



COMITÉ ESPAÑOL
DE REPRESENTANTES
DE PERSONAS
CON DISCAPACIDAD

Guía CERMI de análisis para la evaluación de algoritmos inclusivos



Índice

1	Introducción.....	4
2	Antecedentes y justificación.....	5
3	Guía de análisis para CERMI	10
3.1	Categoría: Inclusión y No Discriminación	11
3.1.1	Categoría de Inclusión y No Discriminación, enfoque de Inclusión:.....	12
3.1.2	Categoría de Inclusión y No Discriminación, enfoque de Sesgos:	13
3.2	Categoría: Transparencia y explicabilidad	13
3.2.1	Categoría Transparencia y Explicabilidad, enfoque de Datos:.....	14
3.2.2	Categoría Transparencia y Explicabilidad, enfoque de Explicabilidad:.....	15
3.3	Categoría: Accesibilidad universal.....	16
3.3.1	Categoría de Accesibilidad Universal, enfoque de Interfaz:.....	16
3.3.2	Categoría de Accesibilidad Universal, enfoque de Accesibilidad:	17
3.4	Categoría: Unidad y participación	18
3.4.1	Categoría de Unidad y Participación, enfoque de Impacto:.....	19
3.4.2	Categoría de Unidad y Participación, enfoque de Participación:	19
3.5	Categoría general: aplicación práctica y responsabilidad	20
3.6	Ejemplo de resultados	21
3.7	Relación con DUDH y CDPD	22
3.8	Guía de análisis simplificada	23
4	Revisión de casos de uso	26
5	Recomendaciones para el uso de la Guía de Análisis	33

5.1	Recomendaciones para un modelo de Gobernanza.....	33
5.2	Como utilizar el cuestionario	35
5.3	Creación de un repositorio común CERMI	36
5.4	Formación en utilización del formulario.....	36
6	<i>Bibliografía</i>	37
7	<i>Anexos</i>	39
7.1	Cuestionario completo	39
7.2	Aproximación a DUDH y CDPD.....	42

1 Introducción

En el contexto actual, donde los sistemas de Inteligencia Artificial (IA) desempeñan un papel cada vez más significativo en diversas áreas de nuestra vida, es esencial abordar su desarrollo y aplicación con una perspectiva ética y accesible. El proyecto **MetaDiverso, CERMI Discapacidad** ha sentado las bases para un análisis exhaustivo de los sistemas de IA, destacando la importancia de evaluar estas tecnologías desde una mirada inclusiva y considerando las necesidades específicas de las personas con discapacidad.

Esta guía de análisis surge como una herramienta esencial para profundizar en esta dirección y se centra en los principios éticos elegidos por el Comité Español de Representantes de Personas con Discapacidad (CERMI) como ejes fundamentales. El objetivo es proporcionar un marco detallado y operativo para evaluar los sistemas de IA, no tanto en cuanto a su cumplimiento técnico y funcional, sino principalmente en términos de su impacto social y ético y su consideración de acuerdo con los criterios de inclusividad y accesibilidad universal.

El enfoque de esta guía es holístico y multidimensional, buscando abarcar cada aspecto relevante que influye en la forma en que los sistemas de IA interactúan con las personas usuarias, especialmente aquellas con discapacidad. Para cada uno de los principios de CERMI, se ha desarrollado una serie de preguntas y criterios de análisis que permiten una evaluación detallada y específica. Estos criterios están diseñados para ayudar a quienes se ocupan del desarrollo, analistas y personas involucradas en la toma de decisiones a identificar áreas de mejora, garantizar la conformidad con los más altos estándares éticos y promover un enfoque más inclusivo y accesible en la creación y aplicación de tecnologías de IA.

Además, esta guía establece una conexión directa entre los principios éticos de la IA y los derechos humanos fundamentales, alineándose con las directrices de la Declaración Universal de Derechos Humanos (DUDH) y la Convención sobre los Derechos de las Personas con Discapacidad (CDPD). Al hacerlo, subraya la importancia de considerar los sistemas de IA no solo como herramientas

tecnológicas, sino como facilitadores o posibles obstaculizadores de derechos y libertades esenciales.

La presente guía es, por tanto, un paso adelante en el camino hacia sistemas de IA más responsables y alineados con los valores de una sociedad inclusiva y respetuosa de la diversidad humana. Su aplicación práctica tiene el potencial de mejorar significativamente la manera en que la tecnología de IA se integra en nuestras vidas, asegurando que beneficie a todas las personas en la sociedad de manera equitativa y justa.

2 Antecedentes y justificación

La Inteligencia Artificial (IA) ya es una herramienta transformadora en numerosos campos y sectores, tales como la educación, la industria, los servicios, o incluso la salud. Al tiempo que ofrece enormes oportunidades, su rápida evolución y adopción ha planteado también preocupaciones significativas relacionadas con la ética, la equidad y la accesibilidad. Desde 2015, se han venido publicando numerosos documentos sobre los principios éticos que deben guiar la IA, y se han generado intensos debates sobre las implicaciones éticas de su uso. Estos debates han llevado a la creación de conjuntos de principios, marcos éticos y cuestionarios de impacto, que persiguen identificar y abordar los desafíos éticos inherentes a la IA.

Se han llevado a cabo avances significativos en técnicas para detectar y mitigar sesgos en algoritmos, así como para mejorar su explicabilidad. No obstante, a pesar del esfuerzo en la valoración del impacto e identificación de problemas éticos, estos esquemas carecen del necesario contexto para clasificar cuestiones éticas relativas a sistemas de IA. Un reciente estudio¹ de revisión de más de cien documentos sirve como base para la clasificación de herramientas utilizadas en el análisis de sistemas de IA. Se realiza una categorización preliminar de enfoques y se analizan las referencias en relación con los principios éticos. A continuación, se describen las diferentes categorías de documentos encontrados (ver resumen en Tabla 1):

- **Visión general:** Suelen ser artículos que incluyen conceptos básicos acerca de la ética de la IA y proporcionan recomendaciones genéricas
- **Casos de uso:** Son normalmente ejemplos reales de aplicaciones de IA en los que se tratan problemas éticos y cómo pueden ser resueltos.
- **Marcos de trabajo:** Generalmente suelen servir como un esqueleto para poder guiar en la resolución de distintas problemáticas o aspectos éticos, y ayudan a entender las fronteras entre distintas cuestiones éticas.

¹ <https://link.springer.com/article/10.1007/s43681-023-00258-9>

-
- **Formularios o guías de análisis:** En la mayoría de las situaciones se usan como una lista de comprobación en el diseño de un sistema.
 - **Métricas e indicadores:** Normalmente se utilizan como referencia para medir los sesgos o la explicabilidad de los algoritmos.
 - **Modelado de procesos:** Generalmente sirven como descripción visual, que ayuda a entender los posibles problemas éticos dentro de un flujo de trabajo.
 - **Recomendaciones y códigos de prácticas:** Un conjunto de reglas, normalmente publicadas por una asociación, que señalan aspectos clave para el diseño de una IA ética. A veces los códigos de prácticas no son un patrón, pero sí que hacen recomendaciones que pueden ser usadas dentro de una comunidad de práctica concreta.
 - **Métodos algorítmicos:** Normalmente son técnicas computacionales para implantar o mejorar un sistema de IA. Pueden incluir pseudocódigo, representaciones gráficas o descripciones lingüísticas.
 - **Software:** Ya sean librerías de todo tipo o software comercial que puede ayudar no sólo a la detección de los problemas éticos sino también a la resolución o mitigación de estos.
 - **Conjunto de datos:** Grupos de datos que pueden ayudar a entrenar de manera “ética” distintos tipos de algoritmos.
 - **Auditorías:** Exámenes formales de los aspectos éticos de un sistema de IA incluyendo los componentes, requerimiento, comportamiento...etc.
 - **Modelos de licencia:** Normalmente son patrones que puede ser usados para crear las reglas de uso del sistema de IA.

Figura 1. Tabla que describe la clasificación de guías de análisis en base a principios.

Tipo de aproximación	Tipo de problema ético
Guías de análisis y formularios	Información acerca de los principios éticos proporcionada por los grupos de interés
Algoritmos	Privacidad Explicabilidad Equidad y no discriminación
Métricas	Equidad y no discriminación
Modelado de procesos	IA responsable Revisión de respuestas de los grupos de interés
Librerías de software	Transparencia Explicabilidad Equidad y no discriminación
Auditorías	Privacidad Equidad y no discriminación
Software	Privacidad
Conjunto de datos	Diversidad Equidad y no discriminación

Desde otra perspectiva, si consideramos el ciclo de diseño de las aplicaciones de *machine learning*, siguiendo el modelo de CRISP-DM, también se puede hacer una clasificación de cuáles serían los métodos recomendados en cada una de las partes del proceso, como se describe en la Tabla 2.

Figura 2. Tabla que describe la clasificación de guías de análisis en base a CRISP-DM.

Etapas de desarrollo de un sistema de IA	Tipo de aproximación
Entendimiento del negocio	Marco de trabajo
Diseño del sistema	Marco de trabajo
Creación de datos	Guía de análisis o formularios
Comprensión de los datos	Auditoría
Entrenamiento del modelo	Algoritmos
Evaluación y test	Auditorías y métricas
Despliegue	Licenciamiento del modelo

Como se puede observar, en la mayoría de los casos cuando se trata de entender cómo un determinado caso de uso se está comportando respecto a los principios éticos de la organización, se recurre en una primera aproximación a las **guías de análisis** o formularios. El objetivo es valorar el tipo de riesgo, los principios éticos que sean de aplicación y aspectos específicos sobre qué método posterior (según la Tabla 1 y la Tabla 2) puede resultar más conveniente aplicar. Las tablas no representan una relación biunívoca sino más bien una recomendación de que aproximación usar en cada momento, pero dependerá en gran medida del caso de uso y el tipo de organización.

Entre las conclusiones de este análisis, conviene destacar el hecho de que mientras algunos problemas éticos concretos como la explicabilidad, la no discriminación y la privacidad suelen abordarse con herramientas prácticas, otras cuestiones se manejan principalmente con marcos conceptuales, guías de análisis o modelos de procesos. Es por tanto necesario disponer de una guía detallada y práctica que permita a las empresas y organizaciones evaluar sus sistemas de IA no solo por su eficiencia y efectividad, sino también por su conformidad con los principios éticos fundamentales, y con relación al caso de uso específico que se aborde.

Efectivamente, según el Instituto de Ada Lovelace², las guías de análisis son el método más usado junto con las auditorías algorítmicas. Generalmente las guías de análisis y formularios van a tratar de evaluar el impacto del sistema de IA para evitar las posibles implicaciones éticas antes de desarrollarlo, mientras que en el caso de las auditorías se enfocan más en los sistemas ya en funcionamiento.

Las guías de análisis suelen abordar cuestiones menos técnicas, pero a la vez tienen que ser capaces de reconocer los aspectos clave de los sistemas de IA: diseño, los datos, el modelado y el despliegue. Este tipo de herramientas ayuda a establecer una cultura de diseño de IA ético, y concienciar a todos los participantes a comprometerse

² Lovelace, A., & DataKind, U. K. (2020). *Examining the black box: Tools for assessing algorithmic systems*. Technical report, AdaLovelace Institute, <https://ico.org.uk/media/about-theico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf>.

con el uso ético de los datos y los sistemas de IA. Diferentes organismos y organizaciones han creado guías de análisis que pueden servir de referencia. Algunas de las guías de análisis más relevantes y usadas son:

- **ALTAI (List for Trustworthy Artificial Intelligence)**³ es una herramienta práctica que ayuda a organizaciones, empresas y personas desarrolladoras a evaluar la confiabilidad de sus sistemas de IA. Fue creada por las personas expertas HLEG (*High Level Expert Group*) de la Unión Europea en el marco de la publicación del documento Guías éticas para una IA confiable publicado por la Unión Europea en 2019.⁴
- **Herramienta de “Self-assessment” del Gobierno de Dubai**⁵ no solamente ofrece un cuestionario sino también medidas que se podrían aplicar para mitigar los problemas identificados.
- **Autoevaluación ética de IA para emprendedores del Banco Interamericano de desarrollo**⁶, es también un cuestionario que ayuda a emprendedores a identificar las áreas principales donde pueden estar los problemas éticos y prevenirlos.
- **Autoevaluación de IA ética del gobierno de Australia**⁷ que permite identificar procesos y soluciones sobre cómo se han aplicado los ocho principios éticos que han decidido evaluar.
- **Herramienta de análisis de impacto de la Unesco**⁸ que proporciona una autoevaluación con base en seis principios éticos, además de proporcionar una guía para mitigar el impacto.

Además de estas guías hay otro tipo de herramientas, más parecidas a la tipología de marco de trabajo como el *Data Ethics Framework* del Gobierno de Reino Unido⁹, la

³ AI HLEG (2020) The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, European Commission, Brussels, <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence>

⁴ <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

⁵ Digital Dubai Office (2020) AI System Ethics Self-Assessment Tool, <https://www.digitaldubai.ae/self-assessment>

⁶ Autoevaluación ética de IA para actores del ecosistema emprendedor: Guía de aplicación. Banco Interamericano de Desarrollo, <http://dx.doi.org/10.18235/0003269>

⁷ Australian Government: Department of industry, Science, Energy and Resources. (2019). Australia’s Artificial Intelligence Ethics Framework, <https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework>

⁸ https://www.unesco.org/en/articles/ethical-impact-assessment-tool-recommendation-ethics-artificial-intelligence?TSPD_101_R0=080713870fab20004d3ed789c1195c05fcb31c4e68014c214df5fdf21a6acc8e3c87ce6be930c1b0838384d4014300017da76e3fa8cbd0745482cf7ea21112272852a25d3ea00dcc5b68a1f295f5095fec75b125ecb1bb132ad2061e45e5891

⁹ Data Ethics Framework. United Kingdom Government, <https://www.gov.uk/government/publications/data-ethics-framework>

guía de diseño responsable para sistemas de IA del instituto de Alan Turing¹⁰ o el recurso proporcionado por el World Economic Forum¹¹ que consta de 13 módulos destinados a fomentar el conocimiento y actuar en las preocupaciones éticas del uso de los sistemas de IA.

La mayoría de estos recursos son muy generales y no tienen en cuenta el ámbito de trabajo, el sector o el contexto donde se desarrolla el sistema de IA. Por ello, el presente documento se desarrolla para proporcionar un marco detallado para evaluar los sistemas de IA en términos de los principios éticos desarrollados por el CERMI: inclusión y no discriminación, igualdad entre mujeres y hombres, transparencia y explicabilidad, accesibilidad universal, unidad y participación, y sostenibilidad. Esta guía busca ofrecer un enfoque práctico y aplicable, alineado con los estándares éticos y los derechos fundamentales, asegurando que la IA sirva al bienestar de toda la sociedad, incluyendo especialmente a las personas con discapacidad.

