



La adopción de tecnologías disruptivas en las organizaciones a partir de la creación e implementación de un toolkit ético-legal.

María Lorena Flórez Rojas

Andrés Felipe Rodríguez Murillo

David Orlando Niño Muñoz

Ana María Rey Grajales

Para citar este informe utilice el siguiente formato

Flórez Rojas, M. et al. «La adopción de tecnologías disruptivas en las organizaciones a partir de la creación e implementación de un toolkit ético-legal.», Documentos GECTI. Universidad de los Andes. Agosto 2022.

El Grupo GECTI desde hace 22 años ha buscado dar un aporte académico independiente sobre diferentes aspectos del ciberespacio, la economía digital y la realidad socio tecnológica contemporánea. El grupo ha realizado diversas investigaciones, consultorías, publicaciones y programas académicos de alto nivel especializados en derecho y tecnología.

 @GECTIXXI

María Lorena Flórez Rojas

PhD cum laude Scuola Superiore Sant'Anna en Italia, LLM en Derecho y Tecnología de la Universidad de Tilburgo en Holanda y Abogada de la Universidad de Los Andes. Actualmente es Profesora Asistente de la Universidad de Los Andes y Directora del Grupo de Estudios en Comercio Electrónico, Telecomunicaciones e Informática (GECTI) e Investigadora del Centro CINFONIA de Inteligencia artificial de la misma Universidad.

Correo electrónico: ml.florez66@uniandes.edu.co

Andrés Felipe Rodríguez Murillo

Ingeniero de Sistemas y Computación, Estudiante en Maestría de Ingeniería del software, Asistente Graduado de Maestría en Docencia.

Correo electrónico: af.rodriguezm1@uniandes.edu.co

David Orlando Niño Muñoz

Estudiante de Derecho & Gobierno y Asuntos Públicos, Opción en Psicología. Asistente de Investigación y Monitor de las clases de Títulos Valores, Métodos Cualitativos para Asuntos Públicos, Contratos, Law & Tech, y Constitución y Democracia en la Universidad de los Andes.

Correo electrónico: do.nino@uniandes.edu.co

Ana María Rey Grajales.

Estudiante de Derecho. Opción en Tecnologías de la Información en las Organizaciones. Opción en Edición. Asistente de Investigación y Monitorea de Derecho Procesal General en la Universidad de los Andes.

Correo electrónico: a.reyg@uniandes.edu.co

1. Resumen

El presente documento se plantea como una Guía para la evaluación ética y legal de proyectos que involucran tecnologías disruptivas en especial los proyectos que incluyen inteligencia artificial (IA). Esta caja de herramientas ha sido el fruto de la investigación financiada por el Fondo de Apoyo para Profesores Asistentes (FAPA) de la Universidad de los Andes. La metodología aquí propuesta se ha desarrollado sobre la base de textos especializados en este campo. Esta caja de herramientas ofrece directrices y orientaciones metodológicas para la evaluación previa de aquellos proyectos de cualquier sector que involucren el uso o implementación de IA.

Los servicios de IA basados en el uso de algoritmos se extienden con rapidez tanto en el sector público como en el privado. Un sistema de IA, como explica el Grupo de Expertos en IA de la OCDE, es un “sistema basado en una máquina que puede, para un conjunto dado de objetivos definidos por el ser humano, hacer predicciones, recomendaciones o decisiones que influyen en entornos reales o virtuales”¹. El sistema utiliza datos de entrada basados en la máquina y/o en el ser humano para percibir entornos reales y/o virtuales, abstraer dichas percepciones en modelos, de forma automatizada o manualmente; y utilizar la inferencia del modelo para formular opciones de información o acción.

Además de los beneficios, la IA plantea consideraciones de política pública y se necesitan esfuerzos para garantizar sistemas de IA fiables y centrados en el ser humano. La IA plantea nuevos tipos de preocupaciones éticas y de equidad. El principal de ellos es el respeto a los derechos humanos y a los valores democráticos. Así, se puede afirmar que el uso e implementación de la IA trae consigo diversas implicaciones de carácter ético, social, jurídico y, también, empresarial.

Así, esta caja de herramientas no pretende realizar una definición técnica exhaustiva de este tipo de tecnologías, ni establecer una metodología de evaluación específica para cada una de ellas. Su objetivo

¹ Organización para la Cooperación y el Desarrollo Económico. (2019). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

es exponer una metodología general que sirva como hoja de ruta para los diversos sectores que pretenden incluir en IA como solución a una determinada necesidad.

2. ¿A quién está dirigida?

La caja de herramientas busca ser un insumo de autoevaluación para cada empresa o agencia que pretenda incluir IA dentro de sus proyectos. Así, está especialmente dirigida a las y los responsables del uso de algoritmos, de tratamiento de datos y de la dirección de proyectos de tecnología, aunque también pretende ampliar el conocimiento del público general, cada vez más interesado en comprender estas cuestiones, en especial estudiantes de diversas facultades que estén incursionando con estas tecnologías.

Al tratarse de una serie de herramientas que se adoptan de forma voluntaria y autónoma por diversas entidades y empresas, no se pretende que estas herramientas solucionen los sesgos o riesgos que los diversos proyectos puedan llegar a tener, sino por el contrario, se busca brindar una serie de puntos clave para tener en cuenta y crear alertas tempranas, así como sugerencias sobre el relacionamiento con los diversos actores que se involucran en el proyecto como usuarios finales, desarrolladores, tomadores de decisiones empresariales, equipo legal entre otros. Se trata de la versión beta de autodiagnóstico que está dividida en tres áreas las cuales se identificaron dentro del proceso de ideación del proyecto de IA. Dichas áreas no constituyen conceptos rígidos preestablecidos, sino que ofrecen una orientación para cada equipo al interior de una organización con el fin de interactuar entre los diferentes actores para crear una verdadera gobernanza de la IA.

