrar en su seno a investigadores, gestores, instituciones educativas, estamentos del Estado y organismos no gubernamentales que se interesen en la reflexión ética y su aplicación práctica a los problemas morales, sociales, políticos, biomédicos y ambientales que enfrenta la sociedad en el siglo XXI; y, promover acciones que fortalezcan la presencia de la Ética como elemento a considerar en todos los procesos sociales. Entre estos problemas presenta un muy importante reto la vigencia y desarrollo de la Inteligencia Artificial en todos los ámbitos de la vida humana, así como los numerosos dilemas éticos que todos los días plantea. Así pues, el objetivo inicial del proyecto fue: Analizar las relaciones entre los desarrollos de la Inteligencia Artificial y las propuestas de la nueva ética. Estas propuestas de la nueva ética fueron presentadas por primera vez en un Webinar realizado por IESALC-UNESCO:

<https://www.youtube.com/watch?v=X7-cCp0yiw&t=481s>

Siendo fundamentalmente una experiencia investigativa, sus objetivos son esencialmente teóricos y no es posible determinar los potenciales beneficiarios de sus desarrollos. La elección de este campo de investigación se justifica por el desarrollo de las aplicaciones de inteligencia artificial, así como su vigencia en muchos de los campos de la vida humana.



4. Descripción, planificación, objetivos y desarrollo

El objetivo general planteado en el presente proyecto fue analizar las relaciones entre los desarrollos de la Inteligencia Artificial y las propuestas de la nueva ética. Los objetivos específicos fueron dos: Perfilar algunas de las prescripciones éticas necesarias para la regulación de la Inteligencia Artificial en el ámbito nacional; y, proponer las consecuencias jurídicas de estas prescripciones éticas.

El proyecto de investigación se desarrolló desde el mes de abril de 2020 hasta febrero de 2021; es decir, durante dichos meses se recopiló información de distintos ensayos publicados en revistas indexadas (Springer, UNESCO, SCOPUS, entre otras), páginas web de organizaciones internacionales (UNESCO, UE, entre otras), libros, enciclopedias y ponencias de expertos en el tema (Bostrom y otros) en lo pertinente a la ética de la inteligencia artificial.

Luego, se realizó un análisis de la información recabada en la que se diferenció lo referente a inteligencia artificial fuerte, bajo un plan hipotético en que una máquina puede tener mente, conciencia y estados mentales; e , inteligencia artificial débil, como un sistema de procesamiento de la información que incorpora modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y adopción de decisiones en entornos reales y virtuales. (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2020)

De acuerdo a Bostrom y Yudkowsky (2011) refieren que los actuales algoritmos de inteligencia artificial se resumen en dos: rendimiento humano equivalente y rendimiento humano superior; sin embargo, ambas competencias se programan en un único dominio restringido y carecería de la generalidad como característica y por tanto, la inteligencia humana sería la única generalmente aplicable. Los problemas de seguridad se pueden prever de aristas como el propósito de la maquina representado en la mente del diseñador y la falta de capacidad de prever eventos específicos que pueden suscitarse en un futuro.



Son enfáticos en subrayar que la primera inteligencia general es la de los humanos y algunas consecuencias evidentes frente al desarrollo de la sociedad ha sido el cambio climático y otros. Por lo que, el construir inteligencia artificial bajo el criterio de superethics se basaría en tres puntos: a) Poseer una propia experiencia moral; b) Reconocer al menos parte de la ética de la actualidad; y, c) Progreso moral.

Bringjord et. al, (2020) afirman que se puede percibir el continuo del agente sofisticado de la siguiente manera:

Vista basada en agentes: Se concibe como agente al que recibe la percepción del entorno y realización.

1. Agente inteligente con capacidad de pensar con anticipación en base al entorno.
2. Agente de conocimiento declarativo y capaz de razonar.
3. Agentes que manejan la incertidumbre razonando de manera probabilística.
4. Agentes capaces de aprender.
5. Agentes que se comuniquen.

