



Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España

Lucía Ortiz de Zárate Alcarazo
Ariana Guevara Gómez

Lucía Ortiz de Zárate

Investigadora predoctoral y Candidata a Doctora en la Universidad Autónoma de Madrid (UAM). Es graduada en Física y también en Filosofía (Premio Extraordinario de Fin de Grado y Premio Alumni UCM de excelencia en estudios de grado) por la Universidad Complutense de Madrid (UCM). Además, ha realizado un Máster en Astrofísica (UCM) y otro en Democracia y Gobierno (UAM). Ha sido becaria JAE-Intro en el Instituto de Filosofía del CSIC, en el Departamento de Ciencia Política y Relaciones Internacionales de la UAM, e investigadora visitante en el Harvard-Smithsonian Center for Astrophysics. Actualmente realiza una tesis doctoral sobre ética y gobernanza de la Inteligencia Artificial.

Ariana Guevara Gómez

Candidata a doctora en el Departamento de Ciencia Política y Relaciones Internacionales de la Universidad Autónoma de Madrid (UAM). Tiene un Máster en Democracia y Gobierno, y un Máster en Gobernanza y Derechos Humanos, ambos de la UAM. Es licenciada en Comunicación Social por la Universidad Central de Venezuela. Forma parte del equipo editorial de Agenda Estado de Derecho para Latinoamérica, y es integrante del Lab Grupo de Investigación en Innovación, Tecnología y Gestión Pública de la UAM. Investiga sobre género, tecnología y administración pública.



Ninguna parte ni la totalidad de este documento puede ser reproducida, grabada o transmitida en forma alguna ni por cualquier procedimiento, ya sea electrónico, mecánico, reprográfico, magnético o cualquier otro, sin autorización previa y por escrito de la Fundación Alternativas.

© Fundación Alternativas
© Lucía Ortiz de Zárate
© Ariana Guevara Gómez

Maquetación: Belén Avilés González
ISBN: 978-84-18677-16-8

ÍNDICE

Resumen ejecutivo

- 1. ¿Qué es la Inteligencia Artificial y por qué debería importarnos?**
 - 1.1. Historia, definición y contexto
 - 1.2. Beneficios de la IA en el sector público
 - 1.3. Riesgos de la IA en el sector público
- 2. La Inteligencia Artificial desde una perspectiva de género**
 - 2.1. Los sesgos de género en la IA
 - 2.2. Casos y ejemplos
 - Chatbots y asistentes de voz
 - Softwares de reconocimiento facial
 - Datos sesgados y machine learning
 - 2.3. Los impactos diferenciados de la IA en mujeres y hombres
- 3. El marco europeo para una Inteligencia Artificial ética (y feminista)**
 - 3.1. Una Inteligencia Artificial europea y ética
 - 3.2. Criterios para la evaluación y el análisis de la perspectiva de género
 - 3.3. Una Inteligencia Artificial europea y feminista
- 4. El caso de Suecia como referente en la igualdad de género**
 - 4.1. Suecia y la Inteligencia Artificial
 - 4.2. Inteligencia Artificial y género
 - Potenciales usos de la IA para alcanzar la igualdad de género
- 5. España, Inteligencia Artificial e Igualdad de Género**
 - 5.1. España: una IA con sello europeo
 - 5.2. Género e Inteligencia Artificial en España
- 6. Propuestas para promover la igualdad de género en el uso de la IA en España**
 - 6.1. Propuestas transversales
 - 6.2. Propuestas concretas
- 7. Conclusiones**

Referencias

RESUMEN EJECUTIVO

La Inteligencia Artificial (IA) está transformando las empresas, las administraciones públicas y la vida cotidiana de las personas. Desde sistemas de análisis de datos y proyección de situaciones futuras, hasta chatbots y asistentes de voz, los sistemas de IA se están implementando cada vez más para automatizar procesos, incrementar la eficacia y eficiencia de las instituciones, y producir oportunidades de negocios. Si consideramos el uso cada vez más extendido de estas tecnologías disruptivas, se hace necesario analizar sus potencialidades, beneficios y riesgos.

En este documento, nos concentramos específicamente en los riesgos y oportunidades que ofrece la IA para la igualdad de género. En primer lugar, el vínculo entre el desarrollo de la IA y las cuestiones de género sólo es posible si se considera una perspectiva socio-técnica de la tecnología: entendemos que existe una relación entre el diseño de los avances tecnológicos y los asuntos sociales y éticos que se asocian a ellos. Si se habla, además, de unos sistemas que emulan ciertas características de la inteligencia humana, no es posible desligar estos desarrollos de los valores éticos de la sociedad, en este caso concreto de Europa y España.

En el marco del sector público, la IA tiene el potencial de incrementar la eficacia y eficiencia, favorecer los procesos de detección de problemas y predicción de soluciones, avanzar hacia la personalización de los servicios que se ofrecen a la ciudadanía, y favorecer la automatización de procesos y trámites. Pero, al mismo tiempo, hay retos en relación con la producción y perpetuación de sesgos, la rendición de cuentas, la transparencia, la privacidad y la explicabilidad de los algoritmos, que pueden tener un efecto en el desarrollo de la democracia y en el cumplimiento de los valores europeos y los derechos humanos.

Más concretamente, se ha demostrado que, en muchas ocasiones, los datos usados en los sistemas de IA reproducen estereotipos de género, y que el diseño de estas tecnologías disruptivas pueden contribuir con la perpetuación de roles asociados tradicionalmente a mujeres y hombres. Desde chatbots y asistentes de voz, hasta algoritmos en buscadores y sistemas de reconocimiento facial, existen ejemplos que demuestran cómo los estereotipos de

género permean en el diseño de la IA y cómo se hace necesaria la incorporación de una perspectiva que respete el principio de la igualdad de género en los desarrollos tecnológicos.

En Europa existe un abordaje de los principios éticos que deben seguirse en el desarrollo de una IA centrada en los seres humanos, que sirve de marco de referencia para incluir la perspectiva de género en la IA. En un análisis de los principales documentos oficiales europeos, que abordan los aspectos éticos del diseño y la implementación de la IA, se demuestra que la UE ha incluido referencias implícitas y explícitas que sirven de base para la incorporación de este enfoque de género en los avances tecnológicos de los Estados Miembros. El caso de Suecia, por ejemplo, es paradigmático, no sólo por la publicación de su estrategia nacional de IA casi a la par de la UE, sino porque ha desarrollado iniciativas concretas para usar estas tecnologías disruptivas de una forma proactiva a favor de la igualdad de género, en estas áreas: la influencia y el poder, la educación, la distribución igualitaria de los cuidados y el trabajo no remunerado, la igualdad económica, la lucha contra la violencia de género, y el acceso y disfrute igualitario a la salud.

España, por su parte, ha emitido recientemente algunos documentos que sirven de base para el diseño e implementación de la IA, entre ellos la estrategia nacional, publicada a finales de 2020. Estos documentos oficiales siguen la estela de la UE, con su apuesta por una IA centrada en los seres humanos. También los textos españoles incluyen referencias implícitas y explícitas al género, lo que se traduce en un buen punto de partida para diseñar e implementar sistemas de IA con un enfoque de género.

En este sentido, este documento ofrece cinco recomendaciones para concretar y profundizar en las propuestas generales que ha hecho España en relación con el género y la IA: la elaboración de guías para la implementación de la IA en las administraciones públicas (AAPP) con una perspectiva de género; la elaboración de un informe anual o bianual de IA con perspectiva de género, que contenga cifras y datos sobre el avance de proyectos que incluyan este enfoque; IA contra la violencia de género (diseñar e implementar un chatbot para ayudar a las mujeres que sufren violencia de género); IA contra la brecha salarial de género (impulsar el desarrollo de sistemas de IA para favorecer la transparencia retributiva en las empresas); IA

para favorecer la salud de las mujeres (Diseñar e implementar un chatbot para ayudar a las mujeres a resolver sus dudas sobre salud).

1. ¿Qué es la Inteligencia Artificial y por qué debería importarnos?

1.1. Historia, definición y contexto

Aunque el sueño de crear máquinas inteligentes no es reciente, el término Inteligencia Artificial (IA) para referirse a este tipo de entidades no se acuñó hasta el año 1956 durante una conferencia celebrada en Darmouth a la que asistieron algunos de los principales fundadores de lo que hoy se conoce como Inteligencia Artificial. La IA surge a partir del deseo de desarrollar máquinas que puedan simular la inteligencia humana. Normalmente, se considera al matemático Alan Turing (1912-1954) como el “padre” fundador de esta disciplina. El propio Turing contemplaba y reconocía que el intento por construir máquinas inteligentes no solo depende del ámbito de la matemática, la física y la ingeniería. Es decir, el intento de simular la inteligencia humana no solo requiere de una parte científico-tecnológica, sino también del estudio de lo que es en sí misma la inteligencia humana: cuáles son sus características, sus limitaciones, etc. La IA surge así de ideas y personas provenientes de distintas disciplinas. Algunas de las figuras más destacadas en este ámbito han sido: en el campo de la ingeniería Norbert Wiener y sus trabajos sobre cibernética; en biología destacan las contribuciones de W. Ross Ashby, Warren McCulloch y Walter Pitts sobre redes neuronales (neural networks), desde la psicología las aportaciones más relevantes son las de Allen Newell y Herbert Simon (premio Nobel de Economía en 1978); en el campo de la Teoría de Juegos encontramos a John Von Neumann y Oskar Morgenstern; en lógica y filosofía al ya mencionado Alan Turing, pero también a Alonzo Church y Carl Hempel; finalmente, en el área de la lingüística las contribuciones más destacadas son las del pensador Noam Chomsky¹.

En lo que se refiere a la Inteligencia Artificial como disciplina científica, es importante desterrar la idea recurrente de que ésta es una cuestión puramente tecnológica. Desde sus inicios, esta nueva área del conocimiento se ha nutrido de disciplinas tanto del ámbito de la ingeniería y las ciencias naturales, como de las ciencias sociales y las humanidades. Si el objetivo de la IA es

¹ Buchanan, B. G. (2006). A (Very) Brief History of Artificial Intelligence. AI Magazine, 26(4).

simular la inteligencia humana, entonces es imposible reducir los intentos al campo de la ingeniería y la matemática, es decir, aunque estas disciplinas son una condición necesaria para su desarrollo a nivel tecnológico, no son condición suficiente per se. Dado que la inteligencia humana no es tan solo mera materia interconectada, sino la conjunción de un sinnúmero de complejidades de carácter biológico, pero también social, su total simulación escapa al ámbito de la ciencia y la tecnología.

Más allá de la Inteligencia Artificial como disciplina científica o área del conocimiento y de estudio, es importante señalar que actualmente no existe una definición consensuada respecto a qué es la IA como tecnología. A pesar de la falta de acuerdos, la mayor parte de las definiciones sobre IA la describen como tecnologías que hacen uso de algoritmos inteligentes de tal modo que son capaces de reproducir y/o simular ciertas capacidades humanas.

Definición IA: Tecnologías que hacen uso de algoritmos inteligentes de tal modo que son capaces de reproducir y/o simular ciertas capacidades humanas (aprender de su entorno y, con base en él, tomar decisiones).