La primera herramienta es un listado de preguntas con retroalimentación para que los interesados puedan hacer un autodiagnóstico, esta es la herramienta que se pretende retroalimentar a través del conversatorio con expertos. La segunda herramienta se trata de una serie de formatos interactivos para que los sujetos involucrados en el proyecto puedan determinar posibles soluciones a los posibles riesgos. Finalmente, la última herramienta se trata de una guía para empresas o entidades evidenciando los riesgos más comunes del uso de esta tecnología con recomendaciones para adopción al interior que cualquier empresa o entidad.

El listado de preguntas no corresponde a una lista rígida, sino por el contrario, cada empresa o entidad puede adaptar las preguntas según considere que su proyecto se encuentra avanzando o en etapa temprana. Esta versión beta busca ser retroalimentada y comentada por expertos legales, en tecnología y en diseño con el fin de garantizar la facilidad de uso y pertinencia de este. Para la versión final se busca iterar el producto con emprendimientos y entidades de diversos sectores para tener un margen de diagnóstico. Así, esperamos que esta caja de herramientas sea de utilidad, en especial para estudiantes y todos aquellos interesados en el uso ético y responsable de la IA.

3. Introducción

En las sociedades de la información, las transacciones, las decisiones y las elecciones que antes se dejaban en manos de los humanos se delegan cada vez más en los algoritmos, que pueden aconsejar, si no decidir, cómo deben interpretarse los datos y qué acciones deben tomarse como resultado. Cada vez

con más frecuencia, los algoritmos median los procesos sociales, las transacciones comerciales, las decisiones gubernamentales y la forma en que percibimos, entendemos e interactuamos entre nosotros y con el entorno. Las lagunas entre el diseño y el funcionamiento de los algoritmos y nuestra comprensión de sus implicaciones éticas pueden tener graves consecuencias que afecten tanto a individuos como a grupos y sociedades enteras.

Desde el desarrollador, hasta el usuario final, tienen un conglomerado de derechos y deberes frente a la IA y otras tecnologías disruptivas, incluso, este mismo sistema debe desenvolverse bajo los parámetros de la ética y la moral². Así, en palabras de Mark Coeckelbergh “[d]ado que los sistemas de IA ya se usan hoy en día, estas preocupaciones no solo son filosóficamente interesantes, sino que también son muy prácticas y urgentes”³. El proyecto se busca implementar o testear en por lo menos una o dos organizaciones del sector público o privado haciendo uso de la metodología de diseño legal. Esta combinación parte de la creencia que el derecho necesita ser aplicado más allá de los textos y debe ser accesible por sus destinatarios para que pueda ser ejercido de forma autónoma y para que pueda jugar un rol de impulsor de la innovación y salvaguarda de los derechos de los ciudadanos.

De este modo, sea intencionado o no, estamos rodeados de tecnología. Tecnología que puede llegar a excluir a grupos de personas y explotar la psicología humana la cual ha impactado la sociedad de formas que nunca imaginamos⁴. Muchas de las preocupaciones sobre las tecnologías basadas en la IA han sido bien publicitadas, solo para citar algunos ejemplos: Los algoritmos de IA integrados en las tecnologías digitales y de medios sociales pueden reforzar los prejuicios sociales⁵, acelerar la difusión de rumores y desinformación⁶, amplificar las cámaras de eco de la opinión pública, secuestrar nuestra atención y perjudicar el bienestar mental⁷. Así, aunque el temor a que las tecnologías ‘inteligentes’ se ha estudiado en diferentes frentes como el trabajo, la economía, entre otros, es una realidad a la cual nos debemos adaptar.

Teniendo en cuenta los diversos insumos para empresas y entidades gubernamentales que han salido en los últimos años, se busca que esta caja de herramientas se sume a los diversos esfuerzos nacionales e internacionales con miras a orientar empresas, estudiantes, innovadores y reguladores sobre los aspectos éticos y legales que se deben considerar en el desarrollo e implementación de proyectos que involucren tecnologías de IA.

Las personas o entidades que decidan usar esta caja de herramientas lo pueden hacer de forma autónoma con ayuda del formato online en la página web del GECTI. Como resultado de ese ejercicio de autodiagnóstico cada entidad podrá identificar los elementos de mejora que permitirán crear alertas

² Monasterio Astobiza.

³ Coeckelbergh, M. (2019). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 1-18. <https://doi.org/10.1007/s11948-019-00146-8>

⁴ Gran, A. B., Booth, P., & Bucher, T. (2020). To be or not to be algorithm aware: a question of a new digital divide? *Information Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1736124>

⁵ Coeckelbergh, M. (2019). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 1-18. <https://doi.org/10.1007/s11948-019-00146-8>

⁶ Bathaee, Y. (2020). Artificial Intelligence Opinion Liability. *Berkeley Technology Law Journal*, 113. <https://doi.org/10.15779/Z38P55DH32>

⁷ High Level Independent Group on Artificial Intelligence (AI HLEG). (2019). Ethics Guidelines for Trustworthy AI. In *European Commission*.

tempranas frente a tres grandes áreas que se han identificado: I) Desarrollo; II) Ético-legal y III) Organizacional.

Lo innovador de este documento radica en dos aspectos: I) Es un proyecto interdisciplinar que busca la retroalimentación constante de diversas áreas del conocimiento, siguiendo de cerca el llamado de las múltiples partes interesadas. II) Se basa en preguntas orientadoras en las tres áreas identificadas ofreciendo una retroalimentación y recomendaciones en cada una de ellas. Finalmente, es importante resaltar que la elaboración de este producto se basa en los siguientes principios éticos descritos por diversos grupos entre ellos la OCDE⁸, UNESCO⁹, IBM¹⁰ y el Gobierno de Colombia¹¹.

