

Modelos de Inteligencia Artificial

*Conforme a contenidos del «Curso de Especialización
en Inteligencia Artificial y Big Data»*



**Modelos de
Inteligencia_Artificial**

Universidad de Castilla-La Mancha

Escuela Superior de Informática
Ciudad Real

8

Capítulo

Aspectos éticos y jurídicos de la Inteligencia Artificial

José Angel Olivas Varela

En este capítulo se presentan algunos aspectos éticos y jurídicos de la Inteligencia Artificial. La inteligencia artificial se está desarrollando rápidamente. Cambiará nuestras vidas mejorando la asistencia sanitaria (por ejemplo, haciendo más precisos los diagnósticos, permitiendo una mejor prevención de las enfermedades), aumentando la eficiencia de la agricultura, contribuyendo a la mitigación del cambio climático y a la adaptación al mismo, mejorando la eficiencia de los sistemas de producción mediante el mantenimiento predictivo, aumentando la seguridad de los europeos, y de muchas otras maneras que sólo podemos empezar a imaginar. Al mismo tiempo, la Inteligencia Artificial (IA) conlleva una serie de riesgos potenciales, como la opacidad en la toma de decisiones, la discriminación por razón de género u otros tipos de discriminación, la intrusión en nuestra vida privada o su utilización con fines delictivos.

La Comisión Europea, especialmente en su ‘White Paper on Artificial Intelligence: a European approach to excellence and trust’¹ se compromete a permitir el avance científico, a preservar el liderazgo tecnológico de la UE y a garantizar que las nuevas tecnologías estén al servicio de todos los europeos, mejorando sus vidas y respetando sus derechos. También es recomendable consultar el documento sobre ‘Nuevas normas sobre la inteligencia artificial: preguntas y respuestas’² y el de ‘Communication on Fostering a European approach to Artificial Intelligence’³.

¹https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

²<https://ec.europa.eu/commission/presscorner/detail/es/QANDA211683>

³<https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence>

8.1. La ética de la inteligencia artificial.

Se considera la ética de la inteligencia artificial a una rama de la ética que analiza y evalúa los elementos morales que se derivan del desarrollo y aplicación de estas tecnologías en la sociedad. Desde hace siglos la idea de seres conscientes y pensantes ha sido utilizada por la filosofía para analizar características propias de los seres humanos y si estas son trasladables a otros objetos. Así, se estudiaron conceptos como los de consciencia y su importancia en esta rama, teniendo en cuenta que a través de nuestra consciencia individual tomamos conocimiento de nuestros principios morales más arraigados, nos motivamos a actuar conforme a ellos y evaluamos nuestro carácter, nuestro comportamiento y a nosotros mismos conforme a estos principios.

En Los últimos años se ha derivado una nueva aproximación a la ética de la inteligencia artificial, basada en la ética de los datos (data Ethics). Esta ética se centra en el uso y analítica de los datos y los distintos sistemas e innovaciones que interactúan con estos recursos. Dada la importancia que la analítica de datos y el auge que el big data ha tenido en los últimos años, esta es una de las éticas que mayor predominancia ha tenido y muchas de las propuestas éticas y de principios se derivan de este estudio.

Esta ética se basa en los fundamentos de la ética de la computación y la informática, pero, al mismo tiempo, enriquece la perspectiva considerada hasta el momento en este campo de investigación, al cambiar el nivel de abstracción de las preguntas éticas, de estar centrado en la información a estar centrado en los datos, especialmente aquellos que son utilizados como base para el desarrollo e implementación de sistemas de inteligencia artificial.

De esta forma, se propone una definición de la ética de los datos que a su vez divide esta disciplina en tres criterios de análisis: una nueva rama de la ética que estudia y evalúa los problemas morales relacionados con los datos (incluyendo su generación, registro, adaptación, tratamiento, divulgación, diseminación y uso), algoritmos (incluyendo IA, agentes artificiales, machine learning y robots) y prácticas correspondientes (incluyendo innovación responsable, programación, hacking y códigos profesiones), con el fin de formular y apoyar soluciones moralmente buenas, por ejemplo códigos adecuados o valores correctos.

8.1.1. Retos y efectos no deseados.

Como ya se ha señalado, la implementación de la inteligencia artificial ya ha supuesto una serie de retos éticos, al igual que ha generado inquietudes alrededor del impacto que puede tener en algunos casos.