¹⁰ A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.3240529>

¹¹ Empowering AI Leadership, <https://www.weforum.org/publications/empowering-ai-leadership-ai-c-suite-toolkit/>

3 Guía de análisis para CERMI

Los seis principios éticos establecidos por el Comité Español de Representantes de Personas con Discapacidad (CERMI) representan una guía esencial para el desarrollo y uso responsable de la Inteligencia Artificial (IA), enfocándose en asegurar la inclusión, la equidad y el respeto por los derechos humanos. Dichos principios son los siguientes:

- Inclusión y no discriminación.
- Igualdad entre mujeres y hombres.
- Transparencia y explicabilidad.
- Accesibilidad universal.
- Unidad y participación.
- Sostenibilidad.

Para el desarrollo de esta guía de análisis se han identificado subdimensiones o enfoques de estos principios, y a continuación se ha desarrollado un cuestionario con las preguntas fundamentales a valorar. Dado el número de cuestiones resultante se ha decidido elaborar finalmente dos cuestionarios: un cuestionario estándar con todas las preguntas, que permite un análisis más detallado, y un cuestionario simplificado con una serie de cuestiones clave, que permite un análisis preliminar más rápido. Para las preguntas se ha optado por la utilización de una escala de 1 a 5. La adopción de una escala de 1 a 5 para valorar las preguntas en nuestra guía de análisis de sistemas de IA ofrece una evaluación matizada y detallada, permitiendo diferenciar entre varios grados de cumplimiento o implantación de los principios éticos, siendo el valor de 1 de menor cumplimiento y el valor de 5, el de mayor cumplimiento. Esta escala es intuitiva y fácil de usar, facilitando la comparación y el seguimiento del progreso a lo largo del tiempo. Permite cuantificar aspectos cualitativos y proporciona una base sólida para decisiones informadas, fomentando la reflexión crítica y la mejora continua en prácticas éticas, inclusivas y accesibles en el ámbito de la IA.

En primer lugar, con base en los principios, se ha desarrollado un conjunto de categorías y enfoques, que se resumen en la Tabla 3. A partir de estos enfoques, se han desarrollado las preguntas del cuestionario.

La sostenibilidad es un aspecto crucial en el desarrollo de sistemas de IA, pero su inclusión en nuestra guía presenta ciertos desafíos. La evaluación de la sostenibilidad a menudo requiere un análisis técnico complejo y detallado, lo que podría sobrepasar el alcance de esta guía inicial y la experiencia de quienes la van a usar. Por ello, hemos optado por excluir preguntas directas sobre sostenibilidad, con la intención de evitar evaluaciones superficiales o incorrectas. Sin embargo, mantenemos la sostenibilidad como un ideal subyacente y recomendamos consultas con especialistas para aquellas personas interesadas en profundizar en la contribución de sus sistemas de IA a los objetivos de sostenibilidad global.

Figura 3. Tabla que recoge las categorías de análisis.

Principio Ético IA – CERMI	Categoría resultante	Enfoques
Inclusión y no discriminación	Inclusión y Equidad	Inclusión
Igualdad entre mujeres y hombres		Sesgos
Transparencia y explicabilidad	Transparencia y explicabilidad	Datos
		Explicabilidad
Accesibilidad universal	Accesibilidad universal	Interfaz
		Accesibilidad
Unidad y participación	Unidad y participación	Impacto
		Participación
Sostenibilidad	-	

A continuación, se detalla cada categoría, sus enfoques y las preguntas que conforman el cuestionario estándar. El cuestionario completo se puede consultar en el Anexo.

3.1 Categoría: Inclusión y equidad

A partir de los principios de inclusión y no discriminación, y de igualdad entre mujeres y hombres, se ha propuesto una categoría unificada denominada “inclusión y equidad”, que refleja una comprensión integral de cómo la IA debe desarrollarse y utilizarse para respetar, proteger y promover los derechos humanos de todas las personas, con un énfasis particular en las personas con discapacidad y en las mujeres y niñas con discapacidad. Esta categoría unificada aboga por sistemas de IA que luchen activamente contra la discriminación y promuevan la diversidad y la inclusión, prestando especial atención a quienes están en mayor riesgo de exclusión. Al integrar estos principios en esta categoría, se adopta un enfoque más amplio que reconoce la inclusión no solo como la ausencia de discriminación, sino también como la promoción activa de la igualdad y la equidad.

La brecha digital, un problema central en ambos principios, afecta a la inclusión y equidad. Unificar estos principios subraya la importancia de combatir esta brecha para todos los grupos en situación de vulnerabilidad, especialmente para las mujeres y las niñas con discapacidad.

Dentro de esta categoría se diferencian dos enfoques que comprenden diferentes cuestiones: inclusión y sesgos. El enfoque de inclusión se centra en garantizar que los sistemas de IA beneficien a todas las personas usuarias, especialmente a las personas con discapacidad. Evalúa si el sistema de IA contempla y sirve efectivamente a una amplia gama de personas usuarias, incluyendo aquellas con discapacidad, promoviendo la representación diversa en su diseño y funcionamiento. El enfoque de sesgos se ocupa de identificar y mitigar cualquier forma de sesgo o discriminación en los sistemas de IA, para evitar la discriminación y asegurar la equidad en el tratamiento de todas las personas usuarias.

A continuación, se incluyen las preguntas para cada uno de los enfoques:

3.1.1 Categoría de Inclusión y Equidad, enfoque de Inclusión:

- **Diversidad y representatividad en datos:** Evalúa si se ha considerado la diversidad y representatividad de las personas usuarias finales en la recopilación de datos utilizados por el sistema de IA, lo cual es fundamental para garantizar que el sistema funcione equitativamente para una amplia gama de personas usuarias.
- **Pruebas para grupos específicos de discapacidad:** Pregunta si se han realizado pruebas específicas para grupos de personas con discapacidad o para casos de uso que presenten desafíos particulares, asegurando que el sistema sea inclusivo y accesible.
- **Adaptabilidad del sistema a diversas necesidades:** Evalúa si el sistema de IA es lo suficientemente flexible para adaptarse a un amplio espectro de preferencias y habilidades individuales, incluyendo la compatibilidad con tecnologías de apoyo para personas con discapacidad.
- **Pruebas y validación con grupo diverso:** Mide si el sistema de IA ha sido probado y validado con un grupo diverso, incluyendo específicamente personas con discapacidad, para garantizar su efectividad y accesibilidad para todas las personas usuarias.
- **Interacción del equipo de desarrollo con personas usuarias finales:** Investiga si el equipo responsable de construir el sistema interactuó con personas usuarias finales potenciales, incluyendo personas con discapacidad, para comprender mejor sus necesidades y validar la inclusión.
- **Consultas con grupos afectados:** Valora si se ha consultado con los grupos de personas con discapacidad afectadas para definir y comprender mejor los aspectos de inclusión, justicia y equidad en el desarrollo del sistema de IA.
- **Inclusión de personas con discapacidad en el desarrollo:** Pregunta cómo se valoraría la inclusión de personas con discapacidad en el equipo de desarrollo e implantación del sistema de IA, para asegurar que sus perspectivas y necesidades sean consideradas en el diseño y evaluación del algoritmo.
- **Igualdad de género en datos y desarrollo:** Esta pregunta busca evaluar el compromiso y las acciones realizadas para fomentar la igualdad de género dentro de la inteligencia artificial. Se centra en la inclusión de mujeres y hombres con discapacidad en los conjuntos de datos y en los equipos de desarrollo, con la intención de reflejar una representación equitativa y aprovechar una variedad de experiencias y necesidades para enriquecer el sistema de IA.

3.1.2 Categoría de Inclusión y Equidad, enfoque de Sesgos:

- **Evaluación de impacto en grupos específicos:** Esta pregunta se centra en si se ha evaluado si ciertos grupos, especialmente las personas con discapacidad, podrían verse afectadas

desproporcionadamente por los resultados del sistema de IA. Es crucial para garantizar que el sistema no perpetúe ni exacerbe las desigualdades existentes.

- **Mecanismo para señalar problemas de prejuicios y discriminación:** Evalúa si se ha implantado un sistema efectivo para identificar y abordar problemas relacionados con prejuicios, discriminación o mal desempeño del sistema de IA, lo que permite una respuesta rápida y adecuada a tales problemas.
- **Uso de variables y 'proxies' no discriminatorias:** Indaga si el sistema de IA ha sido diseñado para evitar el uso de variables o 'proxies' que puedan resultar en discriminación injusta, asegurando que todas las decisiones se tomen de manera justa y equitativa.
- **Medidas de inclusión y equidad:** Pregunta sobre las medidas específicas implantadas para probar el sistema de IA en diferentes grupos afectados y para asegurar que no existan tasas de error desproporcionadas, lo que es crucial para garantizar la equidad en el trato a todas las personas usuarias.
- **Procedimientos para identificar y mitigar sesgos:** Evalúa si se han implantado y evaluado procedimientos efectivos para identificar y mitigar sesgos en el sistema de IA, lo cual es fundamental para prevenir la creación o el refuerzo de discriminaciones o prejuicios.
- **Revisión y monitorización continua de sesgos:** Se centra en si existe un proceso continuo para revisar y monitorear el sistema de IA a lo largo de su ciclo de vida con el fin de detectar y corregir cualquier sesgo y discriminación.
- **Iniciativas formativas sobre sesgo potencial:** Investiga si se han desarrollado programas de formación y concienciación para personas que diseñan y desarrollan IA, con el fin de concienciarlas sobre el potencial de sesgo en sus trabajos y cómo mitigarlo.
- **Evaluación de medidas contra sesgos de género:** Esta pregunta invita a calificar la eficacia de los procedimientos establecidos para abordar y neutralizar los sesgos de género en los sistemas de IA. Se enfoca en la capacidad del sistema para producir resultados justos y no discriminatorios, considerando especialmente el impacto en las mujeres y niñas con discapacidad y promoviendo la igualdad de género en todas las operaciones y resultados del sistema.

3.2 Categoría: Transparencia y explicabilidad

La transparencia y explicabilidad son pilares esenciales en el uso ético de la inteligencia artificial, exigiendo que cada fase y componente de los sistemas de IA sea accesible y comprensible. Es imprescindible que las personas usuarias, particularmente las personas con discapacidad, comprendan cómo se toman las decisiones y cómo los datos son seleccionados y procesados. Las personas usuarias tienen derecho a recibir información detallada en formatos que sean universalmente

accesibles, asegurando que estén plenamente informadas sobre cómo interactúan con la IA y cómo esta afecta los resultados que les conciernen. Además, la protección y privacidad de los datos no solo es una expectativa, sino un requisito fundamental para preservar la dignidad y los derechos de todos los individuos en el ecosistema de IA. La categoría de Transparencia y Explicabilidad es por tanto fundamental en la ética de los sistemas de IA. Esta categoría se desglosa en dos enfoques principales: datos y explicabilidad.

El enfoque de datos subraya la importancia de la claridad y el acceso a la información referente a los datos que alimentan la IA. Se trata de garantizar que cada paso en el manejo de los datos, desde la recolección hasta su uso final en la toma de decisiones, sea documentado y transparente. Se verifica que los datos reflejen una representación justa y diversa, y que haya medidas en marcha para rastrear los cambios y justificar las modificaciones hechas durante el ciclo de vida del sistema. Este nivel de transparencia es crucial para que las personas con discapacidad comprendan cómo se utilizan sus datos y cómo pueden influir en los resultados que les afectan directamente.

El enfoque de explicabilidad se centra en la capacidad del sistema de IA para ofrecer justificaciones claras y comprensibles de sus procesos y decisiones. Este enfoque asegura que las personas usuarias puedan entender y, si es necesario, cuestionar las decisiones tomadas por el sistema. Es vital que las explicaciones sean accesibles y estén disponibles en formatos que las personas con discapacidad puedan utilizar y comprender. La auditabilidad es otro componente esencial, proporcionando la capacidad para una revisión independiente del sistema y para verificar que las operaciones y los resultados sean justos y no discriminatorios.

Ambos enfoques, en su conjunto, fortalecen la confianza en los sistemas de IA al promover la apertura y facilitar la comprensión y la supervisión de sus operaciones por parte de todas las personas usuarias, incluyendo a las personas con discapacidad. La protección y privacidad de los datos son consideraciones intrínsecas a estos enfoques, asegurando que la información personal sea manejada con el máximo cuidado y respeto a la privacidad individual.

A continuación, se incluyen las preguntas para cada uno de los enfoques:

3.2.1 Categoría Transparencia y Explicabilidad, enfoque de Datos:

- **Documentación y accesibilidad de información de datos:** Esta pregunta evalúa si el sistema de IA ha sido documentado adecuadamente en términos de los datos que utiliza. Se enfoca en si esa documentación es fácilmente accesible para las personas interesadas, detallando la procedencia de los datos y cualquier cambio realizado durante el ciclo de vida del sistema.
- **Inclusión de datos de personas con discapacidad:** Indaga si se han incluido datos específicos de personas con discapacidad durante la fase de entrenamiento del algoritmo, lo que es crucial para desarrollar sistemas de IA inclusivos y que no perpetúen sesgos o exclusiones.
- **Diversidad de categorías de discapacidad en datos:** Se refiere a si se han considerado los diferentes tipos de discapacidad al recopilar datos, asegurando que el sistema de IA sea representativo de la diversidad de la población de personas usuarias.
- **Validación de datos demográficos:** Pregunta si, al utilizar datos demográficos, el sistema de IA valida la representatividad de esos datos en el contexto de uso, posiblemente utilizando información estadística relevante para confirmar su validez.
- **Evaluación y garantía de calidad de datos:** Evalúa si se han implantado medidas para asegurar y evaluar continuamente la calidad de los datos utilizados por el sistema de IA, lo cual es fundamental para la integridad y la fiabilidad del sistema.
- **Trazabilidad de los datos utilizados:** Investiga si es posible rastrear y acceder a la información específica sobre qué datos fueron utilizados por el sistema para llegar a decisiones o recomendaciones particulares, lo que es un componente clave de la transparencia.
- **Protocolo para trazabilidad de datos:** Evalúa si se ha establecido y comunicado un protocolo claro para la trazabilidad de los datos que también apoya la gestión del riesgo asociado con el sistema de IA.
- **Información transparente y accesible para las personas usuarias:** Mide si se proporciona a las personas usuarias información clara y accesible sobre cómo se manejan sus datos y cómo pueden ejercer sus derechos, incluyendo el derecho a ser informadas, el derecho a acceder a sus datos y el derecho a rectificarlos.

3.2.2 Categoría Transparencia y Explicabilidad, enfoque de Explicabilidad:

- **Comprensión de conclusiones y resultados:** Esta pregunta evalúa si el proceso mediante el cual el sistema de IA llega a sus conclusiones y resultados es transparente y si las personas

usuarias finales pueden entender y seguir el razonamiento detrás de las decisiones tomadas por el sistema.