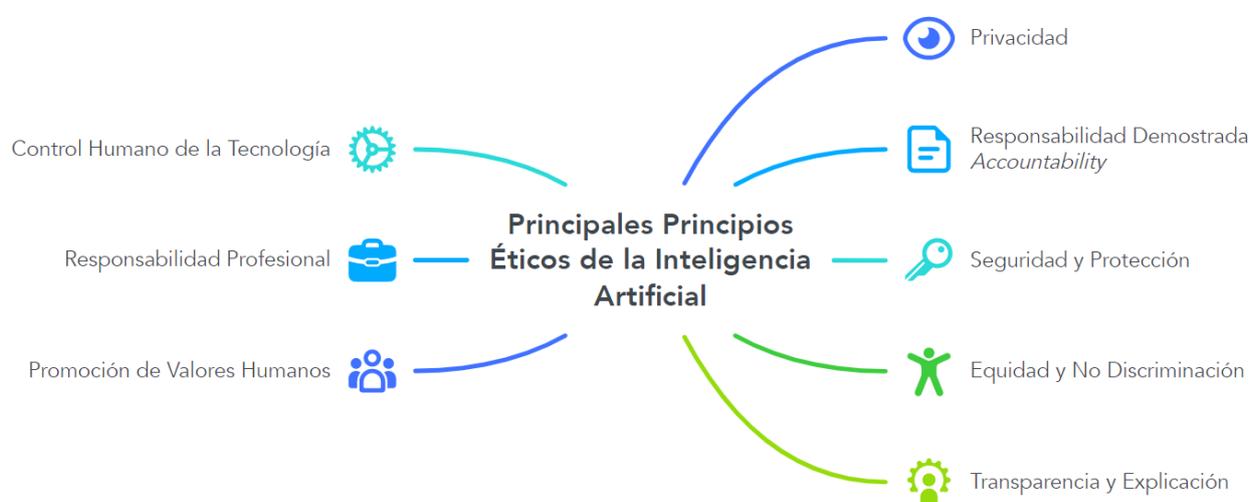


Imagen 1. Principales Principios Éticos de la IA. Elaboración por los autores.

4. Pensamiento de Diseño

A lo largo del proyecto se buscó implementar el pensamiento de diseño con miras a centrar las preguntas orientadoras en el usuario tal y cómo se efectúa en los métodos de privacidad por diseño y por defecto¹²; los análisis de impacto regulatorios y tecnológicos¹³, así como el diseño legal¹⁴.

⁸ Organización para la Cooperación y el Desarrollo Económico. (2019). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

⁹ Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2022). Recomendaciones sobre la ética de la inteligencia artificial. UNESCO 2022. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

¹⁰ IBM. (s.f). Pillars. IBM Artificial Intelligence. <https://www.ibm.com/artificial-intelligence/ai-ethics-focus-areas>

¹¹ Gobierno de Colombia. (2021). Marco ético para la inteligencia artificial en Colombia. Versión final. <https://dapre.presidencia.gov.co/TD/MARCO-ETICO-PARA-LA-INTELIGENCIA-ARTIFICIAL-EN-COLOMBIA-2021.pdf>

¹² Cavoukian, A. (2012). Privacy by design. *IEEE Technology and Society Magazine*, 31(4), 18-19. <https://doi.org/10.1109/MTS.2012.2225459>

¹³ Cruz-Castro, L., & Sanz-Menéndez, L. (2005). Politics and institutions: European parliamentary technology assessment. *Technological Forecasting and Social Change*, 72(4), 429-448. <https://doi.org/10.1016/j.techfore.2004.01.007>

¹⁴ Mayer, S., Schwemmler, M., Nicolai, C., & Weinberg, U. (2021). Assessing the Impact of Design Thinking in Organizations: Foundations of a Framework. *Design Thinking Research*, 255-272. https://doi.org/10.1007/978-3-030-76324-4_14

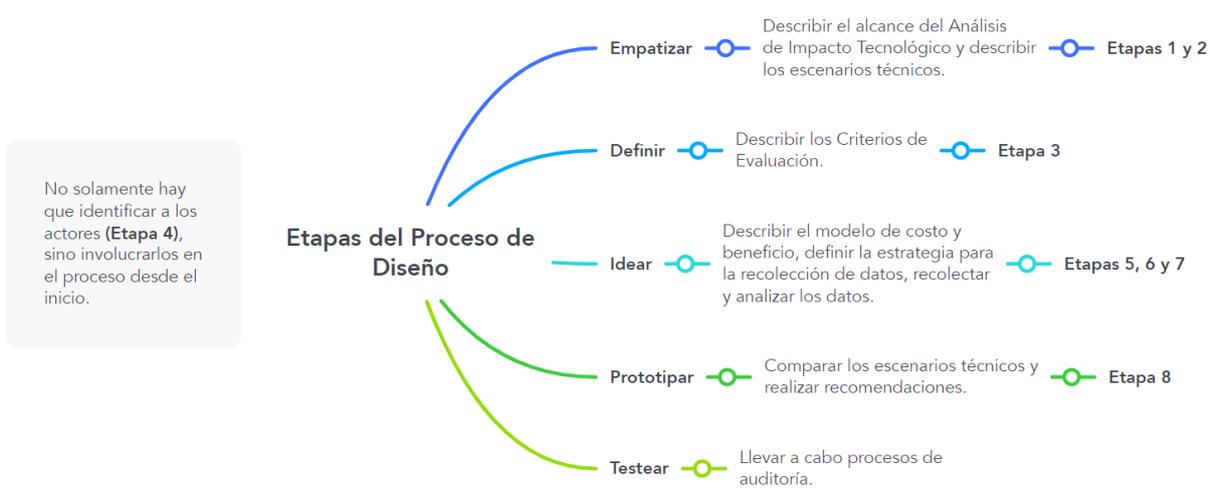


Imagen 2. Etapas del proceso de diseño en contraposición con el análisis de impacto. Elaboración por los autores.

Las preguntas orientadoras se han basado en la similitud entre estas metodologías y las diversas etapas. Las cuales se describen a lo largo de este documento.

1.1. Alcance

El proyecto inicialmente se estructuró dentro de un esquema de uso e implementación de la IA. Sin embargo, con el paso del tiempo nos dimos cuenta que las preguntas orientadoras y en general la información ética y jurídica recolectada podía ser usada por diversas organizaciones que no necesariamente estuvieran usando o implementando IA.

Ese fue uno de los grandes retos, pues encontrábamos literatura que le apuntaba a mencionar que en ocasiones empresas venden un cierto tiempo de software afirmando que es IA pero, que en realidad no lo es. Una de las características comunes en diversos proyectos que se analizaron fue entender en qué parte de la IA se encontraba el proyecto para así entender también los riesgos de este.