Tal y como lo ha señalado el gobierno francés, existen los siguientes retos específicos que se deben abordar:

1. Posibles amenazas a la libertad de voluntad y responsabilidad;
2. Sesgos, discriminación y exclusión;

3. Perfilamiento algorítmico: personalización versus beneficios colectivos;
4. Buscar un nuevo balance entre bases de datos masivas mientras se aumenta la IA;
5. Calidad, cantidad y relevancia: el reto de los datos adaptados para IA, y
6. La identidad humana frente al reto de la inteligencia artificial.

Según la Web Foundation⁴, los principales efectos negativos que puede tener un sistema de inteligencia artificial se dividen en daños y discriminación. En cuanto a los daños algorítmicos, la Web Foundation (2017) considera la definición de daños está vinculada a los valores de cada sociedad. Al definir lo que un algoritmo no debe hacer (daño), se establecen límites claros para lo que debe ser una función de optimización de un algoritmo (objetivos amplios). Asegurarse de que los algoritmos sean compatibles con la diversidad de valores que existe alrededor del mundo es un reto. ¿Quién debe definir y determinar si los algoritmos han producido un daño? ¿En qué casos debemos promover que quienes hayan podido ser afectados por un algoritmo se integren al proceso de diseño?. La Web Foundation toma la definición de daño de la esfera legal, que define daño como retrocesos que además se considera que están mal. Los riesgos antes mencionados se tratan de abordar desde el concepto de legitimidad que han propuesto distintas entidades internacionales.

Por otra parte, la discriminación puede ocurrir de dos formas. Dos personas pueden ser iguales en aspectos relevantes, pero ser tratadas de forma diferente, o las diferencias relevantes entre ellas no son reconocidas o tenidas en cuenta y las dos personas son tratadas igual. En el segundo escenario, al no tener en cuenta estos detalles relevantes, el resultado es injusto y en consecuencia. De esta manera, una persona puede esperar razonablemente un resultado que es injustamente impedido por un algoritmo, constituyendo un daño. Los países con altos y bajos ingresos enfrentan las mismas categorías de daños y amenazas provenientes de la toma de decisiones algorítmica.

Ahora bien, el impacto de estos daños puede ser muy diferente, dependiendo de las protecciones legales existentes y los mecanismos de responsabilidad demostrada implementados, especialmente para los grupos marginados. En algunos países, la discriminación algorítmica y las predicciones inexactas pueden resultar en publicidad no deseada o en otros inconvenientes en las experiencias de los consumidores, pero para los grupos marginados en contextos frágiles, se argumenta que la discriminación algorítmica puede llevar a agresiones sin control e incluso a exclusiones fatales de servicios y recursos públicos.

Como se puede observar, la mayoría de las preocupaciones está alrededor de la discriminación que puedan generar estos sistemas, como pueden profundizar las desigualdades y la posibilidad de tener sistemas que tomen decisiones automáticamente, sin que exista control y orientados por los prejuicios y la discriminación que ha sido establecida en su diseño o en los datos que usan.

⁴<https://webfoundation.org/>

Un ejemplo han sido los sistemas de reconocimiento facial que se valen de esta tecnología. Hay una amplia evidencia de que esta tecnología causa daños graves, con frecuencia a personas de color y gente pobre. En consecuencia, debe haber una suspensión en todos los usos de reconocimiento facial en los dominios públicos y sociales sensibles, donde este tipo de reconocimiento tiene riesgos y consecuencias que no se pueden remediar retroactivamente. Los legisladores deben suplementar esta suspensión con (1) requisitos de transparencia que permita a investigadores, hacedores de política pública y comunidades asesorar y entender la mejor forma de restringir y regular el reconocimiento facial, y (2) protecciones que brinden a las comunidades sobre las que se utiliza esta tecnología con el poder de hacer sus propias evaluaciones y rechazar la implementación.

Estos riesgos no solo pueden terminar afectando a la sociedad general, sino también a los derechos humanos fundamentales y propios de cada individuo. A través de una investigación de casos de estudio [1], miembros del Berkman Klein Center de la Universidad de Harvard determinaron que los sistemas de inteligencia artificial que habían sido implementados en los Estados Unidos en distintos sectores habían tenido un impacto en distintos derechos humanos, como la libertad de expresión, el derecho de asociación e igualdad, etcétera.