- **Accesibilidad de explicaciones y justificaciones:** Indaga sobre la implantación de procesos y mecanismos que aseguren que cualquier persona usuaria pueda comprender y acceder a las explicaciones y justificaciones de las decisiones tomadas por el sistema de IA, y si pueden solicitar y recibir aclaraciones si es necesario.
- **Opciones de explicaciones adaptadas:** Se refiere a la disponibilidad de explicaciones que pueden ser simplificadas o más técnicas según las necesidades e intereses de la persona usuaria, permitiendo una personalización de la información proporcionada por el sistema de IA.
- **Información sobre la influencia de los datos en decisiones:** Evalúa si se proporciona a las personas usuarias información clara sobre cómo sus datos personales influyen en las decisiones tomadas por el sistema de IA.
- **Prácticas de registro de decisiones:** Mide si se han implantado prácticas adecuadas de registro que documenten de manera detallada y sistemática las decisiones o recomendaciones emitidas por el sistema de IA.
- **Retroalimentación continua de personas usuarias:** Investiga si existe una metodología establecida, como encuestas continuas, para recopilar la opinión de las personas usuarias sobre si entienden las decisiones tomadas por el sistema de IA.
- **Auditabilidad completa del sistema:** Evalúa si se han creado procesos y mecanismos que permitan la auditoría completa del sistema de IA por entidades independientes, incluyendo la capacidad de rastrear el desarrollo del sistema, el manejo de los datos y el registro exhaustivo de todas las operaciones, resultados e impactos del sistema.

3.3 Categoría: Accesibilidad universal

La categoría de Accesibilidad Universal en sistemas de IA es esencial para cerrar la brecha digital y asegurar que todas las personas puedan interactuar efectivamente con la tecnología. Esta accesibilidad no solo mejora la experiencia de la persona usuaria, sino que también es un derecho para las personas con discapacidad. Esta categoría destaca la importancia de incorporar principios de diseño inclusivo y accesibilidad desde el inicio del desarrollo de sistemas de IA. Este enfoque no solo es un imperativo ético sino también una práctica que promueve la inclusión y la igualdad de oportunidades para todas las personas, permitiéndoles beneficiarse plenamente de las ventajas que ofrece la inteligencia artificial.

Esta categoría consta de dos enfoques: interfaz y accesibilidad. El enfoque de Interfaz se centra en la usabilidad y la adaptabilidad de la interfaz de usuario del sistema de IA. Se pone especial atención en cómo los elementos de la interfaz y las funciones interactivas se alinean con las necesidades de las personas usuarias con diferentes tipos de discapacidad, abarcando desde la claridad visual y auditiva hasta la compatibilidad con tecnologías de apoyo. Las preguntas bajo este enfoque buscan evaluar si la interfaz promueve una experiencia de usuario intuitiva y accesible, y si permite personalizaciones que respetan las preferencias individuales y los requisitos de accesibilidad.

El enfoque de Accesibilidad amplía la lente para evaluar la accesibilidad del sistema de IA en su conjunto. Se examina si el sistema cumple con los estándares de accesibilidad, como las Directrices de Accesibilidad al Contenido Web (WCAG), y si se han realizado auditorías independientes para asegurar la conformidad. Las preguntas también abordan la disponibilidad y la accesibilidad de los recursos de capacitación y soporte técnico, asegurando que todas las personas usuarias, incluidas las personas con discapacidad, puedan utilizar el sistema de manera efectiva y autónoma.

A continuación, se incluyen las preguntas para cada uno de los enfoques:

3.3.1 Categoría de Accesibilidad Universal, enfoque de Interfaz:

- **Consideración de diversas discapacidades en diseño de interfaz:** Evalúa si se han tenido en cuenta los diferentes tipos de discapacidad durante la recopilación de datos para el diseño de la interfaz del sistema de IA, garantizando así una experiencia de usuario inclusiva.
- **Compatibilidad con tecnologías de apoyo:** Indaga si el sistema es compatible con tecnologías de apoyo y otras herramientas utilizadas por personas con discapacidad sensorial, física, intelectual o psicosocial y si hay un compromiso de mantener esta compatibilidad.
- **Consideración de necesidades de las personas con discapacidad en la interfaz:** Examina si la interfaz de usuario del sistema de IA tiene en cuenta las necesidades de las personas con discapacidad.
- **Características específicas para facilitar el uso por las personas con discapacidad:** Valora si existen características o herramientas en la interfaz diseñadas específicamente para facilitar su uso por personas con discapacidad.

-
- **Herramientas de personalización de la interfaz:** Considera si el sistema ofrece herramientas que permitan a las personas usuarias personalizar la interfaz y la interacción según sus necesidades individuales de accesibilidad.
 - **Pruebas de interfaz en diversos dispositivos y navegadores:** Evalúa si las interfaces de usuario del sistema se prueban y validan en una variedad de dispositivos y navegadores para asegurar su funcionalidad y robustez.
 - **Alternativas textuales para contenido no textual:** Interroga si el sistema ofrece alternativas textuales para todos los contenidos no textuales, lo cual es crucial para la accesibilidad de personas con discapacidad visual usuarias.
 - **Ajustes de texto y contraste sin tecnología de apoyo:** Pregunta si las personas usuarias pueden ajustar el tamaño del texto y el contraste de la interfaz sin la necesidad de tecnología de apoyo.
 - **Operación de elementos interactivos sin táctil:** Investiga si todos los elementos interactivos del sistema pueden operarse mediante el teclado o alternativas no táctiles, lo que es esencial para personas con discapacidad física.
 - **Alternativas al audio automático:** Evalúa si el sistema incluye alternativas al audio que se reproduce automáticamente, permitiendo a las personas con discapacidad auditiva usuarias una experiencia de usuario inclusiva.
 - **Descripciones de audio para contenido visual:** Valora si el sistema proporciona descripciones de audio para el contenido visual importante, facilitando su uso por personas con discapacidad visual.
 - **Ajuste de velocidad de interacción:** Indaga si los usuarios pueden ajustar la velocidad de interacción con el sistema según sus necesidades, especialmente importante en interfaces que requieren respuesta rápida de la persona usuaria.

3.3.2 Categoría de Accesibilidad Universal, enfoque de Accesibilidad:

- **Garantía de accesibilidad de funcionamiento:** Evalúa en qué medida el sistema de IA asegura que todas sus funciones sean accesibles para personas usuarias con discapacidad, buscando garantizar una experiencia de usuario inclusiva y sin barreras.
- **Verificación de accesibilidad con personas con discapacidad:** Indaga si se ha validado la accesibilidad del sistema mediante pruebas con personas usuarias con discapacidad, lo cual es crucial para una evaluación realista de su accesibilidad.

-
- **Cumplimiento con estándares de accesibilidad:** Determina si el sistema de IA cumple con las Pautas de Accesibilidad para el Contenido Web (WCAG) 2.1 u otros estándares de accesibilidad reconocidos internacionalmente.
 - **Auditoría de accesibilidad por terceros:** Pregunta si se ha realizado una auditoría de accesibilidad independiente para verificar la conformidad del sistema con estándares internacionales de accesibilidad.
 - **Recursos de capacitación y soporte accesibles:** Evalúa si están disponibles recursos de capacitación y soporte técnico accesibles y adecuados para personas con discapacidad usuarias.
 - **No restricción de opciones personales:** Valora si el sistema de IA está diseñado de manera que no restrinja las opciones de estilo de vida, creencias, opiniones, expresiones o experiencias personales de las personas usuarias en ninguna fase de su ciclo de vida.
 - **Consideración de la brecha digital y acceso social:** Examina si se han tenido en cuenta la brecha digital y el acceso social al implantar el sistema de IA, asegurando que sea accesible para un amplio espectro de personas usuarias, independientemente de su contexto socioeconómico o geográfico.
 - **Métodos de prueba de accesibilidad:** Indaga sobre los métodos utilizados para probar la accesibilidad del sistema en diferentes dispositivos y tecnologías de apoyo, lo cual es fundamental para garantizar una accesibilidad universal.
 - **Herramientas de personalización de la interacción:** Pregunta si el sistema ofrece herramientas de personalización que permitan a las personas usuarias adaptar su interacción con el sistema a sus necesidades individuales de accesibilidad.
 - **Procedimientos para mantenimiento de la accesibilidad:** Evalúa si existen procedimientos establecidos para garantizar que la accesibilidad del sistema se mantenga y se optimice con actualizaciones y mejoras tecnológicas.
 - **Compromiso con actualizaciones para mejorar la accesibilidad:** Valora si hay un compromiso por parte de los desarrolladores para realizar actualizaciones periódicas del sistema que mejoren la accesibilidad, basándose en la retroalimentación de las personas usuarias y los cambios en las normativas de accesibilidad.

3.4 Categoría: Unidad y participación

La categoría "Unidad y Participación" es esencial para garantizar que los sistemas de IA se construyan y operen con un enfoque inclusivo centrado en las personas con discapacidad, sus familias y sus organizaciones representativas. Este enfoque no solo

busca la inclusión de estas personas en la tecnología sino también su activa contribución a la misma. Para que la tecnología sirva verdaderamente a la sociedad, debe ser diseñada y evaluada por una comunidad diversa. Esta categoría persigue que los sistemas de IA no sólo se utilicen en beneficio de las personas con discapacidad, sino con su conocimiento y su aprobación activa, fomentando sistemas más justos, equitativos y efectivos para todas las personas.

Esta categoría comprende dos enfoques: impacto y participación. El enfoque de impacto se concentra en entender y medir las repercusiones que los sistemas de IA tienen en la sociedad y en la vida de las personas con discapacidad. Busca asegurar que los avances tecnológicos se alineen con las necesidades y aspiraciones de este grupo social, mejorando su autonomía y calidad de vida. Las preguntas en este enfoque evalúan cómo los sistemas de IA afectan la inclusión social y laboral de las personas con discapacidad y cómo se están realizando evaluaciones de impacto social para informar y guiar el desarrollo de la IA. El enfoque de participación destaca la importancia de la implicación directa de las personas con discapacidad en el proceso de desarrollo de la IA. No solo como personas sujetos de prueba o usuarias finales, sino como miembros activos de los equipos de desarrollo y como voces en la toma de decisiones. Se enfatiza la necesidad de mecanismos que permitan a las personas con discapacidad ofrecer su retroalimentación y sugerencias continuas, asegurando que la IA se desarrolle de una manera que respete y refleje sus experiencias y conocimientos.

A continuación, se incluyen las preguntas para cada uno de los enfoques:

3.4.1 Categoría de Unidad y Participación, enfoque de Impacto:

- **Autonomía y empoderamiento de personas con discapacidad:** Evalúa si el sistema de IA ha sido diseñado para fomentar la independencia y el empoderamiento de las personas con discapacidad, considerando la diversidad de la discapacidad en su funcionamiento y accesibilidad.
- **Evaluaciones de impacto social:** Determina si se realizan evaluaciones de impacto social antes de lanzar o actualizar el sistema, y si los resultados se comparten con todas las partes interesadas para garantizar transparencia y responsabilidad.
- **Inclusión de personas con discapacidad en decisiones clave:** Verifica si existen procedimientos que permitan la participación activa de personas con discapacidad en las decisiones importantes relacionadas con el funcionamiento y evolución del sistema de IA.

-
- **Mecanismos de retroalimentación para personas con discapacidad usuarias:** Evalúa si se han implantado canales efectivos para que las personas con discapacidad usuarias puedan reportar efectos adversos del sistema y contribuir a su mejora continua.
 - **Formación de comunidades de personas con discapacidad usuarias:** Pregunta si se fomenta la creación de comunidades de personas con discapacidad usuarias para compartir experiencias y mejores prácticas relacionadas con el uso del sistema de IA.
 - **Eliminación de barreras y promoción de igualdad de oportunidades:** Analiza si el sistema de IA ha sido diseñado e implantado con el objetivo de eliminar barreras existentes y promover la igualdad de oportunidades para personas con discapacidad.
 - **Cumplimiento de legislación y normativas:** Interroga sobre las medidas tomadas para asegurar que el sistema de IA cumpla con la legislación vigente, como la RGPD, y otras regulaciones y normativas relevantes.
 - **Marco de autonomía y participación humana:** Examina si existe un marco claro que defina los límites de la autonomía del sistema y el nivel de participación humana requerida, equilibrando la automatización con el control humano.
 - **Evaluación de riesgos y mitigación:** Considera si se han evaluado y documentado los riesgos o daños potenciales que podrían surgir de predicciones inexactas o usos no previstos del sistema, junto con planes de mitigación.
 - **Información clara a personas usuarias finales:** Valora si se proporciona a las personas usuarias finales información adecuada y comprensible cuando interactúan con el sistema de IA.
 - **Revocación del consentimiento por parte de las personas usuarias:** Indaga si las personas usuarias tienen la capacidad de revocar su consentimiento en cualquier momento sin afectar negativamente el funcionamiento del sistema.
 - **Prevención de dependencia y manipulación:** Pregunta si se han tomado medidas para evitar que el sistema de IA cree dependencia excesiva, comportamientos adictivos o manipule a las personas usuarias finales.

3.4.2 Categoría de Unidad y Participación, enfoque de Participación:

- **Participación de las personas con discapacidad en desarrollo de IA:** Evalúa si las personas con discapacidad han sido involucradas activamente en el diseño, desarrollo y evaluación del sistema de IA, y si hay evidencias concretas de su participación.

-
- **Procesos para facilitar la participación de las personas con discapacidad:** Indaga si existen procesos y mecanismos diseñados específicamente para facilitar y fomentar la participación activa y la retroalimentación de las personas con discapacidad en el desarrollo y mejora del sistema de IA.
 - **Consultas con organizaciones representativas de las personas con discapacidad y de sus familias:** Pregunta si se han realizado consultas con organizaciones y representantes de personas con discapacidad y de sus familias para intercambiar experiencias y aprendizajes sobre sistemas de IA similares, lo que puede enriquecer el desarrollo del sistema con perspectivas diversas.
 - **Comunicación de las limitaciones del sistema:** Evalúa cómo se han comunicado las limitaciones del sistema de IA a todas las partes implicadas, incluyendo personas usuarias, desarrolladoras y otras partes interesadas, con un nivel de detalle y claridad adecuados.
 - **Comunicación clara sobre la interacción con IA:** Mide si se informa de manera efectiva y clara a las personas usuarias cuando están interactuando con un sistema de IA y no con un humano, lo que es crucial para la transparencia y la confianza de las mismas.
 - **Efectividad y accesibilidad de herramientas de comunicación:** Valora la eficacia y accesibilidad de las herramientas de comunicación disponibles para que las personas usuarias proporcionen retroalimentación o soliciten soporte o asistencia humana.
 - **Capacitación y documentación para las personas con discapacidad:** Considera si se ha proporcionado capacitación o documentación adecuada a las personas con discapacidad sobre cómo utilizar el sistema, lo cual es esencial para garantizar la accesibilidad y la facilidad de uso.
 - **Mecanismos de revisión sobre responsabilidad:** Investiga si se han establecido mecanismos de revisión periódica que impliquen a diferentes agentes relacionados con el desarrollo y uso del sistema de IA, para garantizar la responsabilidad y la rendición de cuentas.
 - **Participación en análisis y mejora:** Examina si se han abordado los puntos ciegos en el análisis del sistema mediante la participación activa con partes interesadas relevantes, incluyendo a personas con discapacidad, para validar hipótesis y discutir las implicaciones del sistema de IA.