Para ello, creamos un mapa que esperamos ayude a varios a comprender el mundo en el que se están desarrollando la mayoría de las aplicaciones mundiales en diversos sectores:

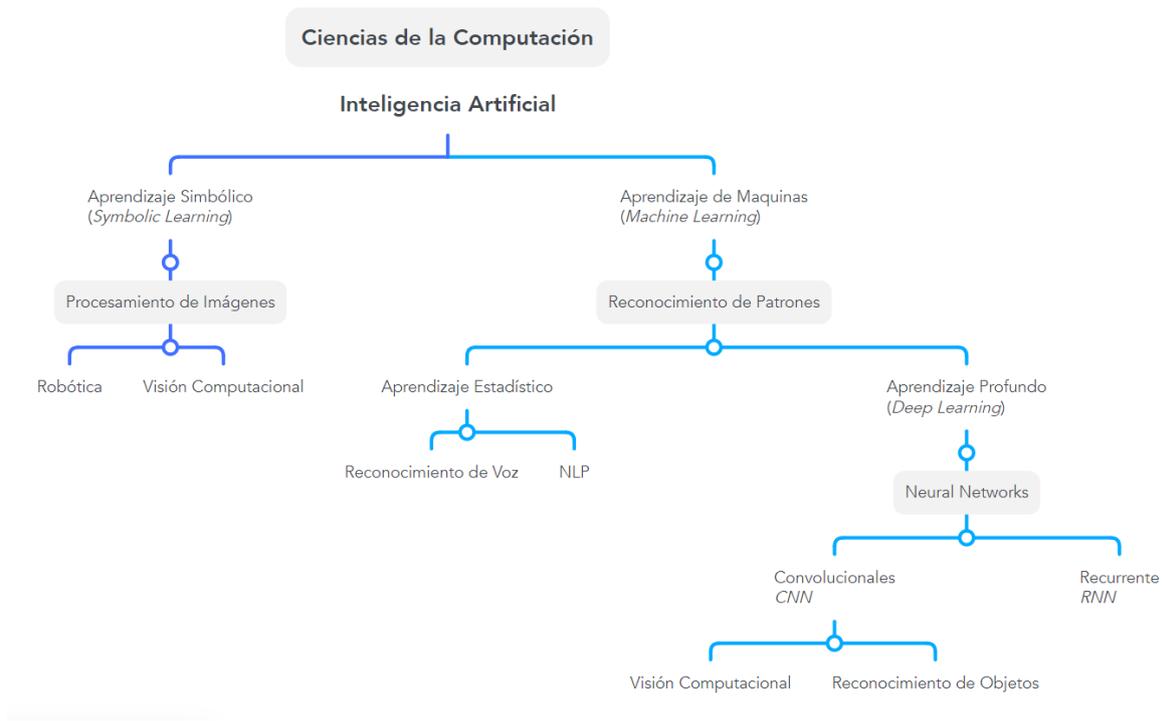


Imagen 3. Estructura de la Inteligencia Artificial. Elaboración por los autores.

¿Qué tareas podría desarrollar un sistema de Inteligencia Artificial

A. Clasificación



Basándonos en probabilidades, los sistemas de IA pueden llegar a clasificar información. Por ejemplo, existen servidores de correo electrónico que filtran automáticamente todos los correos recibidos y nos indican cuales no son tan importantes o son spam.

B. Regresión

Otra labor desarrollada por sistemas de IA consiste en el análisis de múltiples variables con grandes cantidades de datos. Hace parte del «Machine Learning», y establece métodos para relacionar variables.



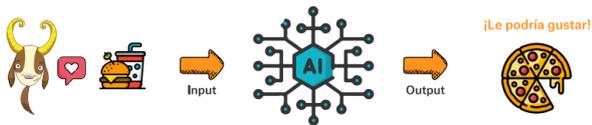
C. Agrupación

La agrupación, o también denominada «clustering», permite identificar ítems en diferentes grupos a partir de ciertas características similares. Un uso de esta funcionalidad se puede evidenciar en la determinación de patrones climáticos.

D. Detección de Anomalías

La detección de anomalías permite reconocer patrones inusuales o con comportamiento inusual a partir de un análisis histórico de datos. Esta aplicación puede ser usada, por ejemplo, para la prevención de lavado de activos o transacciones fraudulentas.





E. Recomendación

Los sistemas de IA también pueden dar recomendaciones a partir del análisis de diferentes variables. Uno de sus usos está en el marketing y sugerencias de compras por internet. El análisis de patrones de búsqueda o productos comprados da nuevas sugerencias similares.

F. Detección de Obtejos

Los modelos de IA tienen la capacidad para reconocer objetos. Así, por ejemplo, existen sistemas para el control de tráfico que pueden diferenciar entre carros, motos, camiones, y dar estadísticas en tiempo real sobre la cantidad de vehículos en una vía.



G. Predicción

Con el análisis de diferentes variables y datos históricos, la IA se ha usado también para predecir los índices de criminalidad en las ciudades. Así, los sistemas indican puntos calientes donde puede enfocarse el esfuerzo policial en determinados lapsos.



1.2. Empatía

En Design Thinking la empatía es una “comprensión profunda de los problemas y realidades de las personas para las que estás diseñando”¹⁵. Implica aprender sobre las dificultades que enfrentan las personas, así como descubrir sus necesidades y deseos latentes para explicar sus comportamientos. Para hacerlo, debemos tener una comprensión del entorno de las personas, así como de sus funciones e interacciones con su entorno.

La empatía nos ayuda a obtener una comprensión y apreciación más profunda de las necesidades emocionales y físicas de las personas y la forma en que ven. También nos ayudará a entender cómo todo esto tiene un impacto en sus vidas en general. A diferencia de la investigación de mercado tradicional, la investigación empática no se preocupa por los datos sobre las personas (como su peso o la cantidad de alimentos que consumen), sino sobre sus motivaciones y pensamientos (por ejemplo, por qué prefieren sentarse en sus casas viendo la televisión en lugar de salir a correr). Es inherentemente subjetivo, ya que hay una buena cantidad de interpretación involucrada en descubrir qué significa la gente en lugar de lo que dicen.