A partir de su investigación se lograron determinar los siguientes impactos específicos de esta tecnología en varios de estos derechos:

- Sistemas de Inteligencia Artificial en el sistema penal para determinar el nivel de riesgo de un individuo y el grado de reincidencia: conforme al estudio estos tienen un impacto negativo en el derecho a una audiencia pública justa, el derecho a ser considerado inocente hasta que se demuestre lo contrario y la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia.
- Sistemas de Inteligencia Artificial en el sistema financiero para determinar el riesgo crediticio de un individuo: conforme al estudio estos tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia, en la libertad de opinión e información, en el derecho a la libertad de asociación y en el derecho al trabajo deseado y de sindicación.

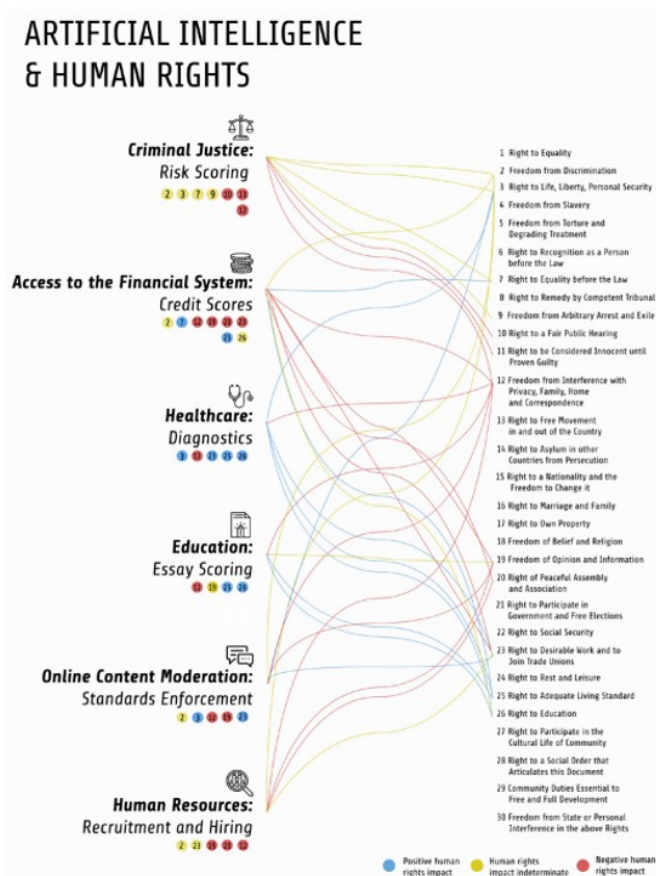


Figura 8.1: Arquitectura de un Mapa de Kohonen.

- Sistemas de Inteligencia Artificial en el sistema de salud para realizar diagnósticos médicos: tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia.
- Sistemas de Inteligencia Artificial en el sistema educativo para la calificación de ensayos: tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia.
- Sistemas de Inteligencia Artificial para moderar contenido online e implementar los estándares de participación propuestos: tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia y en la libertad de opinión e información.

- Sistemas de Inteligencia Artificial en los departamentos de recursos humanos para el reclutamiento y selección de candidatos: conforme al estudio estos tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia, en la libertad de opinión e información y en el derecho a la libertad de asociación.

8.1.2. Los principios predominantes a nivel internacional.

Los principios que han predominado a nivel internacional son los de privacidad, responsabilidad, seguridad, transparencia y explicación, justicia y no discriminación, control y supervisión humana de la tecnología, promoción de los valores humanos y los señalados dentro del marco internacional de derechos humanos. Se debe tener en cuenta:

- La privacidad se constituye como un principio predominante y que llama la atención de todos los sectores.
- La seguridad se ha constituido en una preocupación transversal considerando los distintos ataques que se han producido a esta tecnología.
- Muchas de las preocupaciones están alrededor del uso de los datos.
- Los gobiernos han privilegiado la transparencia.
- El sector privado busca definir responsabilidades en el uso de estos sistemas.
- A la sociedad civil le preocupa la protección de derechos fundamentales y la discriminación.
- Preocupación frente al uso que los gobiernos puedan dar y valerse de esta tecnología para generar políticas públicas y prestar servicios públicos y que pueden llegar a ser discriminatorias frente a un grupo específico.

El análisis mundial de principios aplicables a la inteligencia artificial realizado por el Berkman Klein Center de la Universidad de Harvard [2] permite identificar cuáles son los intereses y preocupaciones que los principales sectores de la sociedad tienen frente a la selección de dichos principios y sus implicaciones. Los reportes, documentos oficiales e investigaciones que los distintos actores han generado frente al tema permiten afirmar esto.