3.5 Categoría general: aplicación práctica y responsabilidad

Esta categoría tiene como objetivo evaluar la integración efectiva de los sistemas de inteligencia artificial en entornos reales y su interacción responsable con las personas usuarias. Esta categoría es crucial para asegurar que personas usuarias y

administradoras del sistema posean una comprensión clara del funcionamiento y finalidad del sistema de IA. Además, esta categoría enfatiza la necesidad de que quienes supervisan el sistema sean personas bien informadas y capacitadas para manejar y mitigar los riesgos asociados con su uso. Asimismo, pone de relieve la importancia de comunicar efectivamente las decisiones y limitaciones de la IA a las personas usuarias, fomentando un entorno de responsabilidad compartida y comprensión mutua entre todas las partes interesadas.

A continuación, se incluyen las preguntas para esta categoría:

- **Claridad del proceso de resultados:** Evaluar si el sistema de IA tiene un proceso claramente definido, documentado y reproducible para la generación de resultados, asegurando que los procedimientos puedan ser revisados y duplicados de ser necesario.
- **Conocimiento del propósito y uso:** Determinar si todas las personas usuarias y las partes interesadas están adecuadamente informadas sobre el propósito y uso específico del sistema de IA para garantizar su correcta aplicación y entendimiento.
- **Adaptabilidad a personas usuarias:** Examinar si el sistema de IA se ha adaptado a las capacidades y habilidades de todas las personas usuarias potenciales, especialmente aquellas con necesidades de apoyo, para facilitar una experiencia inclusiva y accesible.
- **Capacitación sobre riesgos y uso:** Valorar si se ha proporcionado una formación adecuada a las personas usuarias y a quienes supervisan el sistema de IA en cuanto a sus riesgos, usos y marco legal relevante.
- **Supervisión ética:** Confirmar la existencia de una junta de revisión de ética o un mecanismo equivalente para monitorear y guiar las prácticas éticas y la responsabilidad asociadas al uso del sistema de IA.
- **Comprensión del sistema por las personas supervisoras:** Evaluar si las personas responsables de supervisar y controlar el sistema de IA poseen una comprensión adecuada de su funcionamiento y las decisiones que este toma.
- **Impacto en la toma de decisiones organizacional:** Investigar cómo se han comunicado las influencias y los efectos de las decisiones del sistema de IA en los procesos de toma de decisiones de la organización.
- **Mecanismos de compensación:** Verificar si se han implantado canales efectivos para la compensación o la reparación en un caso de daños o impactos negativos resultantes del uso del sistema de IA.

-
- **Responsabilidades de uso:** Asegurar que se informa claramente a las personas usuarias sobre sus responsabilidades en el uso del sistema de IA, así como sobre los límites de las decisiones generadas por el mismo.

3.6 Ejemplo de resultados

En la Figura 1 se representan gráficamente los resultados de la aplicación del cuestionario estándar a un ejemplo ficticio de sistema de IA¹². Se representa cada uno de los 8 enfoques así como el eje de “aplicación práctica y responsabilidad”. Esta representación permite identificar las principales áreas de mejora, y eventualmente la comparación entre varios sistemas. A partir de este primer análisis, podrían plantearse acciones concretas de mejora o solicitudes específicas al proveedor (interno o externo) del sistema de IA.

¹² Nótese que los valores asignados se han producido de manera aleatoria

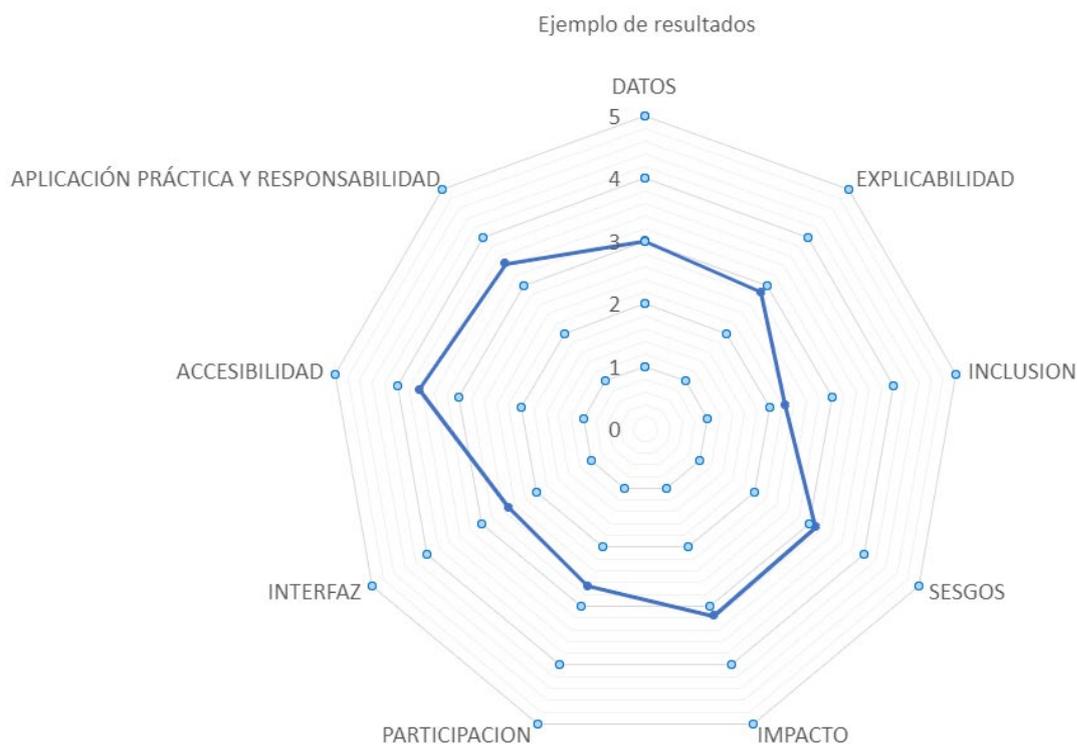


Figura 4. Ejemplo gráfico de resultados del cuestionario estándar.

Descripción figura 4. La imagen representa un gráfico de araña que muestra los resultados de la aplicación del cuestionario estándar a un ejemplo ficticio de sistema de IA.

3.7 Relación con DUDH y CDPD

Dada la relevancia que tiene la Declaración Universal de los Derechos Humanos (DUHD) y la Convención sobre los Derechos de las Personas con Discapacidad (CDPD), se ha realizado un primer ejercicio exploratorio de conexión entre las categorías y enfoques de la guía de análisis y los diferentes artículos de ambos documentos, cuya representación gráfica puede verse en la Figura 5 (el detalle puede encontrarse en los Anexos). Aunque se trata de un ejercicio preliminar, que puede seguir matizándose, su utilidad radica en la posibilidad de detectar algunas cuestiones adicionales para incorporar en el análisis en fases posteriores.

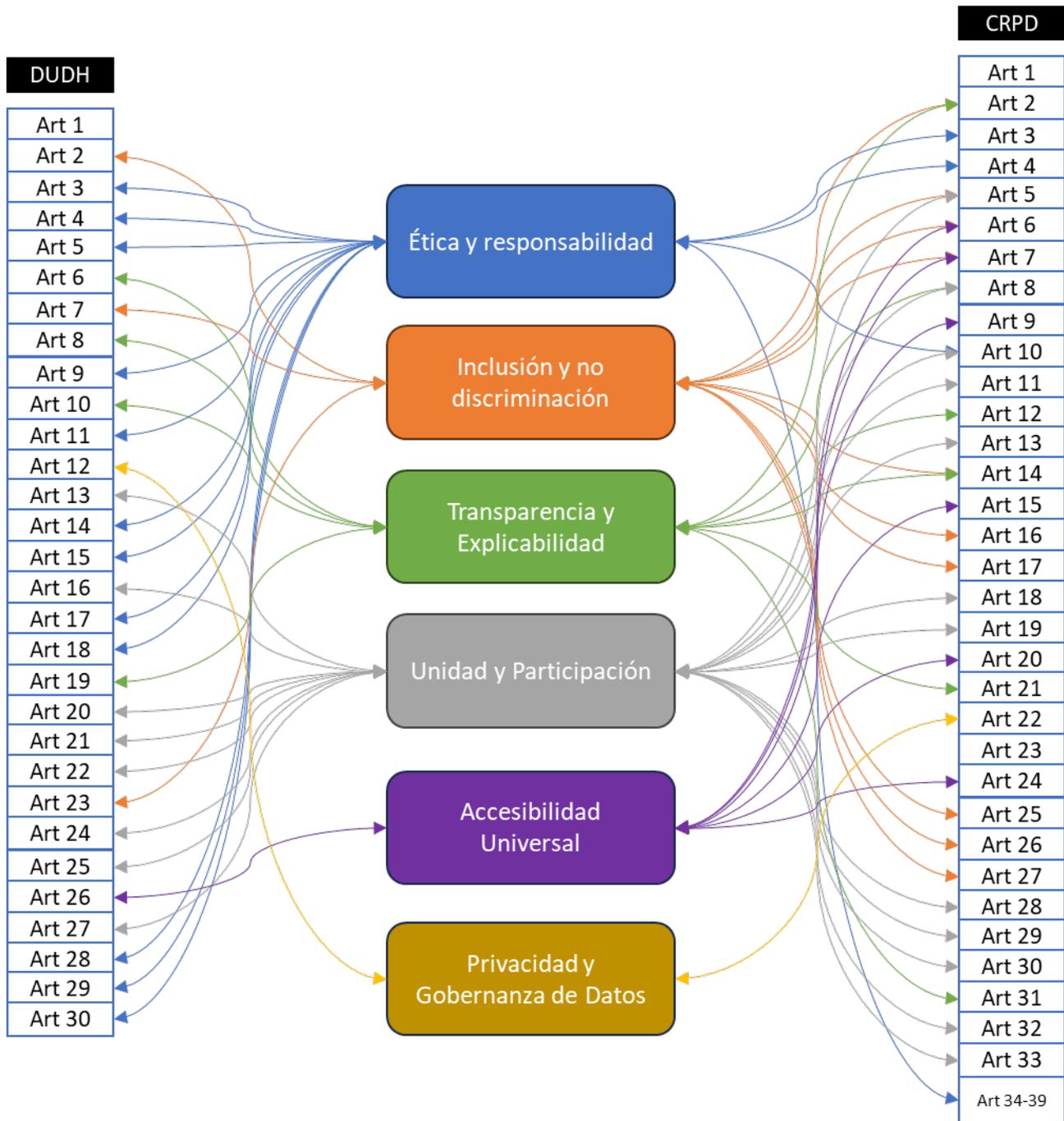


Figura 5. Relación entre categorías, DUDH y CDPD.

Descripción figura 5. La imagen representa gráficamente la conexión entre las categorías y enfoques de la guía de análisis y los diferentes artículos de la Declaración de Derechos Humanos y la Convención sobre los Derechos de las Personas con Discapacidad.

3.8 Guía de análisis simplificada

Hay que considerar que, en algunas ocasiones, puede resultar complicado recabar toda la información para responder el cuestionario estándar, o incluso que parte de dicha información simplemente no esté disponible. De ahí que se ha elaborado un cuestionario simplificado, mediante la selección de determinadas cuestiones relevantes y que abarcan gran parte del objetivo de la guía de análisis.

El cuestionario simplificado se muestra en la Tabla 4. Cada pregunta se valora asimismo del 1 al 5 y se obtiene una puntuación sobre 100 puntos, a partir de la cual se pueden tomar decisiones con respecto al uso del sistema de IA:

Figura 6. Tabla que recoge el cuestionario simplificado.

Cuestionario simplificado

APLICACIÓN PRÁCTICA Y RESPONSABILIDAD

1. ¿El proceso de generación de resultados del sistema de IA es claro, bien documentado y reproducible?
2. ¿Se ha considerado y adaptado el sistema de IA a las capacidades y competencias de todas las personas usuarias potenciales, incluyendo aquellos con necesidades de apoyo?
3. ¿Se han evaluado y comunicado cómo las decisiones del sistema de IA influyen los procesos de toma de decisiones dentro de la organización?
4. ¿Se informa de manera efectiva a las personas usuarias sobre las responsabilidades de uso y los límites de las decisiones generadas por el sistema de IA?

INCLUSIÓN

5. ¿Se ha considerado la diversidad y representatividad de las personas usuarias finales y/o sujetos en los datos utilizados?
6. ¿Se han implantado estrategias para asegurar la igualdad de género en la representación de los datos y en los equipos de desarrollo y evaluación del sistema de IA, asegurando que las decisiones y resultados no introduzcan ni perpetúen desigualdades hacia las mujeres y las niñas con discapacidad?

SESGOS

7. ¿Se ha evaluado si podría haber grupos de personas con discapacidad que podrían verse afectados de manera desproporcionada por los resultados del sistema de IA?
8. ¿Qué medidas de inclusión y equidad ha establecido, como probar los resultados de su sistema de IA para diferentes grupos afectados por tasas de error dispares?
9. ¿Se han desarrollado iniciativas formativas y de concienciación para que las personas diseñadoras y desarrolladoras de IA sean conscientes del sesgo potencial?

DATOS

10. ¿Ha documentado y es accesible la información sobre los tipos de datos, su procedencia y los cambios efectuados durante el ciclo de vida del sistema de IA?

11. ¿Se incluyeron datos específicos de personas con discapacidad durante el entrenamiento del algoritmo?

EXPLICABILIDAD

12. ¿Se han implantado procesos y mecanismos para garantizar que las explicaciones y justificaciones de las decisiones algorítmicas sean comprensibles y accesibles para personas con discapacidad usuarias, incluyendo la posibilidad de cuestionar o solicitar una clarificación sobre dichas decisiones?

INTERFAZ

13. ¿Es el sistema actualmente compatible con tecnologías de apoyo y otras herramientas utilizadas por personas con discapacidad sensorial, física, intelectual o psicosocial, y se compromete a mantener esta compatibilidad?

ACCESIBILIDAD

14. ¿En qué medida garantiza que todas las funciones del sistema sean accesibles para las personas con discapacidad usuarias?