A modo de aclaración. Primero, por tecnología nos referimos tanto a sistemas solamente basados en software, como a sistemas compuestos por software y hardware, es decir, una tecnología de IA puede ser tanto un asistente de voz² de una página web, como un robot inteligente, respectivamente. Segundo, un algoritmo es un sistema, conjunto o secuencia de reglas u operaciones lógicas que permite realizar cálculos de distinto tipo y, por lo tanto, encontrar soluciones a eventuales problemas o demandas. Tercero, decimos que los algoritmos son inteligentes porque éstos, a través del uso de técnicas como el machine learning, el deep learning u otros, son capaces de “aprender” de la experiencia. Algunas tecnologías anteriores a la aparición de la IA también han hecho uso de algoritmos, es el caso de los sistemas expertos. Sin embargo, mientras que los sistemas expertos hacen uso de algoritmos “finitos” o “cerrados”, es decir, que se componen de un conjunto de reglas y directrices, definidas por los programadores, que son inamovibles, las tecnologías de IA usan algoritmos que a partir de un número determinado de algoritmos y nutriéndose de grandes bases de datos (big data) son

² Un ejemplo cercano es el caso de ISSA, el chatbot que utiliza la Seguridad Social para facilitar consultas o trámites en sus servicios. <http://www.seg-social.es/wps/portal/wss/internet/Inicio>

capaces de “aprender” y, por lo tanto, desarrollar nuevas reglas o directrices que van más allá de las inicialmente programadas. Cuarto, debido a esta nueva y genuina capacidad para “aprender” a partir de la experiencia (datos) y que le vale el apodo de “inteligente” a esta nueva tecnología, es por lo que se le atribuye a la IA su capacidad para reproducir y/o simular capacidades hasta ahora solo asociadas, a tan alto nivel, a los seres humanos y algunos pocos animales. Por otro lado, como ya hemos mencionado, existen distintas técnicas que pueden usar los algoritmos para obtener esa “inteligencia”. Algunas de estas técnicas son el machine learning, el deep learning o las neural networks, entre otras, muchas de ellas basadas en el reconocimiento de patrones para hacer inferencias que le permitan derivar y crear nuevas reglas con base en las cuales, nuevamente, evaluar su entorno (datos) y tomar decisiones (procesamiento de datos).

Tipo de IA	Capacidades	Potenciales usos
Análisis de imágenes	<ul style="list-style-type: none"> Identificación de personas Identificación de objetos Clasificación de imágenes y videos Detección de similitudes Reconocimiento de emociones 	<ul style="list-style-type: none"> Reconocimiento facial Detección de armas, tumores, etc. Detección de contenido sensible Detección de casos de plagio Seguridad
Análisis de audio y texto	<ul style="list-style-type: none"> Identificación de personas Conversión audio a texto Detección de sonidos Reconocimiento de emociones 	<ul style="list-style-type: none"> Verificación de identidades online Adaptación para personas con discapacidad Reconocimiento de voces Seguridad
Procesamiento natural del lenguaje	<ul style="list-style-type: none"> Identificación de personas Análisis de sentimientos Generación de contenido 	<ul style="list-style-type: none"> Evitar casos de plagio, atribuir autorías Analizar tendencias en redes sociales Chatbots
Otros	<ul style="list-style-type: none"> Procesamiento de datos ambientales y del entorno Interconectividad Capacidad de adaptación al entorno 	<ul style="list-style-type: none"> Predicciones climáticas y climatológicas Simulaciones a gran escala Asesoramiento político

Fuente: Elaboración propia basada en la Estrategia Sueca de IA.

Estas características únicas de la IA la han convertido en la tecnología protagonista por excelencia en la actualidad. Tanto empresas privadas como gobiernos y administraciones públicas en todo el mundo están apostando por el uso de este nuevo conjunto de tecnologías disruptivas que prometen alcanzar resultados y realizar tareas hasta ahora inimaginables para ninguna máquina. Debido a sus dos cualidades distintivas: primero, que se nutre de grandes

cantidades de datos; segundo, su capacidad para simular comportamientos inteligentes (aprender y tomar decisiones); la IA presenta una capacidad de computación mucho más alta que otras tecnologías previas y una precisión en la realización de evaluaciones, diagnósticos y predicciones hasta ahora desconocida. De este modo, el uso de la IA nos permite soñar con reducciones en los tiempos de ejecución de tareas, con avances y mejoras en el rendimiento, aumentos de la eficacia y la eficiencia, y con el incremento de la oferta de servicios altamente personalizados, entre otras cosas.

En el caso del sector público, se espera que la IA traiga consigo un cambio de paradigma en los modelos de gobernanza. Lo que se ha venido a conocer con el nombre de Gobernanza Algorítmica³. El nuevo paradigma de la Gobernanza Algorítmica propone un modelo de gobernanza en el que gran parte de las decisiones son tomadas o están parcialmente basadas en algoritmos. Este nuevo paradigma conlleva una algoritmización de las administraciones públicas de modo que las antiguas estructuras de control basadas en la burocracia tradicional pasen a basarse en algoritmos inteligentes. La Gobernanza Algorítmica genera tanto grandes promesas e ilusiones, como dudas y reticencias. Los beneficios que se esperan obtener a través de la introducción de la IA en el sector público son muchos y muy variados⁴. Sin embargo, también conlleva serios riesgos y problemas, principalmente, de carácter ético y moral que hay que evaluar, estudiar y resolver si queremos que la nueva Gobernanza Algorítmica sea sinónimo de progreso y no de violación de los Derechos Humanos.

³ También conocida como Gobernanza del Dato o Capitalismo de la Vigilancia (Zuboff, 2019). Aunque si bien es cierto que el uso de unos términos u otros para referirse a estos nuevos modelos conlleva determinadas connotaciones más o menos negativas.

⁴ A este respecto ver Coeckelbergh, M. (2020). *AI Ethics*. MIT Press; Margetts, H., y Dorobantu, C. (2019). Rethink government with AI. In *Nature* (Vol. 568, Issue 7751, pp. 163–165). Nature Publishing Group. <https://doi.org/10.1038/d41586-019-01099-5>

1.2. Beneficios de la IA en el sector público

Aunque en la actualidad la mayor parte de las organizaciones que hacen uso de tecnologías de IA son las empresas privadas, cada vez más gobiernos, tanto a escala nacional como local, apuestan por la implementación de la IA en sus administraciones públicas y demás órganos dedicados a la toma de decisiones. Algunos de los principales beneficios que se esperan a partir de la incorporación de la IA en el sector público son⁵:

- Eficacia y eficiencia

La incorporación de la IA al sector público, si se realiza de manera correcta, puede incrementar la eficacia y la eficiencia de distintos procesos y servicios que son propios del ámbito de la gobernanza. Algunos ejemplos son:

- a) Políticas públicas: Los ciclos de políticas públicas constan de cinco pasos. Primero, está lo que se conoce como la agenda-setting, que consiste en la definición de un problema como relevante o de interés público (debido a distintas circunstancias) y su entrada en el foco mediático. Segundo, la formulación de políticas públicas, que es cuando se plantean las distintas y posibles políticas que pueden dar solución al problema tratado. Tercero, la selección de una política o un paquete de políticas públicas concretas para solucionar el problema con base en una serie de criterios. Cuarto, la implementación de la/las políticas seleccionadas durante un periodo de prueba. Quinto, la evaluación del éxito de la política en la solución del problema. Este ciclo, que normalmente se extiende durante largos periodos, podría ver su duración reducida de forma muy significativa o incluso transformarse completamente a través de la IA. El uso de IA podría modificarlo a través de rápidos y extensos análisis de datos de forma que pudiera identificar de manera más eficaz y eficiente cuáles son los principales problemas a tratar, ordenarlos por nivel de importancia (con base en unos criterios dados), plantear un mayor número de soluciones posibles, e incluso realizar simulaciones de los distintos escenarios en función de la solución elegida. Además, debido a su capacidad de

⁵ Es importante señalar que la mayor parte de los usos de IA que se enumeran a continuación todavía no se han implementado o están en fase experimental salvo en algunos casos concretos.

evaluación en tiempo real, podría acelerar notablemente los procesos de evaluación de las políticas públicas de modo que estos fueran prácticamente constantes. Es decir, que los posibles fallos en el diseño y/o la implementación de las políticas se fueran detectando a medida que se aplican.

- b) Servicios públicos: La IA también puede ser de gran utilidad para mejorar la eficacia y la eficiencia de los servicios públicos que proporcionan las administraciones públicas. Uno de los usos más populares y extendidos actualmente es el de los asistentes de voz inteligentes y los chatbots. Al igual que con otras formas de IA, el uso de chatbots aplicado a servicios de atención al ciudadano para resolver dudas, ayudar en la gestión de trámites, etc., permite que las administraciones públicas estén operativas y accesibles para los ciudadanos las 24 horas del día, los 7 días de la semana.
- c) Administraciones públicas: El correcto y eficaz funcionamiento de las administraciones públicas (AAPP) es fundamental para gozar de democracias sanas y fuertes. Sin embargo, estudios y encuestas sobre la percepción ciudadana acerca del funcionamiento de las AAPP españolas muestran que los/as españoles/as tienden a percibir estas instituciones como grandes y lentas estructuras burocráticas con grandes dificultades para ser dinámicas e innovar. En este sentido, la incorporación de la IA en los procesos internos de la AAPP permitiría acelerar los trámites y operaciones que tienen lugar en estas organizaciones a través de la automatización.
- d) Asesoría: política, técnica, jurídica, etc.: A diferencia de otras tecnologías anteriores, la “inteligencia” desplegada por la IA permite que esta lleve a cabo labores de asesoría política, técnica, jurídica, médica, etc. Debido a su capacidad y velocidad en el procesamiento de información, la IA puede ofrecer más alternativas a barajar en las cuestiones tratadas, así como simulaciones de posibles escenarios en función de la alternativa seleccionada.

- **Detección y predicción**

El uso de IA en el sector público permite un análisis en tiempo real de las ciudades y municipios en los que está operando. Esta característica, sumada a su capacidad de simulación, permite que la IA sea capaz de detectar problemas de distinta índole más rápido de lo que hasta ahora estábamos acostumbrados, incluso en tiempo real. Además, la IA puede realizar predicciones de posibles situaciones de riesgo o potencialmente peligrosas a las que es necesario dedicar más recursos. Algunos ejemplos son:

- a) El uso de IA para realizar diagnósticos médicos. En algunos países ya se está probando el uso de IA para el asesoramiento en tareas relacionadas con la salud. Gracias a su capacidad para identificar patrones, la IA puede detectar con gran precisión anomalías en radiografías, electrocardiogramas, resonancias magnéticas, etc. De esta forma se pueden detectar enfermedades de forma temprana con mayor facilidad, lo que aumenta las probabilidades de curación de los pacientes. La IA también podría ser usada para la prescripción de tratamientos médicos eficaces ya que, en teoría, debería ser capaz de almacenar y procesar la información y experiencia disponible no solo por un médico, sino por toda la comunidad científica.
- b) Crimen y seguridad: A través de los mismos mecanismos que en el ejemplo anterior, la IA también se puede utilizar para detectar ilegalidades e infracciones en tiempo real, así como predecir posibles “puntos calientes” donde se concentre una mayor cantidad de actos delictivos, de forma que se pueda reforzar la presencia policial y/o los recursos para la vigilancia de la zona. También se ha propuesto el uso de la IA para predecir las posibilidades de reincidencia de algunos criminales, de forma que sirva de criterio para otorgar permisos para salir de prisión, etc.
- c) Fracaso escolar: Para la mejora en el sistema educativo, se prevé que la IA pueda detectar de forma precoz casos de fracaso escolar y así reducir las tasas de abandono en las escuelas.
- d) Tráfico: En el caso del tráfico, tanto en ciudad como en carretera, la IA podría detectar accidentes de tráfico e incidencias similares de forma casi instantánea, lo que permitiría dar una respuesta más rápida y aumentar la posibilidad de

salvar vidas y reducir la gravedad de los hechos. Al igual que en los casos de prevención del crimen, la IA puede ser útil para detectar posibles puntos negros en carreteras y vías de circulación, para incrementar la presencia de radares y otras medidas de seguridad.