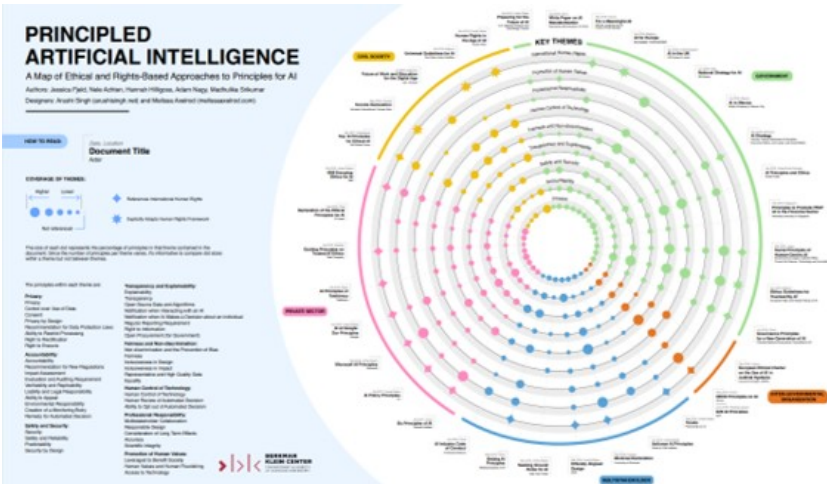


Figura 8.2: Un mapa de las propuestas basadas en ética y derechos [2].

PROPUESTA.

Por lo tanto, en base a todo lo anterior se puede hacer una propuesta general de los principios éticos que se deberían tener en cuenta y los aspectos principales de la ética de los datos, algoritmos y prácticas.

1. **Transparencia.** La transparencia debe ser entendida como la apertura para brindar información completa sobre et diseño, funcionamiento e impacto que tienen los sistemas de inteligencia artificial. Dicha información no debe comprometer la confidencialidad del modelo de negocio y de la innovación que la hagan susceptible de usos no autorizados.
 - En la **ética de los datos:** en la ética de los datos la transparencia tiene como efecto la necesidad de brindar información sobre las fuentes de datos que se utilizan para el diseño de esta tecnología, sus características y los fines para los cuales serán utilizados los datos, especialmente los datos personales. Esto tiene efectos en los datos utilizados para el entrenamiento e implementación de estos sistemas, dado que se debe ser transparente en su recolección y los criterios utilizados para su clasificación y procesamiento.
 - En la **ética de los algoritmos:** significa transparencia en el modelo detrás de los algoritmos, informando de los criterios que llevan a estos sistemas a tener cierto tipo de resultados. En este punto es relevante brindar información a la ciudadanía sobre los insumos que se utilizan en el diseño de los sistemas y tos resultados que puedan presentarse. Así mismo, debe brindarle información completa sobre tos falsos positivos o negativos que ha generado un algoritmo y los porcentajes de precisión.

- En **ética de las prácticas**: consiste en brindar información completa sobre los equipos y personas involucradas en el diseño y desarrollo de estos sistemas, los objetivos que persiguen y los manuales de conducta que hayan generado para desarrollar su trabajo. Estos manuales deben ser de acceso público. Como práctica para los encargados de la implementación de esta tecnología se debe privilegiar el uso de sistemas de código abierto, especialmente dentro de las entidades públicas.
2. **Explicación.** Este principio es visto a veces como complementario a la transparencia, pero se diferencia en que en la información que es compartida y dada a conocer a la ciudadanía de forma transparente debe ser comprensible tanto para los desarrolladores y usuarios del sistema como para aquellos individuos que pueden verse afectados por sus decisiones y resultados. Por lo tanto, va más allá de la transparencia y busca que el contenido de la información y la forma cómo se presenta sea de fácil acceso, considerando las complejidades que rodean a esta tecnología [3].
- En la **ética de los datos**. La explicación hace que las personas puedan comprender la importancia de los datos en el diseño y desarrollo de estos sistemas, la forma como los recolectan y procesan y los fines de hacerlo, en especial, cuando se procesan datos personales.
 - En la **ética de los algoritmos**. Significa que los algoritmos puedan ser comprensibles en el sentido de que se permita entender los objetivos iniciales que buscan y que son propios del modelo, y los resultados esperados y también los obtenidos. Está claro que los sistemas conocidos como de ‘BlackBox’ o ‘Cajas negras’ tienen baja capacidad de explicación y pueden ser conflictivos. Incluso los desarrolladores de este tipo de tecnología y expertos en la materia no logran entender en su totalidad el procesamiento que tiene lugar y la forma cómo los sistemas llegan a ciertos resultados. Así mismo y, en cualquier caso, se debe presentar información comprensible y clara sobre los objetivos que persigue este sistema en su desarrollo e implementación.
 - En la **ética de las prácticas**. Significa brindar información clara y precisa sobre los roles que tienen las personas involucradas en el diseño, desarrollo e implementación de esta tecnología. Unido a lo anterior, debe generarse durante todo el proceso información clara, precisa y comprensible sobre la forma como estos sistemas están siendo evaluados y generar mecanismos específicos para compartir información sobre los resultados obtenidos, especialmente con las comunidades que se están viendo impactadas por estos sistemas.
3. **Privacidad.** La inteligencia artificial debe estar precedida de un respeto por la intimidad de las personas y su esfera privada que impide el uso información que estos no hayan autorizado y el perfilamiento de individuos a través de esta tecnología.