IMPACTO

15. ¿El sistema de IA promueve la autonomía y el empoderamiento de las personas con discapacidad, considerando todos los tipos de discapacidad?

16. ¿Se dispone de procedimientos para incluir a personas con discapacidad en las decisiones clave sobre la funcionamiento y evolución del sistema?

PARTICIPACION

17. ¿Se ha involucrado a las personas con discapacidad en el diseño, desarrollo y/o evaluación del sistema de IA y existen evidencias de ello?

18. ¿Se ha proporcionado formación o documentación a las personas con discapacidad sobre cómo usar el sistema?

PUNTUACIÓN: 54/100

Este cuestionario simplificado resulta más sencillo de aplicar, al haber reducido el número de preguntas, y puede resultar útil para una primera aproximación.

4 Revisión de casos de uso

En la creación de la presente guía de análisis, el objetivo ha sido establecer las preguntas necesarias para entender donde podrían estar las principales alertas o retos respecto al cumplimiento de los principios éticos de CERMI. No obstante, la guía no es un documento estático, sino que es un documento que necesitará actualizarse a medida que se vaya utilizando con diferentes casos de uso, dado que no es posible reflejar todos ellos de manera inicial. Tanto porque la casuística puede variar como porque además la tecnología, en este caso la IA, evoluciona de manera constante.

En esta sección se describen los resultados de una búsqueda exhaustiva de casos de uso, relacionados con empleo, salud y educación. La información hallada se puede agrupar en tres categorías diferentes:

- Informes que muestran ejemplos concretos de aplicación de IA en distintos sectores agrupados por localización geográfica.
- Informes que realizan una agrupación de casos de uso y hacen un estudio de su impacto frente a los derechos humanos.
- Un catálogo de 152 casos de usos a escala mundial agrupados también por distintas categorías.

En el primer grupo de informes son relevantes los estudios de Algorithmwatch¹³ junto con la Fundación BertelsmannStiftung¹⁴, que son organizaciones que normalmente realizan estudios acerca de la aplicación de la IA en diversos sectores y sus posibles implicaciones éticas. En este caso se realiza un análisis en diferentes países de la Unión Europea, incluyendo doce países entre los que se encuentra España. El informe trata de hacer un estudio de la situación actual en referencia a los organismos de supervisión disponible y menciona algunos ejemplos por país, con el objetivo de mostrar modelos reales de implantación del uso de IA. En el caso de España los

¹³ https://algorithmwatch.org/de/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf

¹⁴ <https://www.bertelsmann-stiftung.de/en/home>

ejemplos más relevantes tienen que ver con la salud, la seguridad y la prevención de la violencia:

- **CIBERSAM**¹⁵, para el diagnóstico del trastorno bipolar a partir de neuroimágenes de 3000 pacientes, con una precisión del 65% superior a la humana pero inferior al umbral necesario del 80%, para poder ponerlo en práctica.
- **MEDIKTOR**, una aplicación basada en IA para realizar un prediagnóstico de asistencia médica, desarrollado por una *startup* española, con una precisión alrededor del 91% en estudios con pacientes reales, disponibles en múltiples lenguajes y actualmente con aproximadamente 10 millones de personas usuarias a escala mundial.
- **Veripol**¹⁶, desarrollado para la detección de denuncias falsas en robos violentos con intimidación y hurto basado en el análisis del lenguaje natural. Aunque el sistema parece eficaz, ha recibido diversas críticas respecto a los resultados discriminatorios que podría tener en el análisis del lenguaje en personas extranjeras, con menor dominio del lenguaje, o personas con un menor nivel educativo.
- **SAVRY**¹⁷, que trata de predecir el riesgo de violencia en adolescentes, así como la reincidencia delictiva. Este tipo de aplicaciones no tiene la intención de discriminar, pero estudios previos realizados con aplicaciones similares en Estados Unidos, muestran que puede requerir de revisiones o auditorías necesarias para revisar si existe discriminación.
- **VIOGEN**¹⁸, una aplicación utilizada por el Ministerio del Interior que trata de establecer un riesgo en los casos de violencia de género. Si bien es verdad que presenta un acierto razonable, y que ha contribuido a prevenir muchos casos, también ha sido criticada por multitud de organizaciones en la falta de transparencia del algoritmo, así como en los datos con los que ha sido entrenada.

Estos ejemplos, aun no siendo una categorización exhaustiva, muestran algunos de los posibles riesgos asociados al diseño de algoritmos, así como la necesidad de disponer de más detalles acerca del diseño de éstos. Además, resaltan como están aumentando las peticiones desde todos los ámbitos de la sociedad de la necesidad

¹⁵ <https://www.cibersam.es/noticias/la-combinacion-de-neuroimagen-y-aprendizaje-automatico-nuevas-vias-para-el-diagnostico-del-trastorno-bipolar>

¹⁶ <https://eticasfoundation.org/es/veripol-deteccion-de-denuncias-policiales-falsas/>

¹⁷ <https://bienestaryproteccioninfantil.es/savry-manual-para-la-valoracion-estructurada-de-riesgo-de-violencia-de-jovenes/>

¹⁸ <https://www.interior.gob.es/opencms/es/servicios-al-ciudadano/violencia-contra-la-mujer/sistema-viogen/>

de analizar el impacto en las cuestiones éticas. Asimismo, el informe resalta la necesidad de considerar este tipo de aplicaciones como sistemas sociotécnicos, no como sistemas aislados, que requieren tratar de manera conjunta el análisis y supervisión necesaria de datos y los algoritmos y sistemas donde se implantan. Destaca también la necesidad de formar y preparar tanto a la administración pública como a los ciudadanos para enfrentarse a este tipo de retos, proporcionando tanto la regulación necesaria como el conocimiento adecuado. Por último, destaca la obligación de fortalecer la participación de la sociedad civil e involucrarla de manera adecuada en la supervisión y análisis de este tipo de aplicaciones. Desafortunadamente, el informe no incluye ninguna mención específica a las personas con discapacidad.

En cuanto a los documentos que relacionan los casos de uso con los derechos humanos, el estudio¹⁹ analiza los casos de uso en el ámbito de la justicia, la salud, el acceso a la educación, la gestión del contenido digital y el área de recursos humanos. Como primer paso establece una línea base sobre las implicaciones existentes para los derechos humanos en cualquier ámbito de actividad en el que se vaya a introducir la IA. En un segundo paso trata de identificar si la introducción de la IA ha generado consecuencias negativas, positivas o neutras en los derechos humanos. La Figura 7 muestra gráficamente este análisis.

¹⁹ <https://dash.harvard.edu/handle/1/38021439>. Raso, Filippo, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Kim Levin. 2018. Artificial Intelligence & Human Rights: Opportunities & Risks. Berkman Klein Center for Internet & Society Research Publication.

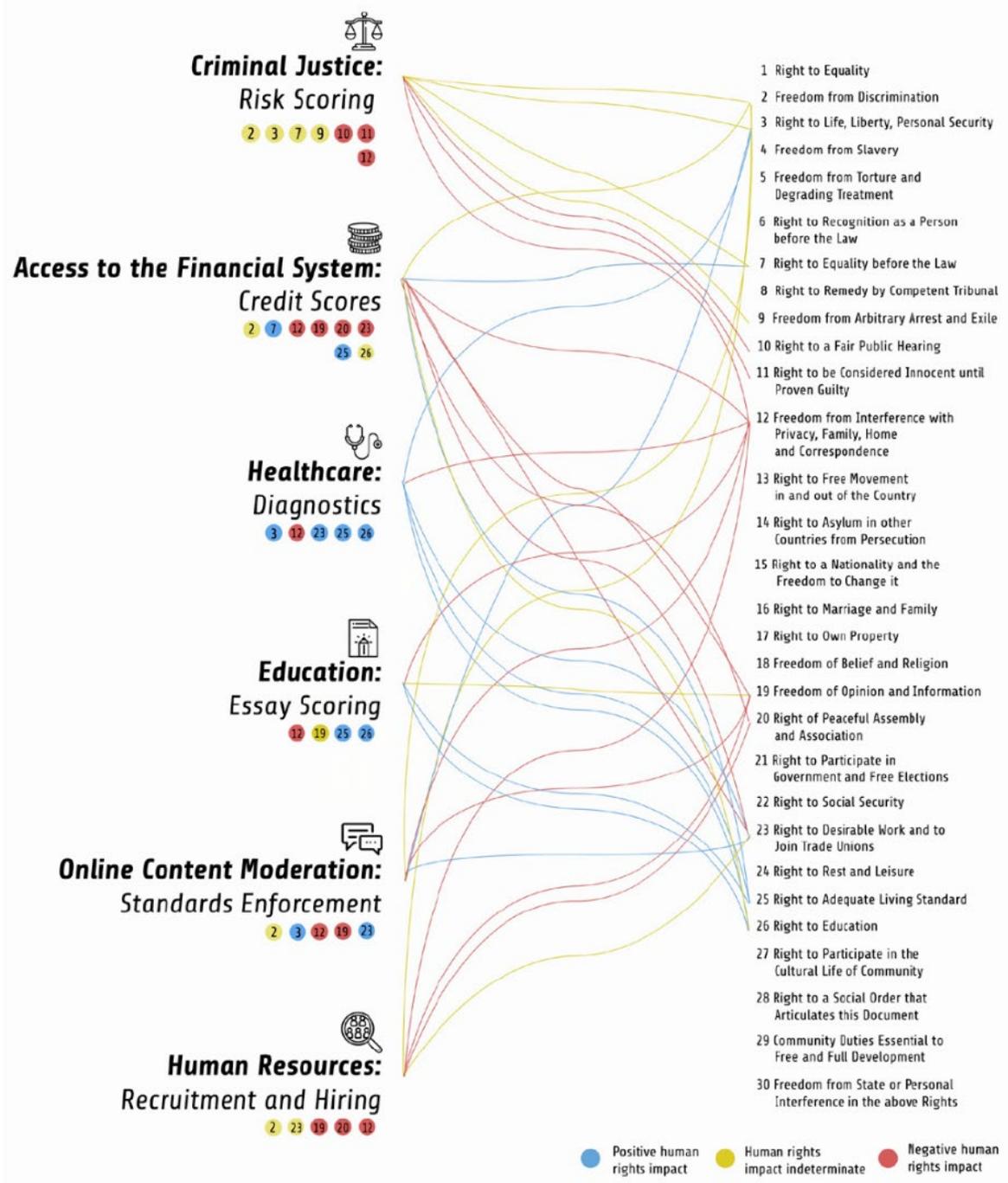


Figura 7. Inteligencia Artificial y Derechos Humanos.

Fuente: Raso, Filippo, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, and Kim Levin. 2018. *Artificial Intelligence & Human Rights: Opportunities & Risks*. Berkman Klein Center for Internet & Society

Descripción figura 7. La imagen representa gráficamente el análisis de la relación entre los sistemas de IA y los derechos humanos en varios casos de uso en el ámbito de la justicia, la salud, el acceso a la educación, la gestión del contenido digital y el área de recursos humanos.

La principal conclusión es que el impacto de la IA en los derechos humanos proviene de tres causas fundamentales. En primer lugar, se refiere a la necesidad de una revisión exhaustiva de la calidad de los datos. En segundo lugar, menciona la necesidad de entender en detalle cómo se ha realizado el diseño del sistema, y si se consideraron los derechos humanos en su diseño. Por último, resalta la complejidad de establecer el impacto en los derechos humanos cuando el sistema se encuentra ya implantado, dado que requiere de sistemas de detección que en la mayoría de los casos no son sencillos de diseñar, por el continuo cambio del entorno. Además, señala que es prometedor que muchas empresas privadas empiecen a utilizar este enfoque basados en los derechos humanos, pero también incide en que no pueden hacerlo solas y que los gobiernos tienen un papel crucial que desempeñar. Por un lado, estableciendo unos criterios aplicables a la calidad de los datos con los que estos sistemas se implantan, por otro lado, estableciendo normativas que motiven a las empresas a garantizar el cumplimiento de los derechos humanos.

Accessnow,²⁰ organismo que participó en la Publicación de la Declaración de Toronto sobre principios éticos de la IA, y que está enfocado a extender y defender los derechos digitales, coincide también en resaltar la importancia del análisis de los sistemas de IA en relación con los derechos humanos. Su informe se centra en mencionar algunos de los derechos humanos que desde su punto de vista se pueden ver más afectados por la implantación de la IA (derecho a la vida, derecho a la igualdad, derecho a la libertad de circulación) y menciona algún ejemplo de caso de uso al respecto. También resalta la importancia de que cualquier enfoque tendrá que ser específico para cada sector y señala que por parte tanto de los organismos públicos como en empresas privadas se debería incluir un análisis de impacto de uso de la IA en los derechos humanos, así como la necesidad de establecer la transparencia, explicabilidad y rendición de cuentas en todos los procesos que usen IA. Por último, subraya la necesidad de continuar investigando en los posibles daños que puede causar la IA en los derechos humanos, así como en la importancia de invertir en medidas concretas para responder a estos riesgos.

²⁰ <https://www.accessnow.org/wp-content/uploads/2018/11/AI-and-Human-Rights.pdf>

Otro informe de la Agencia de los Derechos fundamentales de la Unión Europea²¹, hace hincapié en un enfoque similar respecto al análisis del impacto de la IA y los derechos humanos. La novedad del informe es que se realiza con base en entrevistas a 90 personas de sector público y privado en 5 países diferentes y a 10 personas involucradas en el seguimiento y observación de las posibles violaciones de estos derechos. Las conclusiones son muy similares a las ya mencionadas, respecto a cómo afecta la IA a los derechos humanos. Se resalta la discriminación como uno de los impactos más relevantes de la IA, pero ninguno de ellos menciona la existencia de sistemas que evalúen esta discriminación. Además, este estudio refleja que la mayoría de las personas entrevistadas, fundamentalmente en el sector privado, reconoce que su conocimiento es limitado respecto al resto de derechos humanos que pueden verse afectados por la IA con la excepción del derecho a la privacidad, lo que indica una necesidad clara de reforzar este punto en lo que a las empresas privadas se refiere. Por último, todas las personas entrevistadas coinciden en la importancia de que las decisiones tomadas por la IA sean revisadas por personas.

Respecto a la categorización de casos de uso realizado por la empresa Éticas²², creada en 2012 con el objetivo de estudiar el impacto de la tecnología en los derechos humanos, el estudio muestra un listado de algoritmos que incluyen IA. Es un trabajo de investigación realizado que recopila 152 casos de uso, de los cuales la mayoría (64%) están fuera de Europa y el resto en Europa. Se ha establecido una clasificación, que resulta de utilidad para un análisis exploratorio, identificando el área de la sociedad afectada y el tipo de impacto social. En la figura 8 se representan los casos de uso por área, remarcando la ubicación de los correspondientes a empleo, educación y salud.