- **Personalización**

Otro de los principales beneficios que conlleva la adopción de IA en el sector público es el incremento en la personalización en distintas áreas. La personalización implica que, en vez de ofrecer procesos y/o servicios estandarizados, cada ciudadano/a reciba comunicaciones y tenga acceso a trámites y servicios individualizados con base en sus propias particularidades. Además, la capacidad de personalización de la IA es el elemento fundamental que determinará gran parte de su éxito en labores de detección y predicción como las anteriormente mencionadas. Por ejemplo, en el fracaso escolar, para un tratamiento precoz de los casos, es necesario ser capaz de identificarlos y tratarlos individualmente para que se puedan aportar soluciones específicas a cada uno de ellos. No sería lo mismo evitar un caso de fracaso escolar por motivo de pertenencia a colectivos históricamente discriminados a razón de su sexo, orientación sexual, o religión, que por cuestiones médicas, psicológicas, o de distinta índole.

- **Automatización**

Finalmente, en la base de todos estos beneficios se encuentra la automatización de servicios, trámites y procedimientos de distinta índole y naturaleza que atañen al sector público en toda su complejidad. Aunque la automatización no es un beneficio en sí mismo, sí que es condición necesaria para hacer posibles todas las ventajas y mejoras mencionadas hasta ahora, así como otras que en este informe no hemos tratado. Por ejemplo, uno de los sectores más beneficiados por la automatización es el sector del transporte. El uso de vehículos autónomos, es decir, que no necesitan conductores, es una de las grandes promesas de la IA, que ya es una realidad en muchas partes del mundo.



Fuente: Elaboración propia

Riesgos de la IA en el sector público

Los principales riesgos de naturaleza ética o moral que se derivan del uso de la IA se deben, principalmente, a deficiencias y malos usos de alguna de las dos partes esenciales que componen este tipo de tecnologías: los datos o los algoritmos. Es decir, los problemas pueden deberse o bien a un mal diseño de los datos o los algoritmos, por separado, o a ambas cosas a la vez. A continuación, se exponen algunos de los principales riesgos asociados al uso de la IA⁶⁷.

- Sesgos

- a) En los datos: Las tecnologías de IA hacen uso de grandes cantidades de datos (big data). Las AAPP obtienen los datos y la información con la que operan, principalmente, a través sus propios registros digitalizados y bases de datos, cesiones voluntarias y acordadas por parte de otras administraciones y empresas públicas, fuentes de datos abiertas, depósitos de datos, etc. También

⁶ En este caso tan solo se señalan problemas de carácter ético y/o políticos derivados del uso de la IA. También existen problemas de carácter tecnológico que tienen que ver con diseño y soporte a nivel técnico de la IA que debido a la temática, finalidad y limitaciones del presente trabajo no se van a tratar.

⁷ Es importante señalar que los siguientes problemas éticos se perciben como tal en la medida que nos encontramos dentro de un marco democrático, basado en la libertad y la igualdad de los ciudadanos y ciudadanas, como es el caso de España y la Unión Europea.

es posible la compra de datos a empresas privadas. El problema es que estos datos con los que trabaja la IA están sesgados de partida en la medida que reflejan las discriminaciones presentes en la sociedad. Si, por ejemplo, la IA se nutre de datos en los que de forma sistemática las mujeres cobran menos dinero que los hombres, entonces, la máquina, al naturalizar estos datos, hará predicciones que continúen con las tendencias y patrones que ha identificado si no se le aplican criterios o mecanismos correctores. Casos muy populares en este ámbito son los de la IA aplicada en el sistema judicial norteamericano donde estas tecnologías dictaban sentencias en las que se inculpaba mayoritariamente a hombres negros de forma errónea debido a los datos extraídos de condenas previas. Para evitar estos casos, es fundamental llevar a cabo auditorías de datos que sean transparentes, de forma que los/as ciudadanos/as sepan qué datos se están utilizando y garantizar la representatividad de las muestras. Con ello lo que se pretende es reducir el nivel de desigualdad en los datos.

b) En los algoritmos: En el caso de los algoritmos, estos pueden ahondar en las discriminaciones producidas por los sesgos en los datos, si en ellos no se programan medidas correctoras. Es decir, del mismo modo que en algunos procesos de selección en el ámbito público existen criterios que fomentan la discriminación positiva a favor de colectivos históricamente excluidos, estas medidas deben reflejarse en la programación de los algoritmos si lo que se busca es reducir las desigualdades. De lo contrario, la inacción no solo conlleva la perpetuación de las desigualdades ya existentes, sino que puede incluso acrecentarlas debido a la alta capacidad de la IA.

- **Rendición de cuentas**

Uno de los mayores desafíos que podemos encontrarnos al usar la IA tiene que ver con la capacidad de estas tecnologías para rendir cuentas. Actualmente, cuando algo sale mal y queremos determinar y atribuir responsabilidades por determinadas acciones, la tarea puede ser más o menos sencilla, pero es posible. En el caso de la IA, esta cuestión es mucho más compleja. Si un autobús urbano autónomo se ve involucrado en un accidente de tráfico, ¿quién sería el responsable? Si sale a la luz que un sistema

inteligente ha estado concediendo ayudas públicas con base en criterios discriminatorios, ¿quién debe rendir cuentas? Aunque este es un tema muy complejo que aún hoy se sigue discutiendo ampliamente en los círculos de expertos sobre IA, en general, hay cierto consenso sobre el hecho de que, para que se puedan rendir cuentas por decisiones tomadas por una tecnología de IA, tienen que cumplirse dos requisitos: la transparencia y la explicabilidad.

- **Opacidad y falta transparencia**

Uno de los mayores peligros que conlleva el uso de la IA es la falta de transparencia u opacidad en sus datos y/o algoritmos.

a) En los datos: Debido a las características de los datos de los que se nutren las tecnologías de IA y para evitar la aparición de sesgos y ser capaces de rendir cuentas, es fundamental que los datos de los que hace uso la IA sean transparentes y públicos. Si, por el contrario, los datos son opacos es imposible determinar con precisión la fuente y el origen de los problemas.

b) En los algoritmos: Del mismo modo sucede con el diseño y la programación de los algoritmos. La opacidad algorítmica conlleva la imposibilidad para rendir cuentas y solucionar problemas por parte de las tecnologías de IA. Para resolver cualquier problema relacionado con la IA, es imprescindible que los algoritmos se puedan auditar, como sucede en el caso de otros procedimientos que se llevan a cabo en el sector público, para así poder determinar la fuente del problema. Por ello, garantizar la transparencia algorítmica debe ser una de las prioridades fundamentales de todas las organizaciones públicas que hagan uso de la IA.

- **Explicabilidad**

La explicabilidad, es decir, la capacidad para explicar los algoritmos y los datos que se usan, es la segunda condición necesaria para garantizar la rendición de cuentas y, por tanto, una algoritmización responsable de las administraciones públicas.

a) En los datos: En este caso, se hace referencia a la capacidad para explicar por qué se han usado unos datos y no otros; cuáles han sido los criterios que se han seguido para la recolección de los datos, etc. Los/as ciudadanos/as, y todas las

personas usuarias de la IA, deben tener al menos la posibilidad de comprender los procesos de toma de decisiones en los que se ven involucrados y que tienen repercusiones directas en sus vidas.

- b) En los algoritmos: De forma similar a como sucede en el caso de los datos, los algoritmos que se usan en el sector público, ya sea para conducir un vehículo o para repartir subsidios o ayudas, tienen que poderse explicar de forma que sean comprensibles para la ciudadanía. Es decir, todas las personas tienen que ser capaces de comprender la cadena de “razonamientos” o secuencia de reglas que ha seguido un algoritmo hasta alcanzar una decisión. Esta tarea es especialmente compleja y requiere un extra de atención por parte de tecnólogos, políticos y funcionarios, ya que los algoritmos suelen ser sistemas bastante complejos. En este sentido, resulta difícil creer que la explicabilidad de los algoritmos se solucionaría simplemente haciéndolos públicos, ya que tan solo un porcentaje muy bajo de la población sería capaz de interpretarlos. Por eso, es necesario realizar procesos de “traducción” de los algoritmos, de manera que puedan ser comprendidos por la mayoría y así, públicamente explicables.

- **Privacidad**

Una de las cuestiones más sensibles que atañen al uso de la IA tiene que ver con los problemas y riesgos en relación con la privacidad. Este es un tema especialmente relevante ya que el respeto a la privacidad y la intimidad es un asunto altamente valorado, especialmente en las sociedades occidentales desde la aparición de la Carta de Derechos Humanos.

- a) En los datos: La IA necesita datos e información para aprender a reconocer patrones y tomar decisiones. En este sentido, cuanto mayor sea la cantidad de datos de los que dispone la IA, más precisos serán sus diagnósticos, análisis y decisiones. Sin embargo, la recopilación, no solo de grandes cantidades de datos, sino de ciertos tipos de datos, necesariamente entra en conflicto con la preservación de la privacidad de los/as ciudadanos/as. Por ejemplo, la recolección de datos que hacen referencia a la orientación o prácticas sexuales de las personas está actualmente prohibida; sin embargo, muchos sistemas de

IA son capaces de inferir estos y otros compartimientos estrictamente privados y personales a partir de otros datos que sí están disponibles. En otros casos, se ha alegado la necesidad de recopilación de información biométrica para mejorar los análisis médicos, pero también, por ejemplo, para determinar el estado de salud de las personas y así permitir o prohibir su entrada en determinados lugares en épocas de crisis sanitarias, como durante la pandemia de la COVID-19.

b) Algoritmos: En relación con los algoritmos, la cuestión de la privacidad no es en apariencia un problema tan grave como sucede en el caso de los datos. Sin embargo, también es un asunto que se debe tener en cuenta. Aquí lo más relevante es encontrar un equilibrio entre algoritmos transparentes pero que, al mismo tiempo, respeten ciertas cuestiones que podrían acarrear intrusiones o deterioros de la privacidad de las personas.

- **Otros problemas derivados del uso de IA.** Finalmente, otros riesgos que también se atribuyen al uso de la IA en el sector público, aunque no tienen que ver necesariamente con el diseño de la IA, son: la pérdida de empleo público, la deshumanización del trabajo y la falta de libertad, entre otros. La pérdida de empleo es una de las mayores preocupaciones por parte de la ciudadanía a la hora de afrontar la revolución tecnológica producida por la incorporación de tecnologías inteligentes al ámbito laboral. Aunque no existe consenso sobre las predicciones de las pérdidas de empleo por parte de humanos que verán su trabajo sustituido por máquinas, la Unión Europea (UE) prevé que entre el 50% y el 70% de los trabajos se vean afectados de algún modo por la automatización. Además, se estima que entre el 45% y el 60% de la fuerza laboral europea podría verse reemplazada por la automatización en el año 2030⁸. Otra de las grandes preocupaciones es la deshumanización del sector público debido a la creciente sustitución de personas por máquinas y la automatización de los servicios públicos. Por ello, es imprescindible desarrollar una IA que sea digna de confianza por parte de la ciudadanía. Finalmente, la falta de libertad y el determinismo tecnológico son otras de

⁸ Datos de la Comisión Europea, publicados en: https://knowledge4policy.ec.europa.eu/foresight/topic/changing-nature-work/new-technologies-automation-work-developments_en

las preocupaciones más frecuentes derivadas de la incorporación de la IA a nuestras vidas. Aunque con menos frecuencia, son muchos los que temen que la capacidad de personalización, detección y predicción de estas tecnologías derive en una disminución de la libertad de elección humana producida por un exceso de “guía” por parte de la tecnología, de forma que esta provoque cierto determinismo.