Los agentes adquieren la inteligencia por lectura (texto e idioma); o por codificación insertada manualmente por el ser humano. Dado el caso de algoritmos de aprendizaje automático que se utilizan para analizar grandes volúmenes de datos generados a partir de aceleradores de partículas que se basan en análisis de características y aprendizaje de funciones; las máquinas deberían tener un marco formal utilizado para razonar un código moral mediante tres formas: fuente humana, fuente en que la propia máquina infiere el código y fuente en que la máquina recibe el código directamente (codificación) o indirectamente (lectura).

En relación a la opacidad y sesgos de los sistemas de inteligencia artificial, Müller (2020) refiere que en un soporte automatizado de decisiones y análisis predictivo, como un sistema de aprendizaje automático, ni la persona afectada o el experto (creador) puede identificar el patrón sobre el cual la máquina tuvo la decisión o los datos exacerbados; es decir, se evidencia la opacidad. Por otra parte, al estar un juicio de valor influenciado por una característica irrelevante al tema (preconcepción discriminatoria) se genera ciertos sesgos cognitivos, de confirmación (interpretar información como confirmación de lo que creo e impide el desempeño de un juicio racional); y sesgo estadístico, donde cualquier conjunto de datos dado sólo será imparcial para un solo tipo de problema.



Por ejemplo, el aprendizaje automático reconocería el sesgo y automatizaría el sesgo histórico discriminatorio de mujeres embarazadas porque la compañía tenía un historial de discriminación.

En el último punto de esta investigación se procedió a emitir recomendaciones y prescripciones éticas que se relacionan con las planteadas por la Unión Europea en su Libro Blanco y las Recomendaciones planteadas por UNESCO respecto a la Inteligencia Artificial, esto es:

IA lícita en la que todos los derechos y obligaciones legales que se aplican a los procesos y actividades implicados en el desarrollo, despliegue y utilización de la IA conservan su carácter obligatorio y han de ser debidamente observados; IA ética donde no basta que los sistemas de IA cumplan con la ley; sino que, deben ser coherentes con las normas éticas y adecuadas para abordar sobre determinadas cuestiones; y IA robusta, donde las personas y la sociedad también deben poder confiar en que los sistemas de IA no provocarán un daño involuntario. Dichos sistemas deben funcionar de manera segura y fiable; se deberían prever medidas de protección para evitar cualquier efecto adverso imprevisto. (UE, 2019, p.8)

En este sentido, se debe cimentar un ecosistema de confianza en base a cuatro principios (Wilkinson et. al, 2016):

1. Localizable:

- Los metadatos se les asigna un identificador persistente y único a nivel mundial.
- Los datos se describen con metadatos enriquecidos.
- Los metadatos incluyen clara y explícitamente el identificador de los datos que describen.
- Los metadatos se registran o indexan en un recurso de búsqueda

. Accesible:

- Los metadatos son recuperables por su identificador usando un protocolo de comunicaciones estandarizado.
- El protocolo es abierto, gratuito y universalmente implementable.
- El protocolo permite un procedimiento de autenticación y autorización, cuando sea



necesario.

- Los metadatos son accesibles, incluso cuando los datos ya no están disponibles

3. Interoperable:

- Los metadatos utilizan un lenguaje formal, accesible, compartido y de amplia aplicación para la representación del conocimiento.
- Los metadatos usan vocabularios que siguen principios justos.
- Los metadatos incluyen referencias calificadas a otros meta datos

4. Reutilizable:

- Los metadatos se describen detalladamente con una pluralidad de atributos precisos y relevantes.
- Los metadatos se publican con una licencia de uso de datos clara y accesible.
- Los metadatos están asociados con la procedencia detallada.
- Los metadatos cumplen con los estándares comunitarios relevantes para el dominio.

Los prenombrados principios se complementan con una serie de principios genéricos para el desarrollo, implantación y uso de la inteligencia artificial que garantice los objetivos de desarrollo sostenible, esto es:

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 TIC

En particular, al Grupo de Trabajo le gustaría señalar algunas cuestiones éticas fundamentales en relación con el enfoque específico de la (UNESCO, 2019):

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.