- En la **ética de los datos**. En este campo el principio lleva a la necesidad de tener autorización para el uso de la información personal, cuando estos no sean datos públicos o bajo las excepciones que la ley específicamente señale. describiendo los fines y objetivos específicos perseguidos con el tratamiento (entrenamiento, funcionamiento, etc.). Igualmente, obliga a desarrollar mecanismos para mejorar la calidad de los datos utilizados y para lograr una actualización constante de la información. Esto lleva a que la población impactada tenga la posibilidad de corregir información personal equivocada o con errores que esté siendo utilizada para el desarrollo u operación de estos sistemas, sin que la funcionalidad de la tecnología pueda imitar este tipo de solicitudes.
 - En los **algoritmos**: El diseño de los algoritmos debe ser respetuoso de la intimidad de las personas y por ende los criterios de decisión no deben basarse en características personales y propias de su esfera privada. Debe limitarse el uso de información personal y solo usar aquella necesaria para un adecuado funcionamiento del sistema y que permita evitar los falsos positivos o negativos. Los diseñadores deben evitar el desarrollo de tecnologías que faciliten el perfilamiento de las personas, bajo criterios que no sean previamente conocidos y autorizados por estos. El uso de información para la mejora del funcionamiento o desempeño de estos sistemas debe ser informado a las personas que sean titulares de esa información.
 - En la **ética de las prácticas**: Deben existir procedimientos internos que desarrollen buenas prácticas en el uso de información y en las respuestas y explicaciones que se brindan a los usuarios impactados por estas tecnologías. Deben privilegiarse las medidas de responsabilidad demostrada que permitan la implementación de herramientas de gestión de riesgo para la privacidad, al igual que los mecanismos de análisis de impacto en privacidad. Los equipos de diseñadores y desarrolladores deben generar criterios para identificar aquellos casos en los que pueden presentarse perfilamientos, su impacto y la forma cómo pueden evitarse resultados negativos a partir de este proceso. Así no solo se protege la privacidad en un sentido individual, sino también colectivo, evitando generar clasificaciones o perfilamientos sociales no deseados.
4. **Seguridad**. Los sistemas de inteligencia artificial no deben generar afectaciones a la integridad y salud física y mental de los seres humanos con los que interactúan.
- En la **ética de los datos**: Se deben implementar mecanismos que permitan asegurar que esta información mantendrá su confidencialidad, integridad y que en ningún momento puede verse alterada. Deben generarse mecanismos que puedan evitar este tipo de alteraciones a la información que estos sistemas utilizan y la forma como procesa la misma.

- En la **ética de los algoritmos**: La implementación de algoritmos y su diseño debe seguir un sistema de riesgos que permita establecer las posibles afectaciones que ciertos resultados pueden generar y la posibilidad de evitarlo. En ningún caso, un algoritmo debe llevar a un resultado que ponga en riesgo la integridad de un ser humano. Dichas decisiones solo pueden ser tomadas por seres humanos, y, en cualquier caso, los algoritmos servirán como modelos de guía en la toma de decisiones que los humanos tomen frente a la vida e integridad de otros (ejemplo: sector salud o sector de seguridad nacional).
- En la **ética de las prácticas**: Se deben evitar aquellas prácticas que pongan en riesgo a los sistemas de inteligencia artificial y los códigos de conducta deben generar parámetros para evitar aquellas actividades que pongan en peligro la integridad y seguridad física de las personas.