Todos los beneficios y riesgos que se han planteado, a raíz de la incorporación de tecnologías de IA en el sector público, se consideran como tal beneficiosas o negativas en la medida que están orientadas, o no, a la consecución del bienestar común. En este sentido, por una parte, la mejora en los niveles de eficacia y eficiencia, la mayor personalización de los servicios, las labores de detección y predicción de posibles problemas, etc., y, por otra parte, el evitar riesgos como la falta de privacidad, la opacidad algorítmica o la presencia de sesgos, entre otros, se concibe desde la perspectiva de que todas estas acciones tendrían como consecuencia el aumento de los niveles de justicia e igualdad. Dos pilares fundamentales sobre los que se sostienen las democracias liberales.

2. La Inteligencia Artificial desde una perspectiva de género

Como ocurre en muchos ámbitos de la vida humana, el desarrollo de la Inteligencia Artificial no es neutro desde el punto de vista de género. Si consideramos que, como señala Cecilia Ridgeway⁹, “el género es un marco cultural primario para organizar las relaciones sociales”, no es de extrañar que los roles y estereotipos de género permeen en el diseño e implementación de la tecnología. Concretamente en el caso de la IA, durante los últimos años, los/as investigadores/as y organismos internacionales se han preocupado con creciente ahínco por destacar los riesgos que presenta esta tecnología para perpetuar y aumentar las desigualdades ya existentes entre hombres y mujeres¹⁰. Este es un asunto de primer orden si se considera que la promoción de la igualdad de género es fundamental en las democracias liberales y es uno de los objetivos de la Unión Europea, recogidos tanto en la Carta de Derechos Fundamentales de la UE, del año 2000, como en el Tratado de Lisboa, de 2007.

Desde un punto de vista teórico, se ha señalado que el propio concepto de Inteligencia Artificial contiene ciertos sesgos de género. Lo que se entiende como “inteligencia” en estos desarrollos tecnológicos está más relacionado con el pensamiento lógico y racional, que históricamente se ha asociado de manera simbólica con la masculinidad occidental¹¹. Por su parte, otras formas de inteligencia más ligadas a la emocionalidad o la empatía, comúnmente asociadas a los saberes de las mujeres, se han dejado de lado en estas definiciones. Por supuesto, esto no quiere decir que los hombres no desarrollen estas características, o que haya que perpetuar estereotipos que consideran a las mujeres más emocionales, sino que se deben revisar las propias definiciones desde un punto de vista feminista, para contrarrestar las concepciones dominantes de lo que significa el saber, la inteligencia y la tecnología.

⁹ Ridgeway, C. (2009). Framed Before We Know It: How Gender Shapes Social Relations. *Gender & Society*, 23(2), 145-160. <https://doi.org/10.1177/0891243208330313>

¹⁰ Vale la pena destacar los esfuerzos de la UNESCO, a través de sus informes *I'd blush if I could. Closing gender divides in digital skills through education* (2019) y *Artificial Intelligence and Gender Equality* (2020).

¹¹ Collett, C. y Dillon, S. (2019). AI and gender. Four Proposals for Future Research. [Última visita 20.03.2021]. <https://www.repository.cam.ac.uk/handle/1810/294360>

Por otra parte, desde una perspectiva más práctica, los/as investigadores/as han hecho reflexiones en torno a dos asuntos: los sesgos en los datos y en el diseño de los algoritmos, y el impacto diferenciado de la IA en la vida diaria de hombres y mujeres. En este sentido, se ha destacado la importancia de incluir la perspectiva de género en todo el ciclo de vida de la IA, desde el diseño hasta la implementación, y considerando el contexto¹². Para lograrlo, los sistemas de IA deben ser transparentes, capaces de rendir cuentas y comprensibles para todas las personas.

2.1. Los sesgos de género en la IA

A continuación, se presentan los riesgos de sesgos de género que se han analizado en los desarrollos de la IA, tanto en los datos como en el diseño.

- En los datos: Como se señaló en el apartado dedicado a los riesgos de la IA en el sector público, los sistemas de IA necesitan de una gran cantidad de datos para funcionar adecuadamente. Pero la cantidad no necesariamente está aparejada con la diversidad y la representatividad de esos datos. En muchos casos, los datos no son representativos de lo que realmente ocurre en la sociedad debido, por ejemplo, a la no inclusión de información referida a ciertos grupos sociales, con frecuencia mujeres, que no tienen acceso a la tecnología y que, por lo tanto, no tienen la posibilidad de generar esos datos¹³. Si no se presta atención a esas exclusiones, esas personas seguirán siendo invisibles y los sistemas de IA perpetuarán su condición.

Por ejemplo, la brecha digital de género en el uso de Internet se ha ido reduciendo durante los últimos años. Sin embargo, todavía existen diferencias: a escala mundial, el 55% de los hombres usa Internet en comparación con 48% de las mujeres¹⁴. Hay regiones en las que esta diferencia es aún más acentuada: en África, el 37% de los hombres usa Internet, en comparación con el 20% de las mujeres¹⁵. En el caso de España, el 14% de las mujeres manifestó en 2019 que no había usado Internet en los

¹² Así se destaca en el informe Artificial Intelligence and Gender Equality, publicado por UNESCO en 2020.

¹³ Collett, C. y Dillon, S. (2019). AI and gender. Four Proposals for Future Research. [Última visita 20.03.2021]. <https://www.repository.cam.ac.uk/handle/1810/294360>

¹⁴ De acuerdo con datos de la Unión Internacional de Telecomunicaciones, publicados en 2020: <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf>

¹⁵ Ibid.

últimos tres meses¹⁶. Eso significa que las mujeres siguen en desventaja, no solo en el acceso a las oportunidades que brinda Internet y las tecnologías, sino también en relación con la generación directa de datos con los que se alimentan los sistemas de IA.

Aquí hay que prestar atención a un asunto importante. La representatividad de las mujeres en los datos no solo se refiere a la inclusión de datos generados por ellas o al número de mujeres que se toman en cuenta en las bases de datos. También se refiere a asuntos simbólicos: ¿qué tipo de mujeres están representadas? ¿se generaliza la realidad de las mujeres atendiendo solo a ciertas características geográficas o sociales (es decir, mujeres occidentales y blancas)? Por eso, se ha apuntado la importancia de considerar la interseccionalidad; es decir, las relaciones del género con otras variables como la edad, la condición socioeconómica, la orientación sexual, el sitio en el que viven, etc. Los datos deben reflejar esa complejidad¹⁷.

Por otro lado, el uso de datos que reflejan las desigualdades reales entre hombres y mujeres pueden perpetuar y acrecentar esas desigualdades. Por ejemplo, en Amazon se empezó a utilizar un sistema de IA para reclutar a nuevo personal, que se alimentaba con los datos de los trabajadores de los últimos 10 años. Como en ese periodo la mayoría de los puestos técnicos eran ocupados por hombres, el algoritmo empezó a descartar a las mujeres porque no coincidían con los datos con los que se había programado el sistema¹⁸. La empresa decidió abandonar el proyecto, que demostró cómo los datos, aparentemente neutros, pueden esconder sesgos que obstaculizan la igualdad de género. Por eso, se debe reflexionar críticamente sobre los datos que se utilizan en los sistemas de IA, cuestionar las desigualdades que reflejan y aplicar las correcciones en función de los contextos específico¹⁹

¹⁶ Sáinz, M., Arroyo, L., y Castaño, C. (2020). Mujeres y digitalización. De las brechas a los algoritmos. Instituto de la Mujer y para la Igualdad de Oportunidades. [Última visita 18.03.2021]. https://www.inmujer.gob.es/disenio/novedades/M_MUJERES_Y_DIGITALIZACION_DE_LAS_BRECHAS_A_LOS_ALGORITMOS_04.pdf

¹⁷ UNESCO. (2020). Artificial Intelligence and Gender Equality. <https://unesdoc.unesco.org/ark:/48223/pf0000374174>

¹⁸ https://elpais.com/tecnologia/2019/05/21/actualidad/1558455279_966010.html

¹⁹ Collett, C. y Dillon, S. (2019). AI and gender. Four Proposals for Future Research. [Última visita 20.03.2021]. <https://www.repository.cam.ac.uk/handle/1810/294360>

- En el diseño: No solo los datos que se usan en los sistemas de IA pueden tener un efecto en la perpetuación de estereotipos de género. Si se considera el diseño y la conceptualización de estos sistemas, también se pueden encontrar riesgos. Es decir, quienes diseñan los sistemas de IA pueden volcar sus propios prejuicios en la estructura y las características (tanto en el software como en el hardware) de la IA. El ejemplo más claro se puede encontrar en los asistentes digitales que han desarrollado grandes compañías tecnológicas. En su mayoría, estas herramientas tienen características femeninas y reproducen estereotipos de género²⁰.

El diseño de estas tecnologías puede influir también en la generación de bases de datos y en la creación de algoritmos para la toma de decisiones en las administraciones públicas. Por ejemplo, el Ministerio de la Primera Infancia de la Provincia de Salta, en Argentina, a través de su Plataforma Tecnológica de Intervención Social y con el apoyo de Microsoft Azure, desarrolló un sistema de IA para predecir los embarazos adolescentes. En apariencia, se trataba de un proyecto beneficioso para la comunidad. Pero, en su conceptualización, perpetuaba estereotipos: los responsables crearon una base de datos conformada por niñas y jóvenes de 10 a 19 años, residentes en zonas de pobreza (con sus nombres y domicilios), y diseñaron un algoritmo para predecir cuáles de ellas tenían riesgos de vivir un embarazo adolescente en los próximos años con el objetivo de aplicar medidas preventivas. La iniciativa no solo basaba su diseño en la idea de que la responsabilidad del embarazo recaía únicamente de las mujeres, sino que también reforzaba estereotipos sobre las mujeres en situación de pobreza, además de generar dudas sobre el respeto a la privacidad y la intimidad²¹. En este caso, los sesgos se aprecian en el propio diseño de la política pública, en la que se usa la IA como una herramienta de análisis de datos y predicción. Considerando la capacidad de la IA para potenciar y multiplicar los resultados de una intervención, el daño puede ser mucho mayor que en proyectos que no usen este tipo de tecnologías. Es decir, si bien el diseño de cualquier política pública puede contribuir con la perpetuación de estereotipos de

²⁰ West, M., Kraut, R., y Ei Chew, H. (2019). I'd blush if I could. Closing gender divides in digital skills through education. [Última visita 17.03.2021]. <https://en.unesco.org/ld-blush-if-i-could>

²¹<https://medium.com/codingrights/decolonising-ai-a-transfeminist-approach-to-data-and-social-justice-a5e52ac72a96>

género, si no se tiene en cuenta este asunto, aquellas iniciativas que usen la IA corren el riesgo de generar un impacto aún más fuerte.

En algunos informes se ha destacado que la poca presencia de mujeres en los equipos de diseño de los sistemas de IA puede contribuir con estas conceptualizaciones estereotipadas²². Existen varios datos que demuestran esta realidad. En la Unión Europea, por ejemplo, la brecha digital de género en actividades de programación era de 8 puntos en 2014²³. Del total de autores de artículos en tres importantes conferencias de machine learning en 2017, solo 12% eran mujeres²⁴, y solo 13,83% de los autores de miles de artículos sobre IA son mujeres²⁵. En el ámbito empresarial, un estudio que utilizó datos de 287 compañías que trabajan con IA y machine learning demostró que sólo el 13% de las altas posiciones ejecutivas en empresas tecnológicas son ocupadas por mujeres²⁶. Si se considera el subsector de lenguaje natural, el porcentaje cae a 5%, y en el ámbito legal vinculado a la IA, no hay ninguna presencia femenina en estos altos puestos.