De esta forma, planteamos que un estudio diagnóstico a través de preguntas orientadoras puede acudir más a un estudio objetivo de lista de chequeo, lo cual irrumpiría con el objetivo de la misma fase de empatía. Por esta razón, el diagnóstico preliminar buscar ser el primer paso dentro del análisis final, así, lo que se pretende es que una vez finaliza la primera etapa de diagnóstico inicial, se pasa a una segunda fase para entender a profundidad cada proyecto y sus particularidades. Para luego, en una tercera fase entregar recomendaciones puntuales a cada proyecto.

Es una tarea ardua, sí. Es una tarea que implica tiempo, sí, mucho. Pero, toda cuestión ética que pueda llegar a tener efectos en los derechos humanos de los usuarios finales no puede basarse en un proceso

¹⁵ Mayer, S., Schwemmler, M., Nicolai, C., & Weinberg, U. (2021). Assessing the Impact of Design Thinking in Organizations: Foundations of a Framework. *Design Thinking Research*, 255-272. https://doi.org/10.1007/978-3-030-76324-4_14

estático. Toda vez que, independientemente del tipo de IA que se utilice o se esté explorando en el sector público o privado, los riesgos y las preocupaciones sociales han relucido con miras a adoptar una política para abordar suficientemente estas preocupaciones y diseñar un régimen adecuado de gobernanza de la IA, centrado en la preservación de la relación entre el ser humano y la sociedad¹⁶. De esta forma, **la sensibilización** de los posibles riesgos y preocupaciones no debe ser malinterpretada como un enfoque anti-innovador. Por el contrario, es necesario considerar estas implicaciones de forma adecuada y suficiente para asegurarse de que las nuevas tecnologías, como los sistemas de IA, se construyan y funcionen de forma aceptable para los usuarios y la sociedad en su conjunto.

1.2.1. ¿Cómo entendemos la(s) necesidad(es) de los involucrados en los proyectos?

Cualquier proyecto que involucre nuevas tecnologías o en específico la adopción de IA dentro de algún proceso o proyecto de una empresa o una entidad pasará por diversas áreas. Analizando los casos de estudio más destacados en materia de ética de la IA¹⁷, se podría clasificar las diversas áreas de la siguiente forma:

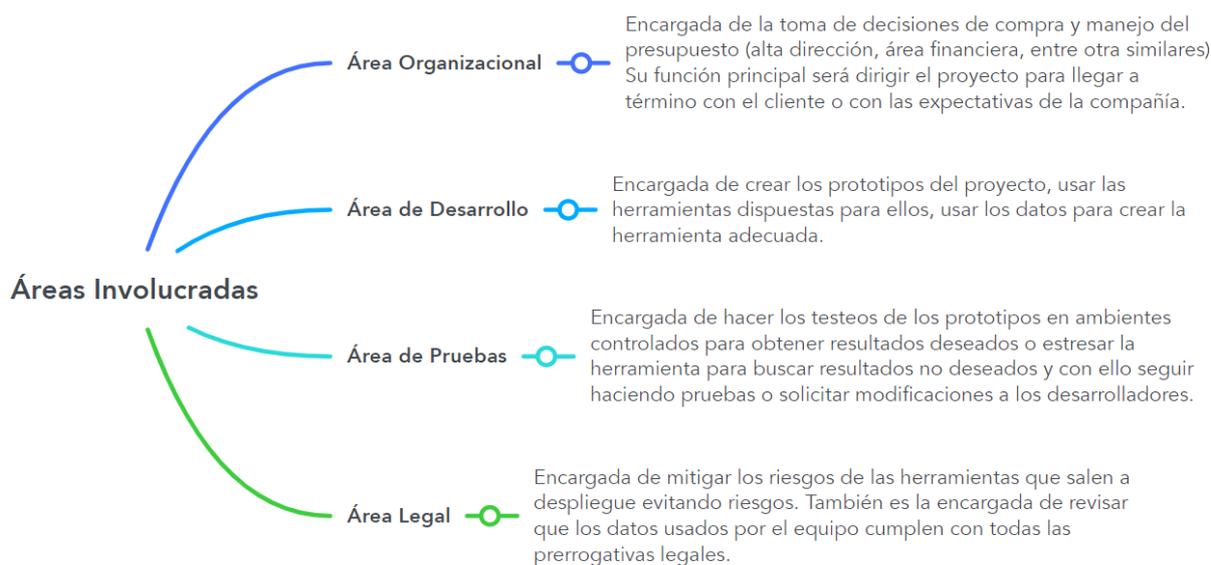


Imagen 4. Áreas Involucradas en Proyectos de IA. Elaboración por los autores.

De esta forma, cada área tiene algún tipo de relación con el proyecto, pero, lo más importante es que cada área en realidad debe estar capacitada para señalar algún tipo de riesgo o posibilidad de riesgo ético. Es así como, actualmente diversas empresas han adoptado un *'chief of ethics'* u oficial de ética. El papel del director de ética de la IA está aumentando en las empresas líderes a medida que la transformación digital se vuelve más compleja y la adopción de la IA crece rápidamente en todas las industrias. Este cargo de reciente revuelo necesita asegurarse de que la tecnología de IA que se desarrolla, utiliza y despliega en su

¹⁶ Nathan, G. (2018). Design-thinking approach to ethical (responsible) technological innovation. *Responsible Research and Innovation*, 286-300. <https://doi.org/10.4324/9781315457291-14>

¹⁷ TayBot, Beauty AI, Watson for oncology, Google photos, entre otros.

organización es digna de confianza. Así, debe asegurarse de que todas las personas de la organización, desde los desarrolladores hasta la alta gerencia, tienen las herramientas, la educación y la formación adecuadas para integrar fácilmente estas propiedades en lo que producen, implementan o prometen.