5. **Responsabilidad.** Existe el deber de responder por los resultados que produzca un sistema de inteligencia artificial y las afectaciones que llegue a generar. Se partirá de la solidaridad en la responsabilidad de los diseñadores, desarrolladores y personas que implementen esta tecnología, por los daños que el uso de esta tecnología tenga en un individuo, salvo que se demuestre de forma suficiente que la responsabilidad recae en uno solo de estos actores.

- En la **ética de los datos**: Las entidades que recolectan y procesan datos para el diseño, desarrollo e implementación de sistemas de inteligencia artificial deben ser consideradas todas como responsables de esta información y deben responder por su integridad y las finalidades de procesamiento. En ningún caso la responsabilidad podrá recaer solo en uno de estos actores.
- En la **ética de los algoritmos**: Existe responsabilidad de las personas que diseñan un algoritmo por aquellos resultados que llegue a generar y los criterios utilizados para llegar a ciertas respuestas. Sin embargo, su responsabilidad no se deriva hasta la implementación, ya que en esta fase los resultados y su impacto serán responsabilidad de la persona o entidad que se encargue del uso de estos sistemas y de tomar decisiones a partir de los mismos.
- En la **ética de las prácticas**: Los involucrados en el desarrollo de esta tecnología deben establecer responsabilidades claras en la cadena de diseño, producción e implementación. Dentro de los equipos de trabajo debe existir una distribución clara de las funciones y de la responsabilidad en su desarrollo y cumplimiento. Se deben restringir prácticas y acuerdos entre actores que limiten la responsabilidad conforme a la forma como se ha establecido en este principio.

6. **No discriminación.** Los sistemas de inteligencia artificial no pueden tener resultados o respuestas que atenten contra el bienestar de un grupo específico o que limiten los derechos de poblaciones históricamente marginadas. Dichas decisiones solo podrán ser tomadas por seres humanos, bajo los criterios que el marco de derechos humanos permita en cada caso. La funcionalidad de un sistema de inteligencia artificial no debe estar limitado a un grupo específico por razón de sexo, raza, religión, discapacidad, edad u orientación sexual.

- En la **ética de los datos:** Los datos utilizados deben ser analizados de tal manera que se mitigue al máximo la posibilidad de usar información que contenga prejuicios o sesgos (clases), ya sea en su contenido, clasificación o el uso que se le ha dado. Se deben privilegiar mecanismos que permitan hacer un análisis previo de un conjunto de datos y de los posibles problemas que pueda tener.
- En la **ética de los algoritmos:** Los algoritmos deben ser capaces de responder a las necesidades e intereses de distintos grupos poblacionales. El adecuado desempeño de un algoritmo no puede verse limitado a un grupo poblacional específico. Debe existir un seguimiento constante de los falsos positivos y negativos que arroje un sistema estableciendo la forma cómo los criterios de sexo, raza, religión, discapacidad, edad u orientación sexual puedan afectar estos resultados.
- En la **ética de las prácticas:** En el diseño debe participar un grupo diverso de la población y se deben generar matrices de impacto que permitan establecer de forma temprana algún tipo de discriminación y corregir los mismos oportunamente. Debe existir un análisis constante de dicho impacto e incluso considerar mecanismos para retirar inmediatamente sistemas que tengan efectos discriminatorios.

7. **Inclusión.** Es la participación activa de poblaciones históricamente marginadas en el diseño, desarrollo e implementación y evaluación de los sistemas de inteligencia artificial. Se deben utilizar sistemas de inteligencia artificial que hayan cumplido con criterios de inclusión y respondan a las necesidades propias y específicas de estos grupos.

- En la **ética de los datos:** Esto implica la utilización de datos que sean representativos y procedentes de distintos grupos sociales, ya sea para el diseño, entrenamiento o funcionamiento de estos sistemas. Para este fin, se debe aumentar la disponibilidad de conjuntos de datos de los grupos históricamente menos representados.
- En la **ética de los algoritmos:** Las variables que han sido incluidas dentro del algoritmo reconocen los efectos que puede tener en contextos particulares y la posibilidad de que se privilegia a un grupo específico, evitando tal tipo de diseño.

- En la **ética de las prácticas**: Los grupos encargados del diseño, desarrollo e implementación deben tener en cuenta distintos sectores de la sociedad y se deben establecer comités de evaluación para evitar prácticas discriminatorias contra grupos como la mujer, afrodescendientes, indígenas o miembros de la comunidad LGTBI+, entre otros. El desarrollo de sistemas de inteligencia artificial que no tengan un género es deseable.