La diversidad en los equipos de diseño de la IA puede ofrecer nuevas perspectivas y traer a colación experiencias que permitan reconocer los sesgos y trabajar para corregirlos. Pero no es suficiente con incluir a más mujeres en estas actividades: hace falta cambiar las normas de las organizaciones y garantizar que las mujeres realmente puedan influir en las decisiones. También es necesario incluir personas expertas en perspectiva de género, porque no todas las mujeres son especialistas en esta área²⁷. Es decir, es importante promover el aumento de mujeres en los equipos de diseño y en las organizaciones que producen sistemas de IA, pero también hay que mirar el asunto con una perspectiva más amplia.

²² Así lo señalan los informes de UNESCO (2020), Collett y Dillon (2019) y West et al. (2019).

²³ Martínez-Cantos, J.L. (2017). Digital skills gaps: A pending subject for gender digital inclusion in the European Union. *European Journal of Communication*, 32(5), 419-438. <https://doi.org/10.1177/0267323117718464>

²⁴ <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/>

²⁵ https://media.nesta.org.uk/documents/Gender_Diversity_in_AI_Research.pdf

²⁶ <https://emerj.com/ai-market-research/women-in-artificial-intelligence-visual-study-leaderships-across-industries/>

²⁷ UNESCO. (2020). Artificial Intelligence and Gender Equality. <https://unesdoc.unesco.org/ark:/48223/pf0000374174>



Fuente: Elaboración propia

2.2. Casos y ejemplos

A continuación, se exponen algunos casos y ejemplos concretos en los que se evidencian los sesgos de género en los sistemas de IA.

- **Chatbots y asistentes de voz:** Cuando se habla de IA, es probable que se piense casi inmediatamente en los chatbots y los asistentes de voz, que han ganado mucha popularidad en los últimos años. Tanto los chatbots como los asistentes de voz son herramientas diseñadas para mantener interacciones con los usuarios a través de conversaciones. La diferencia principal entre ambos es que los primeros funcionan principalmente con texto, mientras que los segundos usan la voz.

En el caso de los chatbots, hay evidencias de que existen sesgos de género en su diseño. Un grupo de investigadores analizó 1.375 chatbots y encontró que la mayoría de ellos presentaban características femeninas: en algunos casos se trataba del nombre, en otros del avatar, y en otros de la descripción. La mayor parte de estos chatbots

femeninos se usan en áreas específicas, como el branding conversacional²⁸, atención al cliente y ventas. Estas herramientas no solo pueden contribuir a perpetuar los estereotipos y roles de género a través de su propio diseño, sino que también pueden reforzar los prejuicios que ya existen en los usuarios durante la relación humano-máquina. Por ejemplo, en un experimento con 60 personas²⁹ se encontró que los participantes le dieron peso a los estereotipos de género al momento de evaluar la eficiencia de un chatbot femenino para la asistencia mecánica, un área asociada tradicionalmente con lo masculino. En España hay varios ejemplos: está Carina, que ofrece información sobre la Covid-19; Helena, que responde dudas sobre los procesos de selección en el Servicio Murciano de Salud y Ámbito Docente; Rosi, que contesta preguntas sobre información básica de la ciudad de Murcia; o Lola, que ayuda a resolver dudas sobre matrículas y preinscripciones en la Universidad de Murcia. Todos ellos han sido desarrollados por la empresa 1 Million Bot, que se dedica a estos desarrollos tecnológicos y que cuenta con unos pocos chatbots con nombres y características masculinas.

En el caso de los asistentes de voz, también se ha demostrado que, en su mayoría, son diseñados para representar a mujeres. Las más conocidas son Siri de Apple, Alexa de Amazon, Cortana de Microsoft, y Google Assistant, que no tiene un nombre, pero sí una voz femenina. Además de sus nombres y voces, las compañías tecnológicas han desarrollado personalidades e historias de fondo, que claramente se asocian con características femeninas. En estos casos, se ha señalado que los asistentes de voz pueden representar y reforzar estereotipos de género³⁰. Primero, porque se asocian las características de asistencia al género femenino. Segundo, porque sus respuestas a mensajes abusivos son tibias o tolerantes, lo que puede consolidar la idea de que las mujeres están en posición de subordinación y son permisivas a los malos tratos.

²⁸ Se trata de un tipo de marketing que usan las marcas para promover sus productos y servicios a través de conversaciones por chatbots y redes sociales, entre otras herramientas digitales.

²⁹ McDonnell, M., y Baxter, D. (2019). Chatbots and gender stereotyping. *Interacting with Computers*, 31(2), 116-121.

³⁰ West, M., Kraut, R., y Ei Chew, H. (2019). I'd blush if I could. Closing gender divides in digital skills through education. [Última visita 17.03.2021]. <https://en.unesco.org/ld-blush-if-i-could>

Uno de los ejemplos más conocidos de este tipo de interacciones se publicó en Quartz, en 2017. Como prueba, le dijeron a Siri “You’re a slot” (“Eres una cualquiera”), y una de las respuestas de la máquina fue: “I’d blush if I could” (“Si pudiera, me sonrojaría”). Dado que este tipo de frases y respuestas de Siri y de otros asistentes de voz se podrían catalogar como complacientes o inadecuadas para responder al tono abusivo, después de esta publicación, las empresas tecnológicas hicieron un esfuerzo por hacer correcciones. Aunque mejoró el problema, faltan todavía mensajes más categóricos contra los abusos verbales³¹.

- **Softwares de reconocimiento facial:** Los softwares de reconocimiento facial usan bases de datos con etiquetas para identificar el color de la piel, la forma del rostro, el grosor del pelo, entre otras características faciales, con objetivos diversos: desde diagnósticos y tratamientos médicos, desbloqueo de dispositivos móviles y cajeros automáticos, hasta investigaciones policiales. Además de los dilemas éticos que supone el uso de estas tecnologías, especialmente en relación con la privacidad y la intimidad³², también existen evidencias de los sesgos que pueden afectar a las mujeres y a personas pertenecientes a minorías.

Un estudio publicado en 2018³³ llegó a dos conclusiones importantes sobre este tema. Primero, las investigadoras analizaron dos grandes bases de datos faciales (IJB-A y Adience) y encontraron que había una sobrerrepresentación de hombres de piel clara, y una subrepresentación de personas de piel oscura en general, y específicamente de mujeres. En vista de estas características descubiertas, decidieron desarrollar una base de datos específica, con imágenes de 1.270 personas provenientes de tres países africanos y tres europeos, para evaluar tres clasificadores comerciales: Face API de Microsoft, Watson Visual Recognition API de IBM, y Face++, una compañía con sede en China. La conclusión del análisis fue inquietante: los hombres fueron identificados de forma más precisa que las mujeres, y las personas de piel clara fueron clasificadas con

³¹ Ibid.

³² https://elpais.com/tecnologia/2019/05/21/actualidad/1558455279_966010.html

³³ Buolamwini, J., y Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research, 81, 1-15.

más precisión que las de piel oscura. En términos generales, los softwares acertaron en menor medida en el caso de mujeres de piel oscura.

¿Qué implicaciones tienen estos datos? Si un software de reconocimiento facial es incapaz de identificar adecuadamente un rostro con determinadas características, entonces los resultados que arrojen este tipo de tecnologías serán incorrectos. Si consideramos que estos softwares se usan cada vez con mayor frecuencia en ámbitos diversos, se podrían generar consecuencias negativas para la vida diaria de las personas y, concretamente, de las más vulnerables. Como señalaron las investigadoras del estudio, para evitar este tipo de problemas se hace indispensable garantizar la transparencia y la rendición de cuentas de la IA.

- **Datos sesgados y machine learning:** La incrustación de palabras o word embedding consiste en entrenar a ciertos programas de computación para que asocien una palabra con otra a través de vectores. En términos generales, por medio de los significados semánticos, se pueden establecer analogías y relaciones entre palabras, lo que permite generar una especie de diccionario para que funcionen los sistemas de machine learning y procesamiento de lenguaje natural³⁴. En este tipo de datos, también se pueden encontrar sesgos de género.

Un grupo de investigadores³⁵ analizó word2vec, un sistema de word embedding, entrenado con textos de Google News. Los resultados mostraron que los contenidos de las noticias presentan estereotipos de género y que estos estereotipos se pueden amplificar cuando se utilizan estos datos para alimentar sistemas de IA. Por ejemplo, en el extremo relacionado con she (ella) se encontraron profesiones como enfermera, ama de casa, recepcionista, bibliotecaria, niñera, estilista y otras. Por su parte, en el extremo relacionado con he (él) están los términos experto, capitán, filósofo, arquitecto, financiero, guerrero, protegido y otros.

³⁴ Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain. <https://papers.nips.cc/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf>

³⁵ Ibid.

Si consideramos que los asistentes de voz usan datos de Internet, se hace necesario que tanto los textos como las imágenes estén libres de sesgos. Un ejemplo en el caso de las imágenes se encuentra en lo que se denomina “el problema de la novia”: al analizar el contenido de ImageNet, una base de datos con más de 14 millones de imágenes que se usan para “avanzar en la investigación de la visión artificial y el deep learning”³⁶, encontró que una mujer blanca con un vestido blanco se etiqueta como novia, mientras que una mujer del norte de la India con un traje típico de boda se etiqueta como una interpretación artística³⁷. Este ejemplo demuestra que hay que mirar la representatividad de género de los datos con lentes de interseccionalidad, y que los sesgos pueden llevar a malas interpretaciones que pueden desembocar en decisiones injustas.

2.3. Los impactos diferenciados de la IA en hombres y mujeres

Los sesgos en los datos y en el diseño de los sistemas de IA pueden traducirse en consecuencias negativas para las personas y, en muchos casos, para las mujeres de forma más concreta. No solo se pueden perpetuar estereotipos que afectan el desarrollo de las mujeres, sino que también se pueden tomar decisiones sesgadas que las dejen al margen. Pero, incluso sin considerar esos sesgos en la creación de la IA, el propio despliegue de estas tecnologías disruptivas puede traer consigo consecuencias que no son neutras desde el punto de vista de género.

El ejemplo más claro es la automatización del trabajo. Los sistemas de IA permitirán la automatización de procesos mecánicos y repetitivos, contribuirán con la toma de decisiones y transformarán de algún modo los espacios de trabajo, si bien hay posiciones dispares. Algunos señalan que los sistemas de IA dejarán a muchísimas personas sin trabajo en un futuro inmediato, mientras que otros consideran que el proceso será más lento y que, en todo caso, siempre se necesitará a los humanos para desarrollar tareas más complejas y creativas³⁸. De

³⁶ <http://image-net.org/index>

³⁷ Collett, C. y Dillon, S. (2019). AI and gender. Four Proposals for Future Research. [Última visita 20.03.2021]. <https://www.repository.cam.ac.uk/handle/1810/294360>

³⁸ <https://www.ft.com/content/c4bf787a-d4a0-11e9-a0bd-ab8ec6435630>

todas formas, no se puede negar que la irrupción de la IA está cambiando la naturaleza de ciertos trabajos, por ejemplo, en la industria y en el mundo de la automoción³⁹.

Estos cambios afectarán especialmente a las mujeres. En 2018, el Fondo Monetario Internacional⁴⁰ estimó que 26 millones de trabajos realizados por mujeres en 30 países están en riesgo de desaparecer en las próximas dos décadas, como consecuencia de la llegada de las nuevas tecnologías. En términos generales, el 11% de los empleos ocupados por mujeres están en riesgo de automatización, en comparación con el 9% de los hombres. Se verán aún más afectadas las mujeres con bajos niveles educativos y de mayor edad, y también las que desarrollan trabajos poco cualificados en áreas administrativas, de servicio y ventas.