Así, teniendo en cuenta a la mayoría de los sujetos involucrados dentro un despliegue de un proyecto de IA¹⁸, decidimos que el área de pruebas y el área de desarrollo usualmente trabajan juntas y adoptan un mismo lenguaje técnico por lo que se decidió unir ambas áreas y enfocar las preguntas para este grupo **(Desarrollo y Pruebas)**. Además, revisando la estructura de algunas compañías se pudo evidenciar que la categoría de 'Responsable ético' de una u otra forma se acercaba más al equipo legal- pensando en mitigar posibles daños- así que decidimos unir estas dos áreas **(Ético-Legal)**. Es importante resaltar que el cumplimiento normativo y el cumplimiento ético son dos cosas claramente diferentes, pero para efectos del diagnóstico inicial se han decidido unir con miras a evidenciar las diversas respuestas desde diversas áreas al interior de la entidad.

5. Ideación

Con miras a analizar cómo cada área entiende la ética de un proyecto de IA, se desarrollaron una serie de preguntas aterrizadas a la realidad Colombiana. En el marco de esta investigación nos encontramos con insumos que se enfocaban por ejemplo en las aceleradoras, en los inversionistas de proyectos, en los reguladores, entre otros. Quisimos también llegar con este proyecto a estudiantes que se encuentran adelantando sus estudios y que de alguna forma sienten que su proyecto podría llegar a tener repercusiones a largo plazo de carácter ético.

El proyecto consistía en analizar datos provenientes del extranjero para identificar una enfermedad en edad temprana que usualmente es más común en niñas que en niños, pero que su detección temprana puede llegar a ser beneficiosa a largo plazo.

Sobre este proyecto, lo primero que identificamos fueron los involucrados en el mismo- estudiantes de la Universidad de los Andes. Sin embargo, con ayuda de algunas preguntas que les planteamos, ellos mismos identificaron que podrían existir más sujetos involucrados como, por ejemplo, la empresa proveedora de los datos de entrada que usaron. La conversación con los estudiantes nos permitió explorar la interdisciplinariedad que se ha buscado desde el inicio con el proyecto.

Algunas de las preguntas que se le formularon en la entrevista a los estudiantes fueron:

1. ¿Cuál es la tarea del proyecto de IA que se desea resolver?

Selección Múltiple, Múltiple Respuesta

- a. Clasificación
- b. Regresión
- c. Agrupación

¹⁸ Es importante aclarar que existen proyectos libres en donde los desarrolladores trabajan en comunidades libres en donde aportan con parte del desarrollo, pero en realidad es un proyecto de comunidad. En estos casos la ética en los proyectos debe igualmente ser implementada.

- d. Detección de anomalías
- e. Recomendación
- f. Predicción
- g. Detección de objetos

2. Los datos utilizados son:

Selección Múltiple, Única Respuesta

- a. Datos estructurados
- b. Datos semiestructurados
- c. Datos no estructurados

3. ¿Hay características sensibles que deba controlar en sus datos para observar su efecto en los resultados?

Selección Múltiple, Única Respuesta

- a. Sí
- b. No

4. ¿Considera que los data sets con los que trabaja son una muestra representativa de la población?

Selección Múltiple, Única Respuesta

- a. Sí
- b. No

5. ¿Qué áreas considera que impacta con este proyecto de IA?

Selección Múltiple, Única Respuesta

- a. Fin de pobreza
- b. Hambre cero
- c. Salud y bienestar
- d. Educación
- e. Igualdad de género
- f. Agua limpia y Saneamiento
- g. Energía asequible y no contaminante
- h. Trabajo decente y crecimiento económico
- i. Industria, innovación e infraestructura
- j. Proyección de las desigualdades
- k. Ciudades y comunicados sostenibles

- l. Proyección y consumo responsable
- m. Acción por el clima
- n. Vida submarina
- o. Vida de ecosistemas terrestres
- p. Paz, justicia e instituciones sólidas
- q. Alianzas para lograr los objetivos
- r. Otra (Texto)

6. ¿Durante todo el proceso de desarrollo se registran todas las decisiones tomadas en un sistema de bitácoras o documentación?

Selección Múltiple, Única Respuesta

- a. Sí
- b. No

7. ¿Puede describir la lógica que conecta las variables con la salida de su ecuación?

Selección Múltiple, Única Respuesta

- a. Sí
- b. No

8. El origen de los datos es:

Selección Múltiple, Única Respuesta

- a. Recolección propia
- b. Datos libres (Open Data)
- c. Datos comprados

9. ¿El conjunto de datos utilizado en el proyecto cuenta con datos sensibles?

Selección Múltiple, Única Respuesta

- a. Sí
- b. No

10. ¿Considera que los data sets con los que trabaja son una muestra representativa de la población?

Selección Múltiple, Única Respuesta

- a. Sí
- b. No

11. ¿Considera que los data sets con los que trabaja son una muestra representativa de la población?

Selección Múltiple, Única Respuesta

a. Sí

b. No

12. ¿Considera que los data sets con los que trabaja son una muestra representativa de la población?

Selección Múltiple, Única Respuesta

a. Sí

b. No

Una vez realizadas las preguntas se realizaron las recomendaciones:

- a. Aplicar el principio de transparencia y establecer una bitácora de eventos para dar explicación de cómo funciona el algoritmo en un lenguaje claro y simple.
- b. Aplicar el principio de igualdad y equidad frente a la población colombiana como población infrarrepresentada y dejar la salvedad de que el algoritmo fue creado y testado con datos principalmente del exterior o si se conoce la jurisdicción (país de dónde provienen los pacientes)
- c. Aplicar el principio de gobernanza de datos para identificar el tipo de datos y su nivel de seguridad en el proyecto.

Este es solo un caso en el cual nos podemos dar cuenta que el llenado de las preguntas orientadoras nos da una noción sobre el proyecto, pero no podemos atarnos a las mismas. Se debe ir más allá y tratar de hacer el ejercicio con las diferentes áreas del proyecto, con diferentes personas y seguir testeando.