8. **Prevalencia de los derechos de niños, niñas y adolescentes.** Los sistemas de inteligencia artificial deben reconocer, respetar y privilegiar los derechos de niños, niñas y adolescentes. En ningún caso está justificada la implementación de un sistema inteligente inicial que vaya en detrimento de su interés superior. Se debe abogar por fortalecer programas y estrategias de educación que faciliten el entendimiento de esta población de esta tecnología y facilite la interacción que éstos tengan con esta innovación.

- En la **ética de los datos**: Los datos de esta población no pueden ser utilizados, salvo en aquellas actividades que se relacionen con su interés superior.
- En la **ética de los algoritmos**: El diseño y desarrollo de los algoritmos debe ser comprensible para los niños, niñas y adolescentes, en especial cuando estos tengan un impacto sobre su desarrollo y bienestar. Se debe evitar el diseño de todo algoritmo que repercuta en un perjuicio para los menores de edad y en especial en prácticas como la intimidación y la discriminación.
- En la **ética de las prácticas**: Los niños, niñas y adolescentes deben ser considerados en el desarrollo de estos sistemas cuando sean propios de sus actividades, estableciendo mecanismos específicos de participación que además les permita evaluar el impacto que estos sistemas tienen en esta población. Se deben generar programas de capacitación y educación que les permita a los niños, niñas y adolescentes conocer y entender las características de esta tecnología y sus implicaciones resaltando la formación ética.

9. **Beneficio social.** Los sistemas de inteligencia artificial que se implementan deben permitir o estar directamente relacionados con una actividad que genere un beneficio social claro y determinable. Dicho beneficio puede verse materializado en la reducción de costos, el aumento de la productividad, la facilitación en la prestación de servicios públicos, entre otros. Los sistemas de inteligencia artificial que persigan otro tipo de fines no deben ser implementados en el sector público y se debe desincentivar su uso en otros sectores.

- En la **ética de los datos**: El acceso fácil a los datos y la infraestructura de datos públicos deben ser priorizados para el desarrollo de sistemas de inteligencia artificial que muestran un claro beneficio social, en el diseño de política públicas y la prestación de servicios públicos.

- En la **ética de los algoritmos**. Los modelos y los algoritmos utilizados deben tener como fin último un resultado ligado a un fin socialmente reconocido, por lo que debe demostrarse cómo los resultados esperados se relacionan con dicho fin social.
- En la **ética de las prácticas**: Las personas que trabajan en el diseño, desarrollo e implementación de esta tecnología deben conocer las principales dificultades sociales y establecer la forma cómo esta innovación y la implementación deseada puede ayudar a resolverlas.

8.2. Marco jurídico de la Inteligencia Artificial.

Hay una necesidad de regulación del uso de la tecnología de IA en Europa, porque cada vez está más integrada en la sociedad. Por ejemplo, el tema de la opacidad en muchos de los algoritmos lleva a incertidumbre y a obstaculizar la aplicación efectiva de la legislación vigente en cuanto a seguridad y derechos fundamentales, como puede verse en el documento sobre “Nuevas normas sobre la inteligencia artificial: preguntas y respuestas”

La Comisión Europea propone cuatro niveles de riesgo: (1) riesgo inadmisible; (2) alto riesgo; (3) riesgo limitado; y (4) riesgo mínimo:

(1) *Riesgo inadmisible*: “Se prohibirá un conjunto muy limitado de usos especialmente nocivos de la IA que contravienen los valores de la unión al violar los derechos fundamentales”. (2) *Alto riesgo*: “Se considera de alto riesgo un número limitado de sistemas de IA definidos en la propuesta y que tienen un impacto negativo en la seguridad de las personas o en sus derechos fundamentales (protegidos por la Carta de los Derechos Fundamentales de la UE). A fin de garantizar la confianza y un nivel elevado y coherente de protección de la seguridad y los derechos fundamentales, se proponen requisitos obligatorios para todos los sistemas de IA de alto riesgo. Esos requisitos se refieren a la calidad de los conjuntos de datos utilizados; la documentación técnica y la llevanza de registro; la transparencia y la divulgación de información a los usuarios; la supervisión humana, y la solidez, la precisión y la ciberseguridad. En caso de infracción, los requisitos permitirán a las autoridades nacionales acceder a la información necesaria para investigar si el uso de la IA cumplió la legislación.”. (3) *Riesgo limitado*: “En el caso de determinados sistemas de IA se imponen obligaciones específicas de transparencia, por ejemplo, cuando exista un riesgo claro de manipulación (por ejemplo, mediante el uso de robots conversacionales). Los usuarios deben ser conscientes de que están interactuando con una máquina. (4) *Riesgo mínimo*: “Todos los demás sistemas de IA pueden desarrollarse y utilizarse con arreglo a la legislación vigente sin obligaciones jurídicas adicionales. La inmensa mayoría de los sistemas de IA utilizados actualmente en la UE pertenece a esta categoría. De forma voluntaria, los proveedores de estos sistemas pueden optar por aplicar los requisitos de una IA digna de confianza y adherirse a códigos de conducta voluntarios.”