Esta realidad, por supuesto, será distinta en cada país, pero es probable que la tendencia sea parecida. Un estudio del Institute for Women's Policy Research⁴¹ encontró que en Estados Unidos hay una mayor representación de mujeres en trabajos con alto riesgo de automatización (como asistentes administrativos, por ejemplo), y el grupo más afectado es el de las mujeres hispanas. Pero también hay un mayor porcentaje de mujeres en puestos con bajos riesgos (como maestras y enfermeras, por ejemplo). Las mujeres, además, trabajan en mayor medida en puestos en los que se necesita algún tipo de habilidad digital, pero la fuerza laboral en áreas de alta especialización, como la computación, la programación y el desarrollo de software, sigue estando dominada por los hombres, e incluso la presencia de mujeres ha disminuido desde el año 2000.

En la medida en la que la IA siga ganando terreno, estos empleos en áreas tecnológicas avanzadas se harán cada vez más necesarios. Y aquí es donde existe otro reto para la igualdad entre hombres y mujeres: de acuerdo con el Instituto Europeo para la Igualdad de Género⁴², de los 8 millones de especialistas en Tecnologías de Información y Comunicación que hay en Europa, solo 17% son mujeres. Eso no solo tiene implicaciones para el diseño de las propias

³⁹ <https://lab.elmundo.es/inteligencia-artificial/trabajo-robots.html>

⁴⁰ Datos incluidos en el informe Gender, Technology, and the Future of Work: <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2018/10/09/Gender-Technology-and-the-Future-of-Work-46236>

⁴¹ <https://iwpr.org/iwpr-issues/employment-and-earnings/women-automation-and-the-future-of-work/>

⁴² <https://eige.europa.eu/publications/work-life-balance/eu-policies-on-work-life-balance/women-in-ict>

tecnologías, como ya señalamos previamente, sino que también incide en la posibilidad de que las mujeres ocupen puestos de trabajo mejor remunerados⁴³.

La falta de mujeres en los estudios TIC y en buena parte de las áreas STEM (Science, Technology, Engineering and Mathematics) es un fenómeno complejo, en el que inciden los estereotipos de género que se van perpetuando desde la infancia. Por eso, algunas autoras recomiendan la promoción del interés en la tecnología desde edades tempranas y las políticas de capacitación a lo largo de toda la trayectoria vital de las mujeres⁴⁴.

⁴³ Ibid.

⁴⁴ Sáinz, M., Arroyo, L., y Castaño, C. (2020). Mujeres y digitalización. De las brechas a los algoritmos. Instituto de la Mujer y para la Igualdad de Oportunidades. [Última visita 18.03.2021]. https://www.inmujer.gob.es/disenov/novedades/M_MUJERES_Y_DIGITALIZACION_DE_LAS_BRECHAS_A_LOS_ALGORITMOS_04.pdf

3. El marco europeo para una Inteligencia Artificial ética (y feminista)

3.1. Una Inteligencia Artificial europea y ética

Con base en las perspectivas y proyecciones previstas para la IA, la Unión Europea publicó su estrategia de IA, Artificial Intelligence for Europe⁴⁵, el 25 de marzo del año 2018, desde la cual invitaba a los países de la Unión a liderar conjuntamente el desarrollo de una Inteligencia Artificial de todos/as y para todos/as, enfocada al bienestar y basada en los valores y principios de la UE. Desde ese momento, la UE ha seguido emitiendo informes y declaraciones en las que reafirma su compromiso con el impulso y el desarrollo de una IA ética y en la que se pueda confiar o que sea digna de confianza (trustworthy). Además, en junio de 2018, la UE creó un Grupo de Expertos en Inteligencia Artificial (HLEG) compuesto, principalmente, por académicos, personas procedentes del sector privado y expertos en políticas públicas, enfocado en el desarrollo de propuestas para delimitar en detalle el marco de la UE respecto a la IA.

En este sentido, se han emitido una serie de documentos en los que se tratan distintas cuestiones, desde la propuesta concreta de una definición para el término Inteligencia Artificial, hasta recomendaciones de políticas públicas a nivel sectorial, pasando por los requisitos y recomendaciones que deben seguir los gobiernos para implementar tecnologías de IA que sean dignas de confianza⁴⁶. De forma adicional, la UE también ha incorporado a su estrategia de IA un observatorio denominado AI Watch, que monitoriza y evalúa los avances en IA dentro de la Unión, y más recientemente una Comisión Especial sobre Inteligencia Artificial en la Era Digital (AIDA⁴⁷). En esta última, España cuenta con la representación del eurodiputado Ibán García del Blanco⁴⁸. Según él, la comisión AIDA, creada el 18 de junio de 2020, es, a diferencia de otros comités o grupos creados con anterioridad, “un órgano con capacidad legislativa” que tiene

⁴⁵ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>

⁴⁶ Estos documentos serían los siguientes: a) A definition of Artificial Intelligence: main capabilities and scientific disciplines ; b) Ethics guidelines for trustworthy AI; c) Policy and Investment Recommendations for Trustworthy AI; d) The assessment list for Trustworthy Artificial Intelligence (ALTAI); e) Sectoral Considerations on the Policy and Investment

⁴⁷ https://www.europarl.europa.eu/doceo/document/TA-9-2020-0162_EN.html

⁴⁸ https://www.europarl.europa.eu/meps/es/197717/IBAN_GARCIA+DEL+BLANCO/home

como finalidad principal “realizar un informe omni-comprensivo sobre el fenómeno [de la IA] y sobre cuestiones que tienen que ver mucho también con la armonización de políticas y planteamientos por parte del Parlamento Europeo”. Añade que, a pesar de la propuesta de duración inicial de un año, la comisión “va a tener una duración de seis meses más”.



DOCUMENTOS SOBRE IA EMITIDOS POR LA UE



25 de marzo del 2018
ESTRATEGIA DE IA: ARTIFICIAL INTELLIGENCE FOR EUROPE*

La Estrategia Europea de Inteligencia Artificial es el primer documento oficial emitido por la UE en el que se plantea el deseo de la UE por desarrollar una estrategia conjunta para impulsar y promover el desarrollo y la implementación de la IA en Europa. En este documento el UE sienta las bases de los que luego se ha venido a conocer con enfoque "humanocéntrico" de la IA para desarrollar una IA que sea digna de confianza por parte de los ciudadanos y ciudadanas europeos.

10 de abril del 2018

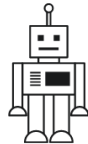
DECLARACIÓN DE COOPERACIÓN EN IA*

En este documento firmado los 25 países de la UE firmantes declaran su compromiso con el desarrollo de estrategias y planes de IA a nivel nacional. De esta forma, los países miembro se comprometen a trabajar unidos para convertir a la UE en un referente a nivel mundial en el uso ética de la IA.



7 de diciembre del 2018
PLAN COORDINADO SOBRE IA*

En el plan coordinado sobre la IA la UE propone una serie de medidas que puedan ser compartidas por los distintos países de la Unión respecto a la IA. Algunas de las propuestas incluyen: el impulso de consorcios público-privados para el desarrollo de la IA; la creación de programas educativos sobre IA; la necesidad de desarrollar un marco ético para la IA, etc.



8 de abril del 2019
BUILDING TRUST IN HUMAN-CENTRIC ARTIFICIAL INTELLIGENCE*
A DEFINITION OF ARTIFICIAL INTELLIGENCE: MAIN CAPABILITIES AND SCIENTIFIC DISCIPLINES**

Se presentan algunos de los requisitos para impulsar una IA que sea digna de confianza. Presentación extensa de la definición de Inteligencia Artificial propuesta por la UE.

Desarrollo extenso de los principios éticos impulsados por la UE para crear una IA que genere confianza.

ETHICS GUIDELINES FOR TRUSTWORTHY AI**



26 de junio del 2019
POLICY AND INVESTMENT RECOMMENDATIONS FOR TRUSTWORTHY AI**

Recomendaciones acerca de la orientación que deben tomar las políticas públicas europeas para promover una IA ética.

19 de febrero del 2020
WHITE PAPER ON ARTIFICIAL INTELLIGENCE - A EUROPEAN APPROACH TO EXCELLENCE AND TRUST*

REPORT ON THE SAFETY AND LIABILITY IMPLICATIONS OF ARTIFICIAL INTELLIGENCE, THE INTERNET OF THINGS AND ROBOTICS*

Presentación del enfoque ético de la UE sobre IA. Compromiso y directrices de la UE para continuar su camino hacia una IA de excelencia y que sea digna de confianza. Informe sobre seguridad y responsabilidad en IA.



17 de julio del 2020
THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI)**

Lista de los criterios de evaluación que propone la IA que aseguran que la IA se usa de forma ética.

23 de julio del 2020
SECTORAL CONSIDERATIONS ON THE POLICY AND INVESTMENT RECOMMENDATIONS

Recomendaciones sectoriales sobre IA.

*Documentos emitidos por la Comisión Europea
**Documentos emitidos por el HLEG

Fuente: Elaboración propia

Hasta ahora, la propuesta de la UE para impulsar y promover una IA centrada en los seres humanos y que sea digna de confianza se basa en tres pilares fundamentales: la legalidad, la ética y la robustez. La legalidad hace referencia a la base jurídica en la que se enmarca el desarrollo de la IA. Tal y como indica la propia UE, “la IA no opera en un mundo sin ley”⁴⁹. Muy al contrario, la Unión tiene importantes marcos legales vigentes, que las tecnologías de IA deben respetar y con los que debe ser compatible. Entre ellos se encuentran, aunque no se limitan a, la Carta de Derechos Fundamentales de la UE, la regulación general sobre la protección de datos, la ley de consumidores, así como la reciente legislación que la UE ha publicado de manera provisional sobre el funcionamiento y uso de la IA.

Adicionalmente, la UE ha propuesto una serie de principios éticos que deben guiar la implementación y el uso de la IA en Europa. Estos principios son: el respeto por la autonomía humana, la prevención del daño, la justicia y la explicabilidad. Cuando se habla de IA que respete la autonomía humana, se está hablando de evitar a toda costa que el uso de estas tecnologías disruptivas coarte o limite de alguna forma la libertad de elección y, por tanto, de autonomía, de los individuos. El principio de prevención del daño se asienta sobre una de las principales consignas liberales: evitar a toda costa causar dolor, tanto físico como mental, a otras personas para desarrollar un proyecto vital individual. La IA justa, o basada en el principio de justicia, se concibe tanto desde la dimensión sustantiva como la procedimental. Finalmente, la explicabilidad es uno de los principios fundamentales para garantizar una IA que sea digna de confianza ya que es una condición necesaria para la rendición de cuentas. Al mismo tiempo, la explicabilidad requiere de transparencia en los datos y los algoritmos (transparencia algorítmica). El objetivo fundamental del principio de explicabilidad es auditar, corregir y juzgar las decisiones tomadas por las tecnologías de IA. La perspectiva ética de la IA es uno de los ejes vertebradores del enfoque de la UE en IA y un punto diferencial respecto a otras grandes potencias como China o EEUU. En este sentido, Ibán García del Blanco señala que la posición de la Unión Europea es clara a este respecto, “a nosotros [la UE] no nos vale cualquier cosa. A nosotros no nos vale que una administración pública pueda utilizar tecnologías de reconocimiento facial a distancia, por ejemplo, sin tener un control previo” afirma el