6. Prototipar

Se analizaron los principales principios y se empezó a desarrollar las preguntas que le apuntaban a cada uno de ellos con miras a acotar diversas áreas y los principios. Sin embargo, para que existiera un diagnóstico preliminar y poder centrarnos en un elemento crucial de cada proyecto se le asignaron puntajes a la cada respuesta para identificar en una escala numérica las preocupaciones éticas más importantes. Así, por ejemplo:

Categoría	Identificador	Estado	Tipo de Pregunta		Pregunta	Opciones de Respuesta			Puntaje de la Respuesta	Max Puntaje de Pregunta
			Tipo	Respuesta Detonante		Tipo de Respuesta	Id	Respuesta		
Organizacional	1	Revisada	Principal	-	¿El equipo de desarrolladores del proyecto en IA cuenta con diversidad en términos de género?	Selección Múltiple, Única Respuesta	A	Sí	1	1
							B	No	0	

Organizacional	2	Revisada	Principal	-	¿El equipo de desarrolladores del proyecto en IA cuenta con diversidad en términos de raza?	Selección Múltiple, Única Respuesta	A	Sí	1	1
							B	No	0	

El puntaje máximo del diagnóstico preliminar es de 45 puntos. Algunas preguntas no tienen ningún puntaje toda vez que se tratan de preguntas sobre el objeto del proyecto o sobre la proveniencia de los datos lo cual no puede clasificarse entre recomendable o no recomendable en términos de los principios escogidos.

Como adjunto al documento se encuentran todas las preguntas.

7. Más allá de la ética

Las metodologías de análisis de impacto se dirigen a las organizaciones que quieren utilizar la IA en sus procesos y quieren analizar las consecuencias legales y éticas. En la fase de diseño, se pueden evitar costosos errores, se puede alertar sobre vulneraciones tempranas, pero, ese mismo ejercicio se puede realzar en diversas fases del proyecto. Pues, así como en materia de datos, la primera finalidad nunca será última, en este caso se puede afirmar que el primer objetivo de un proyecto que involucre IA no necesariamente será el único.

Un proyecto icónico para facilitar este análisis en la práctica es Beauty AI. Este programa se vende como el primer concurso internacional de belleza siendo la Inteligencia Artificial el jurado. A grandes rasgos, los usuarios hacen envío de una foto siguiendo unos parámetros definidos, y el algoritmo “juzgará qué tan bello eres”.

Sin embargo, si pensamos que la recolección masiva de fotos en bruto sobre diversas personas en el mundo es única y exclusivamente para determinar la belleza, muy seguramente estaremos equivocados. Esta recolección masiva puede servir para perfilamiento, sesgos u otro tipo de objetivo no necesariamente concebidos al inicio del proyecto.

Desde la aproximación de las metodologías de análisis de impacto, en un proyecto como Beauty AI, es posible identificar, en etapas tempranas, posibles vulneraciones en temas de sesgos algorítmicos e incluso en la protección de datos. De allí, que un análisis ético desde las primeras etapas del proyecto ayude a mitigar los riesgos asociados.

La ética de la IA es un campo de estudio multidisciplinar en el que el objetivo principal es comprender cómo optimizar el impacto beneficioso de la IA al tiempo que se reducen los riesgos y los resultados adversos para todas las partes interesadas, de forma que se dé prioridad a la agencia y el bienestar humanos, así como al florecimiento del medio ambiente. Para ello, la investigación sobre la ética de la IA se centra en cómo diseñar y construir sistemas de IA que sean conscientes de los valores y principios que deben seguirse en los escenarios de despliegue. También implica la identificación, el estudio y la

propuesta de **soluciones técnicas y no técnicas** para las cuestiones éticas que surgen del uso generalizado de la IA en la vida y la sociedad.

El éxito regulatorio de la IA depende de la gobernanza de IA. Esta se define como proceso de gobierno en el que las reglas y cursos de acción se estructuran, mantienen y regulan, y se asigna la responsabilidad. En general, la gobernanza de la IA se centra en equidad de la IA, ética, responsabilidad y transparencia. El diseño de un modelo particular de gobernanza de IA debe contemplar:

- a. Capacidad de la organización para medir:
 - i. Costos
 - ii. Valor
 - iii. Cumplimiento
 - iv. Auditabilidad
- b. Particularidades del software y los datos que se emplean
- c. Cualquier modificación necesaria del organigrama institucional y su funcionamiento
 - i. Responsable de ética

La ética debe emplearse como herramienta de diagnóstico del contexto en el que se crea, desarrolla o implementa la IA para descubrir puntos ciegos, ampliar el margen de maniobra, y promover la libertad y la autorresponsabilidad. En este sentido, la ética permite identificar cuáles son los asuntos que requieren de regulación legal, y cuáles son los problemas que requieren soluciones consensuadas entre todos los interesados¹⁹. Para fortalecer este papel de la ética, se deben crear indicadores de cumplimiento por cada principio para identificar riesgos y daños asociados a la IA. Si bien esta medida contribuye a la materialización de la ética, no todos los principios son cuantificables, ni comunes a todas las organizaciones y etapas en las que está involucrada la IA. Por eso, la adopción de la ética no puede limitarse a la medición del nivel de riesgo, sino que se requiere de atención constante para detectar riesgos ocultos e incuantificables, y estos indicadores deben apoyarse otras herramientas, como en mecanismos de responsabilidad, documentación, e inventarios de todos los modelos.

A lo largo del proyecto se ha buscado interactuar con diversos agentes involucrados en proyectos de IA los cuales han brindado puntos de vista desde un sector financiero, de gestión de proyectos, de innovación y de beneficio social. Hemos concluido que esta versión beta nos ayuda a dar un diagnóstico preliminar para entender mejor no solo los proyectos que involucran IA sino a su vez las personas que hacen parte de estos proyectos. Comprender que todos son un engranaje de personas que aportan a cada proyecto y

¹⁹ Rességuier & Rodrigues (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data & Society. <https://doi.org/10.1177/2053951720942541>

que a pesar de las dificultades financieras y de tiempo, el beneficio social y la mitigación de riesgos puede potencializar la innovación.