²¹ https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_en.pdf

²² <https://eticas.tech/oasi/algorithms#methodology-and-categories>

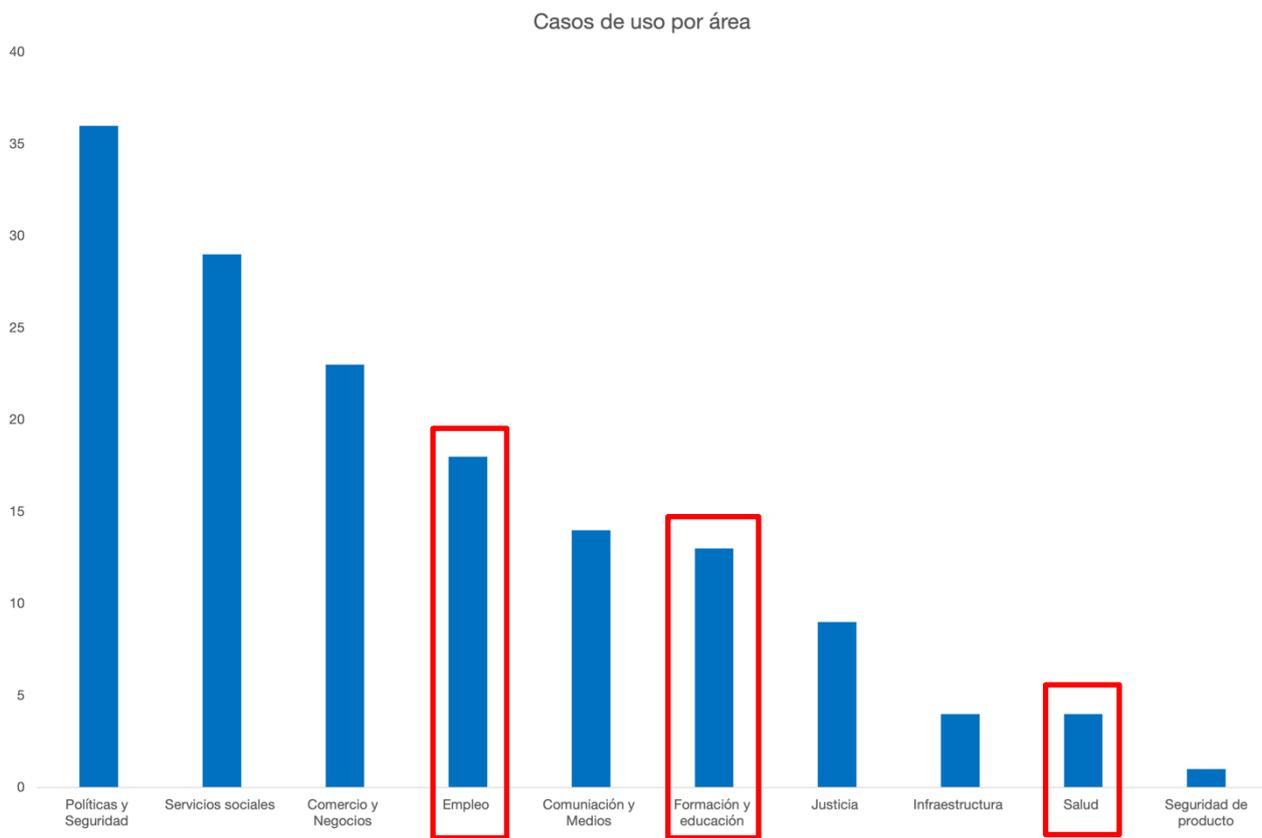


Figura 8. Clasificación de casos de uso por área. Fuente: Éticas

Descripción figura 8. La imagen representa un gráfico de columnas que compara las diferentes áreas de la sociedad y su impacto en los casos de uso estudiados. Se destacan en un recuadro rojo las columnas correspondientes a empleo, educación y salud.

En los casos relacionados con **educación**, el 83% de los casos tienen que ver con la predicción del comportamiento ya sea para alumnado o profesorado, como por ejemplo predecir el riesgo de abandono escolar o la efectividad del profesorado. El resto de los casos están relacionados con automatización de tareas, como por ejemplo un algoritmo de ingreso en la Universidad.

Referente a los casos de **empleo**, más del 70% de los ejemplos se refieren a algoritmos de contratación. HIREVUE es uno de los que se menciona, que consiste en una aplicación de realización de entrevistas mediante video, y su análisis posterior para la contratación. El resto, se enfocan en la evaluación de personas, como por ejemplo para entender la productividad.

Respecto a los tres casos de **salud**, dos están relacionados con la clasificación y perfilado de personas (ej. intentar reducir el tiempo que pasa el paciente en el hospital) y una aplicación está enfocada a intentar reducir el riesgo de suicidio.

Con respecto al análisis de casos de uso relacionados con el tipo de entidad (pública o privada), como se puede ver en la Figura 9, en el caso del sector público se incluyen 86 casos y en el caso del sector privado 56 casos de los 152 registros, fundamentalmente en Estados Unidos.

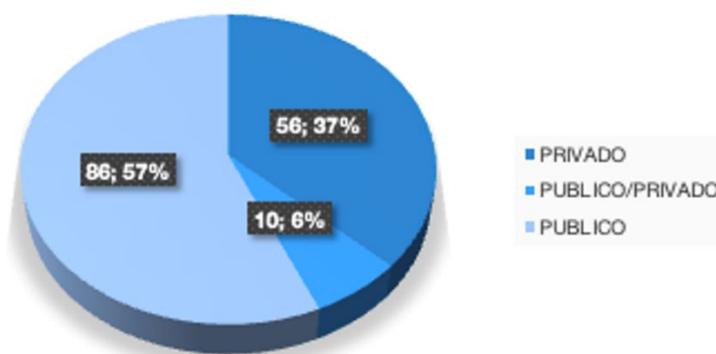


Figura 9. Tipología de casos de uso. Fuente: ETICAS

Descripción figura 9. La imagen representa un gráfico de tarta que describe la relación entre el número de casos de uso según el tipo de entidad, pública o privada

Respecto al análisis de casos de uso en España se incluyen 11 casos de los cuales sólo 2 están relacionados con empleo y no hay ejemplos sobre educación o salud. Respecto al empleo, uno de ellos es un algoritmo del 2012 desarrollado por el SEPE, cuyo objetivo es tratar de automatizar las ayudas a las personas en situación de desempleo. Este algoritmo ha sido objeto de algunas críticas fundamentalmente por el tipo de datos que usa y no se han encontrado datos sobre su estado actual. Otro de ellos, es SEND@²³, implantado desde 2021 un algoritmo que proporciona consejos a las personas en situación de desempleo en su búsqueda de empleo y tipo de

²³ <https://www.expansion.com/economia/2021/07/18/60f4110be5fdea83168b45ab.html>

habilidades necesarias. Según los orientadores laborales la herramienta les permite desarrollar mejor su labor.

A pesar de no haber encontrado información relativa sobre los casos de usos y la discapacidad, la revisión de toda la literatura ha sido muy relevante para refinar el diseño de la guía. Además, el catálogo detallado de ejemplos ha permitido entender multitud de casuística en diferentes países e industrias, así como comprender mejor los casos de uso relacionados con empleo, educación y salud.

5 Recomendaciones para el uso de esta Guía de Análisis

5.1 Recomendaciones para un modelo de Gobernanza

La definición de los principios éticos de la IA es una de las primeras tareas clave que deben realizar las organizaciones que quieren entender el impacto que puede tener la IA en éstas. El diseño de estos principios requiere de una alineación de los distintos departamentos, además de tener un conocimiento detallado de como la IA se utiliza en cada área. Una vez definidos los principios, lo más complicado es su aplicación, es decir, poder llevarlos a toda la organización, así como incluirlos desde el diseño en todos los procesos de la empresa o entidad. Un artículo²⁴ que recopila el estado del arte al respecto y que incluye un estudio llevado a cabo con 22 empresas españolas de distintos sectores, señala las siguientes recomendaciones al respecto:

- Es necesario **establecer una “cultura del dato”** en toda la organización como elemento clave para la aplicación de los principios .
- Es clave establecer la **transparencia y explicabilidad** de los sistemas de IA, con el objetivo fundamental de proporcionar confianza en el uso de ellos.
- Se debe considerar el **ciclo completo del sistema de IA**, no solamente el algoritmo de manera aislada. Desde los requerimientos de su definición hasta su despliegue, pasando por el diseño de los datos y del algoritmo.
- Es necesaria una **formación a toda la organización** en lo que se refiere a la ética y a la IA.
- La **ética** debe estar embebida en los procesos de la empresa o entidad **desde el diseño**.

Una de las claves más importantes para conseguir esa aplicación de los principios es precisamente este último punto, respecto a llevar la ética a los procesos de la empresa o entidad desde el diseño. Esto requiere un proceso de gobernanza para

²⁴ <https://doi.org/10.1007/s00146-021-01267-0>

entender cuáles son los procesos o casos de uso que se están llevando a cabo y como estos casos de uso se alinean con los principios éticos.

Para establecer un gobierno de la IA²⁵ uno de los primeros pasos necesario es establecer un comité de IA ético. Este comité tendrá como objetivo principal asesorar y guiar en el desarrollo y despliegue de la IA en la organización. Una vez tomada la decisión de la creación del comité hay cuatro puntos relevantes que se deben establecer:

- El **propósito del comité**, normalmente enfocado a asegurarse que la toma de decisiones referentes a IA se realiza de manera ética y responsable.
- La **composición del comité** debe incluir una diversidad suficiente entre personas expertas de negocio, de ética, de IA y en la mayoría de los casos es muy recomendable incluir juristas.
- Las **responsabilidades del comité** deben ser acordadas y definidas desde el principio, para acordar su papel respecto al comité de dirección de la organización y entender si será un órgano consultivo o tendrá capacidades ejecutivas propias.
- Los **procedimientos del comité**: su funcionamiento, la periodicidad de las reuniones, aprobación de decisión, metodología a seguir, etc.

Una vez establecido el comité ético, es importante disponer en la organización de personas que puedan actuar como “embajadoras” de esta IA ética, en cada uno de los diferentes departamentos. De esta manera será mucho más sencillo entender los distintos casos de uso o procesos que son llevados a cabo en la empresa de la mano de estas “embajadoras”. Además, el rol de los “embajadoras” debe ser también evangelizar y proporcionar soporte en esta temática a las personas del departamento o departamentos, donde esté asignado. El número de “embajadoras” dependerá del tamaño y la complejidad de la organización.

Para definir el modelo de gobernanza es importante determinar en qué punto del proceso se van a manejar las alertas encontradas en el cumplimiento de los principios éticos. Hay organizaciones que utilizan las guías de análisis o el resultado de

²⁵ <https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee>

auditorías para determinar estas alertas. Es importante establecer esos niveles de alerta en el proceso de gobierno, es decir, que nivel de alerta va a requerir la involucración de las personas “embajadoras”, del comité de IA ético o del comité dirección.

En el caso de CERMI que nos ocupa, nuestra propuesta es que el nivel de alerta venga marcado por la puntuación asignada a los diferentes principios éticos una vez cumplimentada la guía de análisis, explicada en este documento.

En relación, a las personas “embajadoras”, sería importante disponer de las mismas en el resto de las organizaciones con las que trabaja CERMI. Una vez nombradas y formadas convenientemente, nuestra recomendación es que sean éstas quienes completen las preguntas referentes a la guía de análisis.

Durante los primeros meses, es conveniente que el comité asesor de ética de la IA revise todos los casos de uso, antes de aprobar su implantación, porque de esta manera le va a permitir obtener un conocimiento de la tipología de casos de uso, así como establecer una línea base para saber qué tipo de niveles de aprobación respecto a las alertas se debe establecer. Un primer planteamiento podría ser establecer tres niveles de alerta: bajo, medio y alto, que tengan que ver con la puntuación obtenida en cada principio o con una puntuación global del formulario. En el nivel bajo las propias personas “embajadoras” pueden ser quienes aprueben que el caso de uso se puede implantar sin la necesidad de tener la aprobación del comité asesor. En el caso de alerta media sería el comité asesor el que necesita proporcionar la aprobación antes de que la implementación sea llevada a cabo, y por último para las alertas de tipo alto debería ser el comité de dirección.

5.2 Cómo utilizar el cuestionario

Una vez explicado el cuestionario estándar y simplificado a continuación se detalla el proceso que se recomienda seguir a la hora de comenzar a completar el formulario.

Una de las primeras tareas que se debe realizar sería tener nombrada o identificada a la persona que pueda hacer de “embajadora” para la IA ética, como se mencionaba en el punto anterior. Esta persona habrá recibido una formación inicial sobre IA ética

y el formulario y será la persona encargada de dar soporte al resto de personas usuarias de la organización. Para completar los primeros formularios, lo más adecuado, sería que los hiciera de manera conjunta con la persona o las personas encargadas de rellenar el formulario de acuerdo con un caso de uso.

Una vez se disponga del caso de uso que se va a analizar para su puesta en marcha en la organización, es importante contar con el equipo que va a implantar este caso de uso (interno o externo) para poder contestar a las preguntas del formulario. Si la empresa que diseñará el caso de uso es externa es importante compartir el formulario con ellos para que junto con quien actúe como “persona embajadora” de IA ética puedan rellenar las preguntas.

El objetivo sería poder responder a todas las preguntas del **cuestionario estándar**, para poder tener una puntuación total por cada uno de los principios. En su defecto, como mínimo se deberían completar las respuestas del **cuestionario simplificado** y en etapas posteriores del desarrollo del caso de uso volver al formulario para poder completar estas respuestas.

Una vez se disponga de las respuestas se podrá ver de manera gráfica en cuál de los principios se puede tener un nivel de cumplimiento mayor o menor. Según el resultado del formulario la “persona embajadora” presentará este formulario en la comisión ética y allí se decidirá si el proceso puede ser implantado o es necesario diseñar planes adicionales para mejorar la puntuación de alguno de los principios. Por ejemplo, si en la pregunta ¿Se proporciona a las personas usuarias información comprensible como sus datos influyen en las decisiones del sistema?, se obtiene una puntuación baja se podrían definir acciones enfocadas en requerir una mayor explicabilidad al sistema, a los desarrolladores de la solución antes de continuar con la implantación.

Es clave que el comité ético de la IA establezca un sistema de alertas, como se mencionaba en el apartado anterior, para saber qué toma de decisiones respecto a la implantación del caso de uso puede depender de las “embajadoras” y en qué casos esas decisiones tendrán que ser llevadas al comité.