5. Resultados

Los resultados del proyecto, por tratarse principalmente de un proyecto de investigación puede medirse en las ponencias presentadas entre las cuales merecen destacarse dos:

1. Webinar IESALC-UNESCO

https://www.youtube.com/watch?v=X7-_cCp0yiw&t=481s

2. Ponencia en el Congreso Internacional de, Ciencias Jurídicas Sociales y Políticas

<https://www.facebook.com/utpleventos/videos/1048010022343055>

3. Adicionalmente se ha publicado un artículo que contiene algunas de las ideas que se están desarrollando:

<https://studiahumanitatis.eu/ojs/index.php/analysis/article/view/2020-garciatorres>

6. Aportes al nuevo modelo cultural propuesto por el Papa Francisco

El presente proyecto de investigación se relaciona con el modelo cultural propuesto por el Papa Francisco en dos sentidos:

En relación a la carta encíclica Fratelli Tutti del Santo Padre Francisco sobre la fraternidad y la amistad social, el presente proyecto presenta cuatro principios básicos:

1. **Respeto de la autonomía humana:** Las personas que interactúen con sistemas de IA deben poder mantener una autonomía plena y efectiva sobre sí mismas y ser capaces de participar en el proceso democrático. Los sistemas de IA no deberían



subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos de manera injustificada.

2. **Prevención del daño:** Los sistemas de IA no deberían provocar daños (o agravar los existentes), ni perjudicar de cualquier otro modo a los seres humanos. Es decir, deben ser robustos desde el punto de vista técnico.
3. **Equidad:** Distribución justa e igualitaria de los beneficios y costes; y, igualdad de oportunidades.
4. **Explicabilidad:** Los procesos deben ser transparentes (UE, 2019)

Para el correcto desarrollo de los sistemas de Inteligencia Artificial se deben cumplir ciertos requisitos que garantizan una IA Fiable de acuerdo a la Unión Europea (2019):

1. **Acción y supervisión humana:** Vinculada al principio del respeto de la autonomía humana.

Los usuarios deberían ser capaces de tomar decisiones autónomas con conocimiento de causa en relación con los sistemas de IA; y, el control humano se refiere a la capacidad de que intervengan seres humanos durante el ciclo de diseño del sistema y en el seguimiento de su funcionamiento.

2. **Solidez técnica y seguridad:** Vinculada al principio de prevención del daño.

Resistencia a los ataques y seguridad, plan de repliegue y seguridad general, precisión; y, fiabilidad y reproducibilidad.

3. **Gestión de la privacidad y de los datos:** Vinculada al principio de prevención del daño.

Protección de la intimidad y de los datos, calidad e integridad de los datos; y, acceso a los datos

4. **Transparencia:** Vinculada al principio de explicabilidad.

Trazabilidad, explicabilidad; y, comunicación.

5. **Diversidad, no discriminación y equidad:** Vinculada al principio de equidad.

Necesidad de evitar sesgos injustos, accesibilidad y diseño universal; y, participación de las partes interesadas.



6. Bienestar ambiental y social: Vinculada a los principios de equidad y prevención del daño.

IA sostenible y respetuosa del ambiente, impacto social; y, sociedad y democracia.

7. Rendición de cuentas: Vinculada al principio de equidad.

Auditabilidad, minimización de efectos negativos y notificación de estos, búsqueda de equilibrios; y, compensaciones.

En este sentido, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2020) ha emitido ciertos valores y principios:

Valores

1. Respeto, protección y promoción de la dignidad humana, los derechos humanos y las libertades fundamentales
2. Prosperidad del medio ambiente y los ecosistemas
3. Diversidad e inclusión
4. Vivir en armonía y paz

Principios

1. Proporcionalidad e inocuidad

- a) el método de IA elegido debería ser conveniente y proporcional para lograr un objetivo legítimo determinado;
- b) el método de IA elegido no debería repercutir negativamente en los valores fundamentales enunciados en el presente documento;
- c) el método de IA debería ser adecuado al contexto y basarse en fundamentos científicos rigurosos.