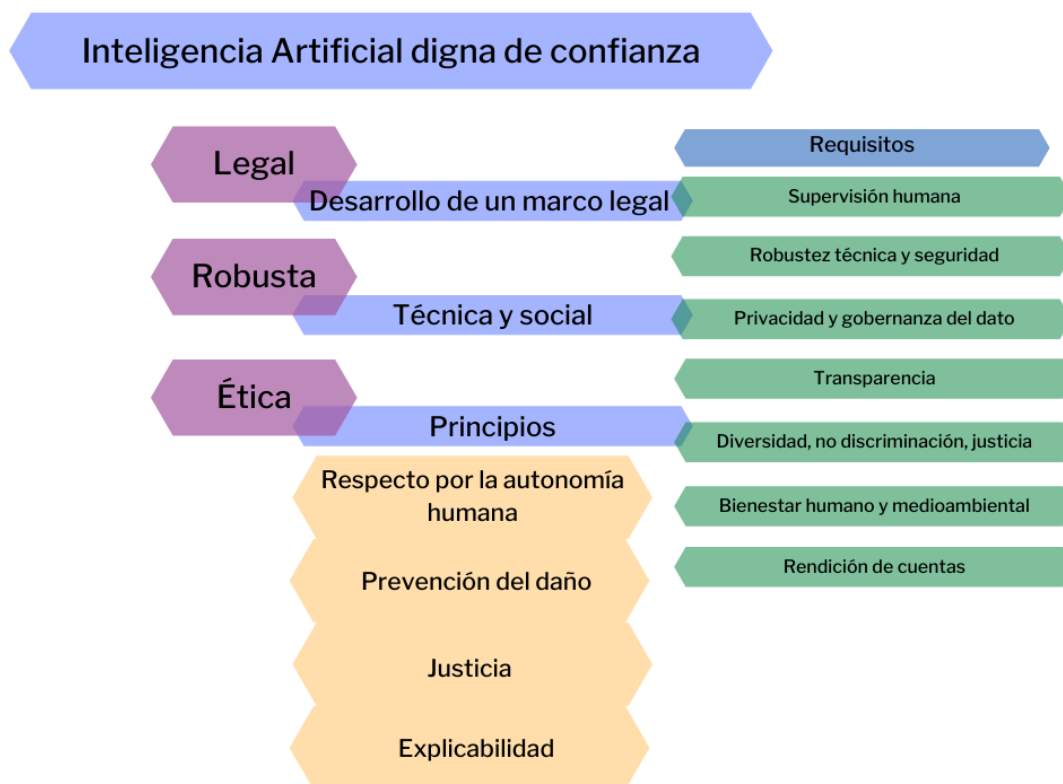
⁴⁹ HLEG. (2019a). A definition of Artificial Intelligence: main capabilities and scientific disciplines. <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>

eurodiputado. Por ello señala que “Europa se plantea primero que lo más importante es tener claras las cuestiones relacionadas con derechos fundamentales y, por tanto, con el proyecto civilizatorio que nosotros pretendemos representar”.

Finalmente, la robustez como otro de los pilares de una IA digna de confianza, hace referencia a la necesidad de que estas nuevas tecnologías disruptivas no causen ningún tipo de daño a la ciudadanía europea. No solo la IA debe tener un asiento legal y ser ética, sino que también debe ser segura y fiable, es decir, robusta, tanto desde un punto de vista técnico (en referencia a las capacidades y componentes técnicos de la IA), como social (que tenga en cuenta el ambiente y el contexto en el que opera la IA).

Para garantizar que la IA se desarrolla y se usa de tal modo que genere confianza, es decir, cumpliendo los tres pilares previamente expuestos, la UE ha propuesto de forma provisional⁵⁰ los siguientes requisitos y requerimientos que sirven como criterios de evaluación para asegurar que la IA funcione adecuadamente. Estos requisitos son: la constante supervisión humana de todos aquellos procesos, acciones y decisiones llevadas a cabo por IA; garantizar la robustez técnica y la seguridad de la tecnología; establecer leyes y mecanismos que salvaguarden la privacidad de la ciudadanía europea y la gobernanza ética del dato; la transparencia; asegurar la diversidad, la no discriminación y la justicia en los datos, algoritmos, y los acontecimientos derivados del uso de IA; buscar el bienestar humano y medioambiental como fines principales de la IA y garantizar el uso de tecnologías que puedan rendir cuentas.

⁵⁰ En este sentido se contempla la posibilidad de modificar y/o ampliar los requisitos actuales.



Fuente: Elaboración propia, con base en documentos oficiales europeos

3.2. Criterios para la evaluación y el análisis de la perspectiva de género

Para analizar de manera adecuada el enfoque de la UE respecto a la igualdad de género y su relación con la IA, en el presente trabajo se ha seguido el enfoque planteado por la UNESCO⁵¹. De este modo, el análisis documental se basa en la identificación tanto de referencias explícitas como implícitas a la igualdad de género. Por referencias explícitas se entienden todas aquellas en las que textualmente aparezcan formulaciones del tipo: IA feminista, igualdad entre hombres y mujeres, promoción del empoderamiento femenino, perspectiva de género, etc. A su vez, las referencias implícitas incluyen referencias a la co-construcción de la IA, el diseño humano de la IA, la perspectiva centrada en los seres humanos, la promoción de la justicia, la igualdad, la equidad, la representatividad, la no discriminación, la reducción de la brecha digital, etc.

⁵¹ UNESCO. (2020). Artificial Intelligence and Gender Equality. <https://unesdoc.unesco.org/ark:/48223/pf0000374174>

El caso de las referencias explícitas no requiere de mayor explicación. Sin embargo, en el caso de las referencias implícitas es necesario aclarar que han sido incluidas porque se considera que, aún sin hacer referencia explícita a la igualdad de género o a las mujeres como colectivo discriminado, su realización implica la reducción de la desigualdad entre hombres y mujeres. Es decir, si hablamos de promover la justicia, garantizar que los datos sean representativos o reducir la brecha digital, se entiende que la consecución de estos objetivos sólo es posible mediante la promoción activa, entre otras cosas, de la igualdad entre hombres y mujeres. No hay justicia si hombres y mujeres no son tratados en igualdad de condiciones, o si las mujeres sufren discriminación a razón de su sexo; la brecha digital no desaparecerá hasta que el acceso y el uso de la tecnología entre hombres y mujeres no sea igualitario; la representatividad de los datos necesita de la inclusión de distintas perspectivas y colectivos de entre los cuales las mujeres tienen un papel esencial, etc. Por todo ello, consideramos las previas formulaciones, o similares, como referencias implícitas a la igualdad de género.

3.3. Una Inteligencia Artificial europea y feminista

Para entender si la UE tiene un enfoque que posibilite la perspectiva feminista de la IA es necesario prestar atención a la definición que este organismo supranacional ofrece de la IA. Esto es importante porque tradicionalmente la tecnología, en sus distintas formas, se ha conceptualizado como una realidad puramente técnica. Es decir, como si hasta cierto punto la tecnología tuviera lugar en una esfera aislada de la sociedad, de tal modo que su diseño, construcción y uso fueran objetivos y neutrales desde un punto de vista social. Con el paso del tiempo esta visión de la tecnología, como ha sucedido en el caso de la ciencia, ha ido cambiando. Actualmente, la visión de que la tecnología es una realidad humanamente y, por tanto, social, cultural, política y económicamente determinada, está muy extendida y es compartida por gran parte (en mayor o menor medida) de la comunidad académica. Es importante señalar que la visión de que la tecnología es objetiva y neutral sigue estando bastante extendida entre gran parte de la población y los tecnólogos.

Aunque este debate en torno a cómo se conceptualiza la tecnología y su respectiva definición puede parecer puramente académico, se trata de una cuestión muy relevante que tiene consecuencias reales en la vida de los/as ciudadanos/as. Por ejemplo, si la IA se entiende desde un punto de vista puramente tecnológico, entonces sería prácticamente imposible analizar su

papel como factor determinante en el desarrollo y condicionamiento de distintas cuestiones sociales. Solo si consideramos la IA desde una perspectiva socio-técnica, es decir, considerando que que la tecnología es una realidad condicionada por los seres humanos y que, por tanto, en ella se ven reflejados todos nuestros prejuicios, normas y costumbres, es posible abordar la IA como un elemento más que marca y condiciona el desarrollo de ciertas dinámicas sociales y culturales de forma decisiva. En esta línea, el eurodiputado Ibán García del Blanco señala que el plantearse “qué es la inteligencia artificial dentro de un determinado concepto tiene unas consecuencias enormes desde la perspectiva de la regulación que se le aplica”.

Dicho esto, la última definición que ha propuesto la UE para la IA, a través de su grupo de expertos, es la de “sistemas de software (y posiblemente también hardware) diseñados por humanos”⁵². En este sentido, entendemos que la condición que posibilita un enfoque feminista de la IA se cumple y que la UE y su comité de expertos reconoce de forma mayoritaria la IA como una tecnología humanamente condicionada. De manera consecuente con este enfoque socio-técnico de la IA, la UE ha remarcado en varios de sus documentos oficiales la necesidad de que la IA se base en los principios y derechos recogidos en la Carta de Derechos Fundamentales de la UE. Esta referencia a la Carta de Derechos Fundamentales se entiende como una referencia tanto explícita como implícita hacia la consecución de una IA feminista, ya que el capítulo tercero versa sobre la igualdad. Dentro de este capítulo podemos encontrar tanto referencias implícitas como explícitas a la igualdad de género. En el artículo 20 se habla de la “igualdad de todos ante la ley”, donde todos (en inglés everyone) hace referencia tanto a hombres como a mujeres; el artículo 21 aborda la no discriminación por razones de sexo, raza, origen social, etc.; y el artículo 23 habla explícitamente de la igualdad entre hombres y mujeres. Por todo ello, consideramos que desde el punto de vista de sus fundamentos, la UE contiene las bases de un enfoque feminista de la IA.

En lo que se refiere a sus objetivos marcados, la UE ha señalado en distintas ocasiones su voluntad de usar la IA para mejorar el bienestar y la libertad humana, así como “facilitar el cumplimiento de los Objetivos de Desarrollo Sostenible establecidos por Naciones Unidas,

⁵² HLEG. (2019a). A definition of Artificial Intelligence: main capabilities and scientific disciplines. <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>, p. 6.

entre los que se encuentra la promoción de la igualdad de género”⁵³. También encontramos referencias implícitas en torno a la igualdad entre hombres y mujeres a través de numerosas manifestaciones en las que se expresa el deseo de que la tecnología contribuya a crear sociedades más justas e igualitarias. En este sentido, el eurodiputado Ibán García del Blanco afirma que la IA es “la herramienta [tecnológica] más potente que tenemos a nuestra disposición” para luchar contra la desigualdad entre hombres y mujeres. De este modo, afirma, “si se tiene en cuenta un control previo sobre el sesgo en sí mismo, se está produciendo una mejora en las condiciones. Se trata del prejuicio que podría tener un ser humano y que, por lo tanto, podría reproducir tomando decisiones injustas y, por cierto, económicamente ineficientes; una tecnología con la autonomía de las características de la Inteligencia Artificial, que esté controlada de base, lo que puede producir al contrario, o por contrario, es una distribución mucho más justa. [...] Como digo, tenemos que analizar que somos una sociedad desigual y tenemos que corregirlo”. Finalmente, en lo que se refiere a la parte propositiva, actualmente la UE no ha formulado ningún tipo de propuesta y/o leyes con carácter vinculante, aunque ese momento parece encontrarse cerca. Lo que sí ha hecho la UE es proponer políticas de carácter general con el fin de que los miembros de la Unión las concreten dentro de sus propias propuestas nacionales y sus marcos legislativos. Algunas de sus propuestas tanto a nivel técnico como no técnico son:

⁵³ HLEG. (2019b). Ethics guidelines for trustworthy AI <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, p. 4.



Fuente: Elaboración propia

Aunque ninguna de las propuestas mencionadas hace referencia explícita a la igualdad de género, todas ellas en su desarrollo (en el que por cuestiones de extensión nos es imposible abordar en este documento) incluyen referencias implícitas a la igualdad entre hombres y mujeres.