8. Preguntas para comentaristas

- a. Bajo su criterio, ¿qué preguntas (2) son las más relevantes para poner en práctica el principio de transparencia?

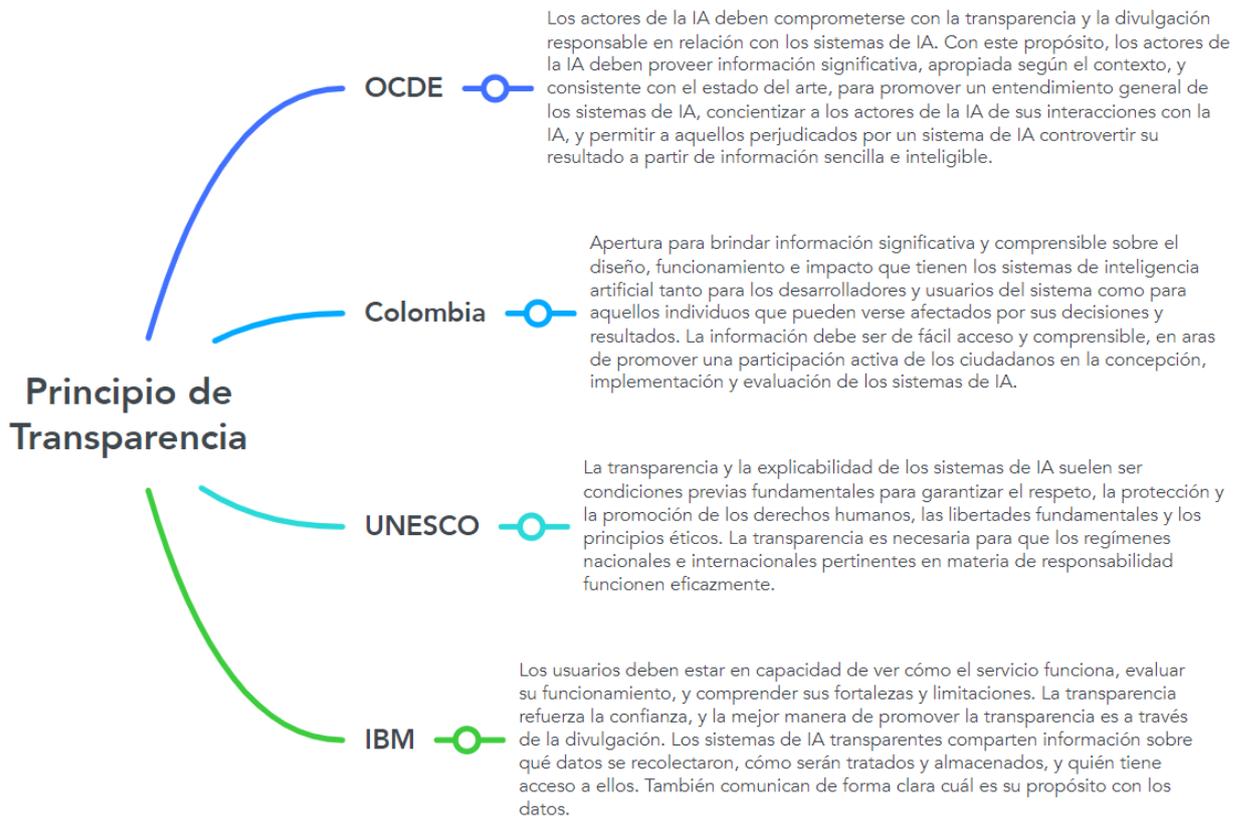


Imagen 5. Aproximaciones al Principio de Transparencia. Elaboración por los autores.

- b. Considera que el estudio de la ética de la tecnología en general, ¿está siendo abordado en las diferentes carreras universitarias?
- c. ¿Considera que el toolkit debe ser implementado en un solo sector o en diversos sectores?

Referencias

- Bathae, Y. (2020). Artificial Intelligence Opinion Liability. *Berkeley Technology Law Journal*, 113. <https://doi.org/10.15779/Z38P55DH32>
- Cavoukian, A. (2012). Privacy by design. *IEEE Technology and Society Magazine*, 31(4), 18-19. <https://doi.org/10.1109/MTS.2012.2225459>
- Coeckelbergh, M. (2019). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 1-18. <https://doi.org/10.1007/s11948-019-00146-8>
- Cruz-Castro, L., & Sanz-Menéndez, L. (2005). Politics and institutions: European parliamentary technology assessment. *Technological Forecasting and Social Change*, 72(4), 429-448. <https://doi.org/10.1016/j.techfore.2004.01.007>
- Gobierno de Colombia. (2021). Marco ético para la inteligencia artificial en Colombia. Versión final. <https://dapre.presidencia.gov.co/TD/MARCO-ETICO-PARA-LA-INTELIGENCIA-ARTIFICIAL-EN-COLOMBIA-2021.pdf>
- Gran, A. B., Booth, P., & Bucher, T. (2020). To be or not to be algorithm aware: a question of a new digital divide? *Information Communication and Society*. <https://doi.org/10.1080/1369118X.2020.1736124>
- High Level Independent Group on Artificial Intelligence (AI HLEG). (2019). Ethics Guidelines for Trustworthy AI. In European Commission.
- IBM. (s.f). Pillars. IBM Artificial Intelligence. <https://www.ibm.com/artificial-intelligence/ai-ethics-focus-areas>
- Mayer, S., Schwemmler, M., Nicolai, C., & Weinberg, U. (2021). Assessing the Impact of Design Thinking in Organizations: Foundations of a Framework. *Design Thinking Research*, 255-272. https://doi.org/10.1007/978-3-030-76324-4_14
- Nathan, G. (2018). Design-thinking approach to ethical (responsible) technological innovation. *Responsible Research and Innovation*, 286-300. <https://doi.org/10.4324/9781315457291-14>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2022). Recomendaciones sobre la ética de la inteligencia artificial. UNESCO 2022. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
- Organización para la Cooperación y el Desarrollo Económico. (2019). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

Rességuier & Rodrigues (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*. <https://doi.org/10.1177/2053951720942541>