Nuestra recomendación, dado que de manera inicial será complejo establecer ese sistema de alertas y dado que no se tienen ejemplos anteriores, sería llevar todos los

casos de uso al comité e ir aprendiendo de los riesgos que puedan ir presentando. Una vez que se hayan evaluado varios casos de uso y en función de la experiencia adquirida con esa tipología de casos de uso, el comité puede decidir establecer esas alertas, para facilitar que la aprobación o denegación de llevar a cabo los casos de uso sea más ágil.

5.3 Creación de un repositorio común CERMI

En la investigación realizada sobre los casos de uso se han encontrado escasas referencias a la discapacidad, tanto en la clasificación de los casos de uso, como en las referencias en los documentos encontrados. Este punto lleva a la reflexión de la necesidad de impulsar a escala local e internacional la creación de un repositorio común de estos casos de uso. Como paso previo a la creación de ese repositorio internacional, CERMI podría crear su propio repositorio basado en los casos de uso sobre los que se haya realizado el formulario, teniendo en cuenta la diversidad de los grupos con los que trabaja.

5.4 Formación en utilización del formulario

Una de las primeras tareas que CERMI debe llevar a cabo, es compartir el formulario con el resto de las organizaciones miembro. En primer lugar, sería relevante compartir el formulario y explicar en detalle su funcionamiento. En segundo lugar, la definición de la figura de “persona embajadora” podría ser clave a la hora de utilizarla como catalizadora de los principios éticos de la IA en CERMI y el resto de las organizaciones.

6 Bibliografía

A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. <https://zenodo.org/records/3240529>

AI and Human Rights. Access Now. <https://www.accessnow.org/wp-content/uploads/2018/11/AI-and-Human-Rights.pdf>

AI HLEG. (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. European Commission. <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal>

AI System Ethics Self-Assessment Tool. Digital Dubai Office. <https://www.digitaldubai.ae/self-assessment>

Artificial Intelligence & Human Rights: Opportunities & Risks. Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Levin, K. (2018). Berkman Klein Center for Internet & Society Research Publication. <https://dash.harvard.edu/handle/1/38021439>

Artificial Intelligence - Summary. European Union Agency for Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_en.pdf

Australian Government: Department of Industry, Science, Energy and Resources. (2019). Australia's Artificial Intelligence Ethics Framework <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework>

Autoevaluación ética de IA para actores del ecosistema emprendedor: Guía de aplicación. (n.d.). Banco Interamericano de Desarrollo. <https://publications.iadb.org/es/autoevaluacion-etica-de-ia-para-actores-del-ecosistema-emprendedor-guia-de-aplicacion>

Automating Society Report 2019. Algorithm Watch. https://algorithmwatch.org/de/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf

Data Ethics Framework. United Kingdom Government <https://www.gov.uk/government/publications/data-ethics-framework>

Empowering AI leadership

<https://www.weforum.org/publications/empowering-ai-leadership-ai-c-suite-toolkit/>

Ethical Impact Assessment Tool for Recommendation on the Ethics of Artificial Intelligence. (n.d.). UNESCO. <https://www.unesco.org/en/articles/ethical-impact-assessment-tool-recommendation-ethics-artificial-intelligence>

La combinación de neuroimagen y aprendizaje automático: nuevas vías para el diagnóstico del trastorno bipolar. CIBERSAM. <https://www.cibersam.es/noticias/la-combinacion-de-neuroimagen-y-aprendizaje-automatico-nuevas-vias-para-el-diagnostico-del-trastorno-bipolar>

La economía de los cuidados puede crear empleo tras la crisis del covid-19. Expansión <https://www.expansion.com/economia/2021/07/18/60f4110be5fdea83168b45ab.html>

Lovelace, A., & DataKind, U. K. (2020). Examining the black box: Tools for assessing algorithmic systems. Ada Lovelace Institute. <https://ico.org.uk/media/about-theico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf>

OASI: AI Ethics Assessment. Eticas Foundation. <https://eticas.tech/oasi/algorithms#methodology-and-categories>

Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study <https://link.springer.com/article/10.1007/s00146-021-01267-0>

Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Levin, K. (2018). Artificial Intelligence & Human Rights: Opportunities & Risks. Berkman Klein Center for Internet & Society Research Publication. <https://dash.harvard.edu/handle/1/38021439>

SAVRY: Manual para la valoración estructurada de riesgo de violencia de jóvenes. FAPMI-ECPAT España. <https://bienestaryproteccioninfantil.es/savry-manual-para-la-valoracion-estructurada-de-riesgo-de-violencia-de-jovenes/>

Sistema VioGén. Ministerio del Interior, España. <https://www.interior.gob.es/opencms/es/servicios-al-ciudadano/violencia-contra-la-mujer/sistema-viogen/>

Veripol: Detección de denuncias policiales falsas. Eticas Foundation <https://eticasfoundation.org/es/veripol-deteccion-de-denuncias-policiales-falsas/>

Why You Need an AI Ethics Committee. Harvard Business Review. <https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee>

7 Anexos

7.1 Cuestionario completo

APLICACIÓN PRÁCTICA Y RESPONSABILIDAD	
1	¿El proceso de generación de resultados del sistema de IA es claro, bien documentado y reproducible?
2	¿Están todas las personas usuarias y partes implicadas informadas del propósito y uso del sistema de IA?
3	¿Se ha considerado y adaptado el sistema de IA a las capacidades y competencias de todas las personas usuarias potenciales, incluyendo aquellos con necesidades de apoyo?
4	¿Se ha formado adecuadamente a las personas usuarias y controladoras del sistema de IA sobre sus riesgos, uso y marco legal aplicable?
5	¿Existe una junta de revisión de ética o un mecanismo similar para supervisar las prácticas de responsabilidad y ética del sistema de IA?
6	¿Las personas encargadas de supervisar y controlar el sistema de IA tienen una comprensión adecuada de su funcionamiento y toma de decisiones?
7	¿Se han evaluado y comunicado cómo las decisiones del sistema de IA influyen los procesos de toma de decisiones dentro de la organización?
8	¿Se han establecido canales para la compensación o reparación en caso de daños o impactos adversos causados por el sistema de IA?
9	¿Se informa de manera efectiva a las personas usuarias sobre las responsabilidades de uso y los límites de las decisiones generadas por el sistema de IA?
INCLUSIÓN	
10	¿Se ha considerado la diversidad y representatividad de las personas usuarias finales y/o sujetos en los datos utilizados?
11	¿Se realizaron pruebas para grupos de personas con discapacidad específicos o casos de uso problemáticos para asegurar la inclusión?
12	¿Se ha evaluado si el sistema de IA es adaptable a un amplio espectro de preferencias y habilidades individuales, incluyendo personas usuarias que requieren tecnologías de apoyo?
13	¿El sistema de IA ha sido probado y validado con un grupo diverso que incluye personas con discapacidad?
14	¿Se evaluó si el equipo involucrado en la construcción del sistema de IA interactuó con posibles personas usuarias finales o sujetos para validar la inclusión?
15	¿Se ha consultado con los posibles grupos de personas con discapacidad afectadas, para definir correctamente aspectos de inclusión, justicia y equidad?
16	¿Cómo evaluaría la inclusión de personas con discapacidad en el equipo de desarrollo e implantación del sistema de IA para asegurar que sus perspectivas y necesidades informen y enriquezcan el diseño y la evaluación del algoritmo?
17	¿Se han implantado estrategias para asegurar la igualdad de género en la representación de los datos y en los equipos de desarrollo y evaluación del sistema de IA, garantizando que las mujeres y las niñas con discapacidad estén adecuadamente representadas y sus experiencias y necesidades sean consideradas?

SESGOS	
18	¿Se ha evaluado si podría haber grupos de personas con discapacidad que podrían verse afectados de manera desproporcionada por los resultados del sistema de IA?
19	¿Se ha implantado un mecanismo que permita señalar problemas relacionados con prejuicios, discriminación o mal desempeño del sistema de IA?
20	¿Se ha asegurado que su sistema de IA no utiliza variables o “proxies” que pueden ser injustamente discriminatorias?
21	¿Qué medidas de inclusión y equidad ha establecido, como probar los resultados de su sistema de IA para diferentes grupos afectados por tasas de error dispares?
22	¿Se han implantado y evaluado procedimientos para identificar y mitigar posibles sesgos en su sistema de IA, a fin de prevenir la creación o el refuerzo de discriminaciones o prejuicios?
23	¿Se ha establecido un proceso de revisión y monitorización continua para detectar y corregir sesgos y discriminación a lo largo de todo el ciclo de vida del sistema de IA?
24	¿Se han desarrollado iniciativas formativas y de concienciación para que quienes diseñan y desarrollan la IA sean conscientes del sesgo potencial?
25	¿Cómo calificaría la efectividad de las medidas implantadas por el sistema de IA para identificar y mitigar sesgos de género, asegurando que las decisiones y resultados no introduzcan ni perpetúen desigualdades hacia las mujeres y las niñas con discapacidad?

DATOS	
26	¿Ha documentado y es accesible la información sobre los tipos de datos, su procedencia y los cambios efectuados durante el ciclo de vida del sistema de IA?
27	¿Se incluyeron datos específicos de personas con discapacidad durante el entrenamiento del algoritmo?
28	¿Fueron consideradas las diferentes categorías de discapacidad (física, sensorial, intelectual, psicosocial) en la recolección de datos?
29	En el caso de utilizar datos demográficos, ¿ha validado su representatividad teniendo en cuenta el contexto más cercano a través de la información muestral o poblacional de una organización de estadística?
30	¿Se han implementado medidas para evaluar y garantizar continuamente la calidad de los datos de entrada en el sistema de IA?
31	¿Es posible rastrear y acceder a la información de qué datos fueron utilizados por el sistema de IA para tomar decisiones o recomendaciones específicas?
32	¿Ha establecido y divulgado un protocolo para la trazabilidad de datos que apoye la gestión del riesgo asociado con el sistema de IA?
33	¿Se proporciona información transparente y accesible a las personas usuarias sobre la gestión de sus datos y cómo pueden ejercer sus derechos?

EXPLICABILIDAD	
34	¿Se puede detallar y entender cómo el sistema de IA llega a sus conclusiones y resultados, y se explican claramente las decisiones tomadas por el sistema de IA a las personas usuarias finales?
35	¿Se han implantado procesos y mecanismos para garantizar que las explicaciones y justificaciones de las decisiones algorítmicas sean comprensibles y accesibles para personas con

	discapacidad usuarias, incluyendo la posibilidad de cuestionar o solicitar una clarificación sobre dichas decisiones?
36	¿Existen opciones para recibir explicaciones simplificadas o técnicas según el interés y capacidad del usuario?
37	¿Se proporciona a las personas usuarias información comprensible sobre cómo sus datos influyen en las decisiones del sistema?
38	¿Se han implantado prácticas adecuadas de registro para documentar las decisiones o recomendaciones del sistema de IA?
39	¿Se realiza una encuesta continua a las personas usuarias para determinar si comprenden las decisiones tomadas por el sistema de IA?
40	¿Se han establecido procesos y mecanismos que aseguren la auditabilidad completa del sistema de IA por entidades independientes, incluyendo la trazabilidad del desarrollo, la adquisición de datos y el registro detallado de operaciones, resultados e impactos?
INTERFAZ	
41	¿Se han considerado los diferentes tipos de discapacidad en la recolección de datos para el diseño de la interfaz?
42	¿Es el sistema actualmente compatible con tecnologías de apoyo y otras herramientas utilizadas por personas con discapacidad sensorial, física, intelectual o psicosocial, y se compromete a mantener esta compatibilidad?
43	¿Considera las necesidades de las personas con discapacidad el diseño de la interfaz de usuario?
44	¿Existen características o herramientas en la interfaz diseñadas específicamente para facilitar el uso por parte de personas con discapacidad?
45	¿Se ofrecen herramientas de personalización para ajustar la interfaz y la interacción según las necesidades individuales de accesibilidad?
46	¿Se prueban y validan las interfaces de usuario en una variedad de dispositivos y navegadores para garantizar su robustez?
47	¿El sistema ofrece alternativas textuales para todo el contenido no textual?
48	¿Puede la persona usuaria ajustar el tamaño del texto y el contraste sin tecnología de apoyo?
49	¿Pueden todos los elementos interactivos operarse mediante el teclado o alternativas no táctiles?
50	¿Incluye el sistema alternativas para controlar el audio que se reproduce automáticamente?
51	¿Proporciona el sistema descripciones de audio para contenido visual importante?
52	¿Pueden las personas usuarias ajustar la velocidad de interacción según sus necesidades, especialmente en interfaces que requieren respuesta del usuario/a?
ACCESIBILIDAD	
53	¿En qué medida garantiza que todas las funciones del sistema sean accesibles para las personas con discapacidad usuarias?
54	¿Se ha verificado la accesibilidad del sistema mediante pruebas con personas con discapacidad usuarias?
55	¿Cumple el sistema de IA con las pautas de accesibilidad WCAG 2.1 u otros estándares de accesibilidad reconocidos internacionalmente?

56	¿Se ha llevado a cabo una auditoría de accesibilidad por un tercero para verificar la conformidad con estándares internacionales?
57	¿Están disponibles recursos de capacitación y soporte técnico accesibles para personas con discapacidad usuarias?
58	¿Puede asegurarse de que el sistema de IA no restringe las opciones de estilo de vida, creencias, opiniones, expresiones o experiencias personales en ninguna fase de su ciclo de vida?
59	¿Se han considerado la brecha digital y el acceso social al emplear e implantar el sistema de IA?
60	¿Se utilizan métodos para probar la accesibilidad del sistema en diferentes dispositivos y tecnologías de apoyo?
61	¿El sistema ofrece herramientas de personalización para adaptar la interacción con el sistema a las necesidades individuales de accesibilidad?
62	¿Existen procedimientos para garantizar la accesibilidad del sistema ante actualizaciones y mejoras tecnológicas?
63	¿Existe un compromiso para realizar actualizaciones periódicas del sistema para mejorar la accesibilidad basándose en la retroalimentación de las personas usuarias y los cambios en las normativas de accesibilidad?
IMPACTO	
64	¿El sistema de IA promueve la autonomía y el empoderamiento de las personas con discapacidad, considerando todos los tipos de discapacidad?
65	¿Se realizan evaluaciones de impacto social antes del lanzamiento o actualización del sistema y se comparten los resultados con las partes interesadas?
66	¿Se dispone de procedimientos para incluir a personas con discapacidad en las decisiones clave sobre el funcionamiento y evolución del sistema?
67	¿Se han implantado mecanismos de retroalimentación para que las personas con discapacidad usuarias puedan reportar efectos adversos y contribuir a la mejora continua del sistema?
68	¿Se promueve la formación de comunidades en las que participen personas con discapacidad usuarias para compartir experiencias y mejores prácticas relacionadas con el sistema?
69	¿Se ha diseñado e implantado el sistema de IA para eliminar barreras existentes y medir su impacto en la promoción de igualdad de oportunidades para personas con discapacidad?
70	¿Ha tomado medidas para facilitar la demostración del cumplimiento de la legislación vigente incluyendo, por ejemplo, la Regulación General de Protección de Datos (RGPD) u otras regulaciones y normativas específicas relacionadas con la IA?
71	¿Existe un marco claro para los límites de autonomía del sistema y el nivel de participación humana requerida?
72	¿Ha evaluado los posibles riesgos o daños que podrían surgir de predicciones inexactas o usos no previstos del sistema de IA y documentado planes de mitigación?
73	¿Se informa a las personas usuarias finales de manera adecuada y comprensible cuando están interactuando con un sistema de IA?
74	¿Pueden las personas usuarias revocar su consentimiento en cualquier momento sin comprometer el funcionamiento del sistema?
75	¿Ha tomado medidas para evitar la creación de dependencia excesiva, comportamientos adictivos o manipulación de las personas usuarias finales?