Para garantizar el cumplimiento, cada Estado miembro designará una o más autoridades nacionales para supervisar la aplicación, ejecución, y realización de actividades de vigilancia de mercado. También designarán una autoridad nacional de supervisión que representará al país en el Comité Europeo de IA.

A pesar de ya existir protección a los derechos fundamentales, debido a la complejidad y opacidad de la IA (“cajas negras”), se establecen requisitos de rendición de cuentas y transparencia para aquellos sistemas de “alto riesgo”. En caso de infracción, se permite a las autoridades nacionales acceder a la información necesaria para investigar si el uso de la IA cumple el derecho de la U.E. En concreto, es importante que no se creen ni produzcan sesgos en cuanto a temas raciales y de género, produciendo resultados falsos. Por ello, los algoritmos y sistemas antes de su comercialización, y después, serán objeto de seguimiento periódico.

El Parlamento Europeo, respecto a cuestiones de interpretación y de aplicación del Derecho Internacional, destaca más de noventa puntos a tener en cuenta, entre los que se pueden resaltar:

“Solicita que se mantenga informado al público sobre el uso de la IA en el ámbito de la justicia, y que dichos usos no den lugar a discriminación derivada de sesgos de programación; subraya que debe respetarse el derecho de toda persona a tener acceso a un funcionario público, así como el derecho del funcionario responsable a tomar personalmente la decisión y a desviarse de la información recibida de la IA cuando lo considere necesario a la luz de los detalles del asunto en cuestión; recalca el derecho de la persona demandada a recurrir la decisión de conformidad con la legislación nacional, sin que se elimine en ningún caso la responsabilidad final del poder judicial.” “Solicita, por tanto, que todos esos usos en el ámbito público y administrativo constituyan información de dominio público y que se evite la discriminación derivada de sesgos de programación”; “Subraya la importancia de hacer posible el despliegue y uso adecuados de la IA; pide a los Estados miembros que formen convenientemente a su personal civil y militar para que le sea posible detectar con precisión y evitar la discriminación y los sesgos en los conjuntos de datos”. “Expresa su profunda preocupación por las tecnologías de ultra falsificación, que permiten falsificar de manera cada vez más realista fotos, audio y video que podrían utilizarse para chantajear, crear bulos, mermar la confianza de la ciudadanía e influir sobre el discurso público; cree que estas prácticas pueden desestabilizar países, pues permiten difundir desinformación e influir en las elecciones; solicita por tanto que se obligue a que en todo documento ultra falseado o video realista realizado mediante técnicas de síntesis sea etiquetado por el creador como no original, limitándose estrictamente su uso para fines electorales y garantizándose firmemente el cumplimiento de esa obligación; pide una investigación suficiente en este sentido al objeto de lograr que las tecnologías para combatir estas prácticas no se queden atrás con respecto al uso malintencionado de la IA”.

A nivel nacional, a pesar de todos los esfuerzos que se siguen haciendo por legislar sobre Inteligencia Artificial, simplemente se cuenta con una ‘ley suave’ centrada en aspectos éticos y estrategias de investigación y desarrollo.

8.3. Bibliografía.

1. Raso, F., Hilligoss, H., Krishnamurthy, V., Krishnamurthy, V., Bavitz, C. and Kim, L. (2018). Artificial Intelligence Human Rights: Opportunities Risks. Berkman Klein Center Research Publication No. 2018-6, Available at SSRN: <https://ssrn.com/abstract=3259344> or <http://dx.doi.org/10.2139/ssrn.3259344>
2. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: <https://ssrn.com/abstract=3518482> or <http://dx.doi.org/10.2139/ssrn.3518482>
3. Hamon, R., Junklewitz, H. and Sanchez Martin, J. (2020). Robustness and Explainability of Artificial Intelligence, EUR 30040 EN, Publications Office of the European Union, Luxembourg.