2. Seguridad y protección: Los daños no deseados (riesgos de seguridad) y las vulnerabilidades a los ataques (riesgos de protección) deberían evitarse a lo largo del ciclo de vida de los sistemas de IA

3. Equidad y no discriminación: Justicia social, respetando la equidad (Beneficios de las tecnologías de la IA); y, los actores de la IA deberían hacer todo

lo posible por reducir al mínimo los sesgos sociotécnicos inadecuados basados en prejuicios.

4. Sostenibilidad: Debería llevarse a cabo una evaluación continua de los efectos sociales, culturales, económicos y ambientales de las tecnologías de la IA con pleno conocimiento de las repercusiones de dichas tecnologías en la sostenibilidad

5. Privacidad: Derecho esencial para la protección de la dignidad, la autonomía y la capacidad de actuar de los seres humanos, debe ser respetada, protegida y promovida a lo largo del ciclo de vida de los sistemas de IA, tanto a nivel personal como colectivo. Es fundamental que los datos de la IA se recopilen, utilicen, compartan, archiven y supriman de forma coherente con los valores y principios enunciados en la presente Recomendación.

6. Supervisión y decisión humana: La supervisión humana se refiere, por tanto, no solo a la supervisión humana individual, sino también a la supervisión pública, según corresponda. La decisión de ceder el control en contextos limitados seguirá recayendo en los seres humanos, ya que estos pueden recurrir a los sistemas de IA en la toma de decisiones y en la ejecución de tareas, pero un sistema de IA nunca podrá reemplazar la responsabilidad y la rendición de cuentas final por parte de un ser humano.

7. Transparencia y explicabilidad: La transparencia suele ser una condición previa fundamental para garantizar el respeto, la protección y la promoción de los derechos humanos fundamentales y los principios éticos. La transparencia es necesaria para que la legislación nacional e internacional pertinente en materia de responsabilidad funcione eficazmente. a explicabilidad de los sistemas de IA también se refiere a la inteligibilidad de la entrada, salida y comportamiento de cada componente algorítmico y la forma en que contribuye a los resultados de los sistemas.

8. Responsabilidad y rendición de cuentas: La responsabilidad ética y la obligación de rendir cuentas de las decisiones y las acciones basadas de alguna manera en un sistema de IA siempre deberían ser atribuibles, en última instancia, a los actores de la IA. Contar con dispositivos tanto técnicos como institucionales deberían garantizar la Auditabilidad y la trazabilidad (del funcionamiento) de los sistemas de IA, en particular para intentar solucionar cualquier conflicto con los



derechos humanos y las amenazas al bienestar del medio ambiente y los ecosistemas.

9. Sensibilización y alfabetización: El valor de los datos deberían promoverse mediante una educación abierta y accesible. El aprendizaje sobre el impacto de los sistemas de IA debería incluir el aprendizaje sobre los derechos humanos, a través de ellos y para ellos,

10. Gobernanza y colaboración adaptativas y de múltiples partes interesadas: Regulación efectiva de los datos sobre la base del respeto del derecho a la privacidad y otros derechos humanos

Nos obstante, tanto los principios y directrices básicos emitidos por la UE y UNESCO, son precisos en referir que se pretende garantizar el desarrollo de agentes de inteligencia artificial con criterios éticos, pues en caso de no hacerlo, los “atajos tomados hoy”, en un futuro se verían reflejados en problemas éticos que generarían consecuencias a toda la humanidad.

De igual manera, el proyecto se relaciona con la carta encíclica Laudato Si' del Santo Padre Francisco sobre el cuidado de la casa común en referencia a dejar la perspectiva del antropocentrismo y contribuir a una ecología integral en la que se desarrolle la tecnología ética y con perspectiva social.