PARTICIPACIÓN

76	¿Se ha involucrado a las personas con discapacidad en el diseño, desarrollo y/o evaluación del sistema de IA y existen evidencias de ello?
77	¿Existen procesos y mecanismos para facilitar y fomentar la participación activa y retroalimentación de personas con discapacidad en el desarrollo y mejoras futuras del sistema de IA?
78	¿Se ha llevado a cabo una consulta con organizaciones representativas de personas con discapacidad y de sus familias para compartir experiencias y aprendizajes sobre sistemas de IA similares?
79	¿Con qué nivel de detalle y claridad se han comunicado las limitaciones del sistema de IA a todas las partes implicadas?
80	¿Se comunica de manera efectiva y clara a las personas usuarias cuándo están interactuando con un sistema de IA y no con una persona humana?
81	¿Son universalmente accesibles las herramientas de comunicación para que las personas usuarias proporcionen retroalimentación o soliciten soporte o asistencia humana?
82	¿Se ha proporcionado capacitación o documentación a las personas con discapacidad sobre cómo usar el sistema?
83	¿Se han establecido mecanismos de revisión periódica sobre responsabilidad y rendición de cuentas con los diferentes agentes involucrados en el desarrollo y uso del sistema de IA?
84	¿Se han abordado los puntos ciegos del análisis mediante la participación con las partes interesadas relevantes, por ejemplo comprobando las hipótesis y discutiendo las implicaciones de su sistema de IA con los grupos de personas usuarias afectadas, en concreto personas con discapacidad?

7.2 Aproximación a la Declaración Universal de los Derechos Humanos (DUDH) y a la Convención sobre los Derechos de las Personas con Discapacidad (CDPD)

A continuación, se incluye el detalle de un ejercicio estimando la relación entre las diferentes cuestiones cubiertas en la guía de análisis y los artículos de la Declaración Universal de los Derechos Humanos (DUDH) y la Convención sobre los Derechos de las Personas con Discapacidad (CDPD). Es una aproximación, no exhaustiva, sujeta a revisión en base a la interpretación de algunas de las cuestiones y artículos, así como pendiente de la evolución de la guía de análisis.

Figura 10. Tabla que describe la interrelación entre la Declaración Universal y la Convención.

Declaración Universal	Categoría/Enfoque	Convención
Art. 1	Todas las categorías	Art. 1, Art. 3, Art. 4
Art. 3	Responsabilidad	Art. 10
Art. 4, Art. 5, Art. 9, Art. 11, Art. 14, Art. 15, Art. 17, Art. 18, Art. 28, Art. 29, Art. 30	Ética y Responsabilidad	Art. 34 - 39
Art. 2, Art. 7, Art. 23	Inclusión y No Discriminación	Art. 2, Art. 5, Art. 6, Art. 7, Art. 14, Art. 16, Art. 17, Art. 25, Art. 26, Art. 27
Art. 6, Art. 8, Art. 10, Art. 19	Transparencia y Explicabilidad	Art. 2, Art. 8, Art. 12, Art. 14, Art. 21, Art. 31
Art. 13, Art. 16, Art. 20, Art. 21, Art. 22, Art. 24, Art. 25, Art. 27	Unidad y Participación	Art. 5, Art. 8, Art. 10, Art. 11, Art. 13, Art. 18, Art. 19, Art. 28, Art. 29, Art. 30, Art. 32, Art. 33
Art. 26	Accesibilidad Universal	Art. 6, Art. 7, Art. 9, Art. 15, Art. 20, Art. 24
Art. 12	Privacidad y Gobernanza de Datos	Art. 22

Figura 11. Tabla que describe la interrelación entre la Declaración Universal y las categorías/enfoque establecido en la guía

Artículo	Descripción Breve	Categoría/Enfoque Relacionado	Explicación Breve
Art. 1	Todos los seres humanos nacen libres e iguales en dignidad y derechos	Todas las categorías	Refleja la importancia de la equidad y la igualdad en todas las aplicaciones de IA.
Art. 2	Derecho a la libertad de discriminación	Inclusión y No Discriminación	Asegura que los sistemas de IA promuevan la inclusión y no discriminen a ningún grupo.
Art. 3	Derecho a la vida, la libertad y la seguridad de la persona	Responsabilidad	Transparencia y Explicabilidad
Art. 4	Prohibición de la esclavitud y la servidumbre	Ética y Responsabilidad	Destaca la prohibición de cualquier forma de explotación a través de la IA.
Art. 5	Prohibición de la tortura y tratos inhumanos o degradantes	Ética y Responsabilidad	Asegura que los sistemas de IA no causen daño y respeten la dignidad humana.
Art. 6	Derecho al reconocimiento como persona ante la ley	Transparencia y Explicabilidad	Garantiza que los sistemas de IA reconozcan a todos los usuarios de manera justa y equitativa.
Art. 7	Igualdad ante la ley sin discriminación	Inclusión y No Discriminación	Promueve la igualdad y lucha contra la discriminación en todos los sistemas de IA.
Art. 8	Derecho a un recurso efectivo ante tribunales	Transparencia y Explicabilidad	Asegura la posibilidad de buscar remedios efectivos en caso de mal uso de IA.
Art. 9	Prohibición de detención o destierro arbitrarios	Ética y Responsabilidad	Enfatiza la protección contra la detención arbitraria por parte de la IA.
Art. 10	Derecho a una audiencia justa	Transparencia y Explicabilidad	Asegura la transparencia y la imparcialidad en los procesos judiciales apoyados por IA.
Art. 11	Presunción de inocencia y defensa en juicio	Ética y Responsabilidad	Destaca la importancia de la presunción de inocencia en sistemas de IA relacionados con la justicia.
Art. 12	Protección de la privacidad y la familia	Privacidad y Gobernanza de Datos	Asegura la protección de la privacidad y los datos personales en la IA.
Art. 13	Libertad de movimiento y residencia	Unidad y Participación	Subraya la libertad de movimiento
Art. 14	Derecho de asilo contra persecuciones	Ética y Responsabilidad	Resalta la relevancia de sistemas de IA justos para la gestión de solicitudes de asilo.
Art. 15	Derecho a una nacionalidad y a cambiarla	Ética y Responsabilidad	Refuerza la importancia de la nacionalidad en sistemas de IA para identificación y registro civil.
Art. 16	Derechos iguales en el matrimonio	Unidad y Participación	Promueve la igualdad de derechos en sistemas de IA relacionados con el estado civil.
Art. 17	Derecho a la propiedad	Ética y Responsabilidad	Asegura que la IA respete la propiedad privada y no favorezca la expropiación injusta.
Art. 18	Libertad de pensamiento, conciencia y religión	Ética y Responsabilidad	Garantiza la libertad de conciencia y la no manipulación a través de la IA.
Art. 19	Libertad de opinión y expresión	Transparencia y Explicabilidad	Asegura la libertad de expresión
Art. 20	Libertad de reunión y asociación	Unidad y Participación	Promueve la capacidad de asociarse y reunirse libremente
Art. 21	Derecho a participar en el gobierno y elecciones	Unidad y Participación	Fomenta la participación democrática mediante sistemas de IA en procesos electorales.
Art. 22	Derecho a la seguridad social y a los derechos económicos	Unidad y Participación	Resalta la importancia de la seguridad social en sistemas de IA para la asistencia social.

Art. 23	Derecho al trabajo y protección contra el desempleo	Inclusión y No Discriminación	Asegura que los sistemas de IA promuevan empleos justos y equitativos.
Art. 24	Derecho al descanso y al ocio	Unidad y Participación	Enfatiza el derecho al ocio y al descanso en el contexto de la IA laboral.
Art. 25	Derecho a un nivel de vida adecuado	Unidad y Participación	Destaca el uso de IA para mejorar el nivel de vida y el acceso a servicios básicos.
Art. 26	Derecho a la educación	Accesibilidad Universal	Enfatiza la importancia de la educación accesible apoyada por sistemas de IA.
Art. 27	Derecho a participar en la vida cultural	Unidad y Participación	Fomenta la inclusión cultural y científica a través de la IA.
Art. 28	Derecho a un orden social e internacional	Ética y Responsabilidad	Subraya la necesidad de un marco ético global para los sistemas de IA.
Art. 29	Deberes a la comunidad y limitaciones en el ejercicio de los derechos	Ética y Responsabilidad	Implica un uso de la IA que considere el bienestar comunitario y los deberes éticos.
Art. 30	Prohibición de suprimir los derechos de los demás	Ética y Responsabilidad	Prohíbe el uso indebido de sistemas de IA que supriman los derechos humanos de otros.

Figura 12. Tabla que describe la interrelación entre la Convención y las categorías/enfoque establecidos en la guía

Artículo	Descripción Breve	Categoría/Enfoque Relacionado	Explicación Breve
Art. 1	Propósito	Todas las categorías	Refleja la necesidad de considerar la inclusión y los derechos de las personas con discapacidad en todas las etapas y aspectos de los sistemas de IA.
Art. 2	Definiciones	Transparencia y Explicabilidad; Inclusión y No Discriminación	Asegura que los términos y el uso de IA sean claros y que no excluyan a las personas con discapacidad.
Art. 3	Principios generales	Todas las categorías	Principios fundamentales que deben guiar todas las acciones relacionadas con la IA.
Art. 4	Obligaciones generales	Todas las categorías	Responsabilidades que deben ser adoptadas al implantar sistemas de IA.
Art. 5	Igualdad y no discriminación	Inclusión y No Discriminación; Unidad y Participación	Fomenta prácticas de IA que no discriminen y que promuevan la unidad y participación activa.
Art. 6	Mujeres con discapacidad	Inclusión y No Discriminación; Accesibilidad Universal	Aborda la doble discriminación y la necesidad de accesibilidad en IA.
Art. 7	Niños y niñas con discapacidad	Inclusión y No Discriminación; Accesibilidad Universal	Considera los derechos de la infancia con discapacidad en la IA.
Art. 8	Toma de conciencia	Unidad y Participación; Transparencia y Explicabilidad	Enfatiza la importancia de la toma de conciencia sobre la IA y su impacto.
Art. 9	Accesibilidad	Accesibilidad Universal	Centra la atención en la necesidad de que los sistemas de IA sean completamente accesibles.
Art. 10	Derecho a la vida	Responsabilidad; Unidad y Participación	Destaca la importancia de proteger la vida a través de sistemas de IA inclusivos.
Art. 11	Situaciones de riesgo y emergencias humanitarias	Unidad y Participación	Reconoce la necesidad de inclusión en situaciones de emergencia.
Art. 12	Igual reconocimiento como persona ante la ley	Transparencia y Explicabilidad	Asegura que los sistemas de IA respeten la igualdad de las personas con discapacidad ante la ley.
Art. 13	Acceso a la justicia	Unidad y Participación	Promueve el acceso igualitario a la justicia en contextos de IA.
Art. 14	Libertad y seguridad de la persona	Transparencia y Explicabilidad; Inclusión y No Discriminación	Salvaguarda la libertad y la seguridad personal en el uso de IA.
Art. 15	Protección contra la tortura	Interfaz	Aboga por interfaces de IA que no causen daño ni sufrimiento.
Art. 16	Protección contra la explotación, la violencia y el abuso	Inclusión y No Discriminación	Busca proteger contra el abuso en entornos de IA.
Art. 17	Protección de la integridad personal	Inclusión y No Discriminación	Protege la integridad y dignidad en el uso de IA.
Art. 18	Libertad de desplazamiento y nacionalidad	Unidad y Participación	Fomenta la libertad de movimiento y la autonomía con el apoyo de IA.

Art. 19	Derecho a vivir de forma independiente y ser incluido en la comunidad	Unidad y Participación	Promueve sistemas de IA que apoyen la vida independiente y la inclusión comunitaria.
Art. 20	Movilidad personal	Accesibilidad Universal	Se enfoca en la accesibilidad de la IA para facilitar la movilidad personal.
Art. 21	Libertad de expresión y de opinión, y acceso a la información	Transparencia y Explicabilidad	Garantiza la transparencia y el acceso a la información en sistemas de IA.
Art. 22	Respeto de la privacidad	Privacidad y Gobernanza de Datos	Asegura la privacidad y la protección de datos en IA.
Art. 23	Respeto del hogar y de la familia	Unidad y Participación	Valora la importancia de la familia y el hogar en el desarrollo de la IA.
Art. 24	Educación	Accesibilidad Universal	Aboga por sistemas de IA que promuevan la educación inclusiva.
Art. 25	Salud	Inclusión y No Discriminación	Promueve el uso de IA para mejorar la salud y el bienestar de las personas con discapacidad.
Art. 26	Habilitación y rehabilitación	Inclusión y No Discriminación	Enfatiza la IA como una herramienta para la rehabilitación y habilitación.
Art. 27	Trabajo y empleo	Inclusión y No Discriminación	Busca asegurar la igualdad en el empleo y el trabajo mediante la IA.
Art. 28	Nivel de vida adecuado y protección social	Unidad y Participación	Se centra en mejorar la vida y la protección social a través de la IA.
Art. 29	Participación en la vida política y pública	Unidad y Participación	Fomenta la participación política y pública con el apoyo de IA.
Art. 30	Participación en la vida cultural, actividades recreativas, el esparcimiento y el deporte	Unidad y Participación	Promueve la inclusión en actividades culturales y recreativas mediante IA.
Art. 31	Recopilación de datos y estadísticas	Transparencia y Explicabilidad	Refuerza la necesidad de transparencia en la recopilación y uso de datos.
Art. 32	Cooperación internacional	Unidad y Participación	Fomenta la colaboración internacional en el desarrollo de IA inclusiva.
Art. 33	Aplicación y seguimiento nacionales	Unidad y Participación	Asegura el seguimiento y la implementación adecuados de la IA a nivel nacional.
Art. 34 - 39	Disposiciones relativas a los órganos de la Convención	Ética y Responsabilidad	Se refiere a la gobernanza y al marco ético de la IA a nivel internacional.