Partiendo del marco propuesto por la UE, centrado en el impulso y el desarrollo de una IA que sea digna de confianza y centrada en las personas, los países y miembros de la Unión han ido desarrollando sus propias Estrategias Nacionales de IA. En este trabajo presentamos las propuestas planteadas por España y Suecia. La elección de España como caso de estudio se debe al interés de la presente propuesta por determinar el estado de nuestro país en relación con la IA, el enfoque ético desde el cual lo enmarca, así como la forma en la que España pretende abordar la relación entre IA y género. La selección del caso de Suecia se debe a que se trata del país con el índice de igualdad de género más alto de la UE de acuerdo con el

Instituto Europeo para la Igualdad de Género⁵⁴. Por ello, consideramos que, en la medida que España desee usar la IA en favor de la igualdad de género, debe medir sus planes y medidas en IA no solo con base en los criterios de la UE, sino también con miras a seguir la estela del país nórdico.

4. El caso de Suecia como referente en la igualdad de género

La consecución de la igualdad de género es una cuestión prioritaria en la agenda política y social de Suecia. El país nórdico cuenta con 83.8 sobre un posible total de 100 puntos en el Índice de Igualdad de Género desarrollado por la UE, situándose así a la cabeza de los países de Unión, que cuenta con una media 64 puntos, y como referente en igualdad a nivel mundial. Como muestra de su compromiso, Suecia ha sido uno de los países pioneros en medidas fundamentales para caminar hacia la igualdad real y efectiva entre hombres y mujeres. En este sentido, Suecia fue uno de los primeros países del mundo en legalizar el sufragio femenino, el aborto, los permisos de maternidad y paternidad iguales e intransferibles, etc. Además, se trata de uno de los países con mayor cuota de mujeres en puestos de responsabilidad política y social, así como en consejos de administración de grandes empresas. Desde principios de los años 70 del s.XX, Suecia ha incluido el término igualdad de género como un eje central en el desarrollo de sus políticas públicas. Es importante señalar que gran parte del éxito de Suecia en la reducción de la desigualdad de debe al reconocimiento temprano, a nivel político, de que la igualdad de género consiste en más que la simple distribución igualitaria de puestos o espacios entre hombres y mujeres, sino que también, y principalmente, se ha centrado en todos aquellos aspectos que se refieren a las actitudes, normas e ideas que afectan la vida de las mujeres de forma diferencial.

Sin embargo, aún quedan muchos ámbitos en los que el país nórdico debe mejorar para alcanzar la igualdad de género. En este sentido, desde el año 2006 el gobierno sueco lleva trabajando en el desarrollo y cumplimiento de líneas estratégicas y metas concretas que le permitan disminuir la desigualdad entre hombres y mujeres. Actualmente, existen seis metas con ejes centrales enfocados en alcanzar la igualdad de género: la división igualitaria de la influencia y el poder, la igualdad educativa, la distribución igualitaria de los cuidados y el trabajo

⁵⁴ <https://eige.europa.eu/gender-equality-index/2020>

no remunerado, la igualdad económica, acabar con la violencia de género, y el acceso y disfrute igualitario a la salud⁵⁵. Teniendo como referencia este enfoque, Suecia ha tomado la delantera a la propia UE y la mayor parte de países del mundo en el desarrollo de una IA con perspectiva de género que ayude a detectar y reducir, de manera proactiva, la discriminación por cuestiones de género.



Fuente: Elaboración propia.

4.1. Suecia y la Inteligencia Artificial: un caso precoz

En mayo de 2018 Suecia lanzó su Estrategia Nacional de Inteligencia Artificial⁵⁶, prácticamente coincidiendo con la publicación de la Estrategia Europea de IA, en marzo de 2018. La Estrategia Nacional de IA sueca busca desarrollar un marco que permita hacer un uso sostenible de la IA.

⁵⁵ Más información disponible en What is gender equality? <https://www.jamstalldhetsmyndigheten.se/en/about-genderequality/what-is-gender-equality>.

⁵⁶ Gobierno de Suecia. (2018). National Approach for Artificial Intelligence. [Última visita 16.03.2021] <https://www.government.se/4a7451/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf>

Para ello ha definido cuatro áreas prioritarias: educación y formación; investigación, innovación y uso; marcos e infraestructuras. Es importante remarcar que frente al uso del término “digna de confianza” (trustworthy), propuesto por la UE para referirse a la IA europea, Suecia se ha decantado por el adjetivo sostenible. En este caso, la IA sostenible se entiende como una IA ética, transparente, segura y fiable.

En relación con las cuatro áreas identificadas por el gobierno sueco para hacer uso de una IA sostenible, la primera área destacada es la educación y formación. A este respecto, Suecia enfatiza en la necesidad de educar a su población en conocimientos, tanto técnico como sociales, sobre IA. Como parte de esta estrategia, se hace especial hincapié en la promoción del estudio de carreras universitarias centradas en el campo de la IA, tanto desde el ámbito de las ciencias como de las ciencias sociales y las humanidades. También se apuesta por la formación de profesionales con perfiles mixtos, principalmente de programadores versados en conocimientos de carácter humanístico para que así puedan hacer frente a los desafíos éticos que conlleva el uso de la IA. Otra de las grandes apuestas de Suecia en este ámbito se centra en la creación de vínculos entre distintos sectores que trabajen o estén involucrados en el campo de la IA como, por ejemplo, la investigación, la educación superior y las organizaciones y empresas, tanto públicas como privadas, centradas en la innovación.

En el ámbito de la investigación, Suecia apuesta por afianzar, fomentar y aumentar la investigación sobre IA para garantizar la existencia de una masa crítica sobre el tema. Para ello se ha establecido como elemento prioritario dentro de su estrategia la necesidad de establecer vínculos y relaciones de colaboración entre las agencias de innovación e investigación suecas y sus homólogos internacionales; así como entre el sector privado y el sector público. Respecto a la innovación y uso en la IA, Suecia busca poner en marcha cuanto antes proyectos piloto y aplicaciones de IA en fase de prueba, tanto en el sector privado como el sector público, para así testar, corregir e implementar la IA de forma segura y responsable en el país. En este sentido, el país nórdico se autopercibe como un actor privilegiado debido al avanzado nivel de digitalización de sus administraciones públicas. Este elemento distintivo y definitorio hace que Suecia cuente con una mayor cantidad de datos disponibles que otros países de su entorno para poner en marcha y probar propuestas que hagan uso de la IA. En su estrategia de innovación en IA, Suecia apuesta por la creación de vínculos y partenariados con otros países

de la UE, así como en el desarrollo constante y conjunto de mecanismos que permitan prevenir los riesgos y malos usos de la IA.

Finalmente, en relación con su cuarta área estratégica, los marcos y las infraestructuras, Suecia considera como una parte fundamental para el uso sostenible de la IA el desarrollo y la elaboración de reglas, estándares, normas y principios éticos que guíen el uso de la IA. Para ello, pretenden basarse en su propia legislación, así como en los estándares y regulaciones internacionales, principalmente los promovidos por la UE. En este sentido, se apuesta por la continua revisión y evaluación de las infraestructuras y los recursos tecnológicos para asegurar el desarrollo de todas las potencialidades ofrecidas por la IA, así como la digitalización a gran escala para así disponer de una mayor cantidad de datos (y de más calidad) que permitan el funcionamiento de la IA de forma que añada valor. Finalmente, Suecia busca jugar un papel activo y estratégico en el desarrollo y la promoción de la IA a nivel europeo.

Cabe destacar que en la Estrategia Nacional de IA sueca publicada en inglés, no se hace ninguna referencia explícita al género. Sin embargo, esta estrategia no se entiende sin el informe previo elaborado por la Agencia Sueca de Innovación (Vinnova⁵⁷) y publicado en febrero de ese mismo año. En ese informe titulado *Artificial intelligence in Swedish business and society. Analysis of development and potential*, se define la IA como una forma de referirse a “las habilidades de una máquina para imitar el comportamiento humano”⁵⁸, y se señalan las principales oportunidades y los retos que la IA trae a las empresas y los servicios públicos suecos, los puntos fuertes y débiles que presenta Suecia para desarrollar la IA, así como una lista de requisitos para hacer realidad el uso de la IA de manera sistemática, potenciales áreas de aplicación y algunos ejemplos de aplicaciones de IA que ha llevado a cabo Suecia.

⁵⁷ <https://www.vinnova.se/en/>

⁵⁸ Innova. (2018). *Artificial intelligence in Swedish business and society. Analysis of development and potential*. https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf, p. 7.

4.2. Inteligencia Artificial y género

El enfoque de género aplicado a la IA en Suecia se resume en un informe publicado también por Vinnova, esta vez junto a la consultora sueca Ramboll, bajo el título *AI for Gender Equality. Addressing inequality through AI*. A la hora de abordar las relaciones entre IA y género, Suecia ha clasificado las problemáticas con base en los dos componentes esenciales de la IA: los datos y los algoritmos; y se ha centrado, principalmente, en el problema de los sesgos presentes en ambos elementos.

En relación con los sesgos de los datos, el enfoque de género se divide en dos partes. Primero, la falta de representatividad en los datos debido al menor acceso que tienen las mujeres a Internet, la falta de diversidad en el estatus socio-económico y el carácter étnico y/o racial de las mujeres, etc. Segundo, la naturalización de los propios sesgos de los datos a través de los cuales se quedan representados y, a falta de acciones correctoras, perpetuados roles, normas y actitudes de género perjudiciales para las mujeres. Para ilustrar este problema, en el informe se hace referencia a problemas reales como, por ejemplo, los prejuicios de género en traducciones del finés, idioma con pronombres de género neutros, al inglés realizadas por algunas tecnologías de IA. En el caso de Google Translate se observó que cuando se traducía la frase “X is a doctor” en finés al inglés, la traducción resultante decía “Él es un doctor”. La misma formulación con el caso de la enfermería, “X is a nurse”, devolvía “Ella es una enfermera”, perpetuando así roles y estereotipos de género muy marcados⁵⁹.

En el caso de los sesgos en el diseño de los algoritmos, los problemas apuntados por Suecia se encuentran en la línea de los ya mencionados anteriormente para el caso de la UE. Además de los sesgos en los datos, si no se aplican criterios correctores en el diseño de los algoritmos por parte de los programadores, entonces estos podrían no solo perpetuarse, sino incrementarse. A modo de ejemplo, el informe hace referencia al caso JAMA en el que se detectó que asistentes de voz de Apple, Samsung, Google y Microsoft no entendían palabras como “violación” o “mi marido me está pegando”. En vez de llamar a emergencias, los asistentes

⁵⁹ Prates, M. O., Avelar, P. H., & Lamb, L. C. (2019). Assessing gender bias in machine translation: a case study with Google Translate. *Neural Computing and Applications*, 1-19.

ponían en marcha búsquedas en páginas webs o incluso, en algunos casos, realizaban bromas o se reían de los usuarios⁶⁰.

Potenciales usos de la IA para alcanzar la igualdad de género

Teniendo en cuenta los tipos de IA, las capacidades que cada uno de ellos presentan y con base en las seis metas de igualdad de género planteadas por el Gobierno de Suecia, se proponen los siguientes posibles (algunos de ellos actualmente en marcha) usos para acabar con la discriminación por razones de género a través de la IA. En las propuestas que se exponen a continuación se observan tanto usos de la IA enfocados en la cuantificación y medida de la discriminación que sufren las mujeres, como medidas de carácter proactivo para reducir la brecha de género:

- **División igualitaria de la influencia y el poder:** Algunas de las formas en las que Suecia propone que la IA puede ayudar a distribuir de manera más igualitaria la influencia y el poder entre hombres y mujeres son, por ejemplo, el análisis de imágenes de vídeo grabadas en reuniones para mostrar y analizar el modo en el que se interactúa en función del género del hablante. También se propone el análisis de audio y discursos para medir con precisión el tiempo que hablan mujeres y hombres, quién interrumpe más y en qué dirección. Análisis de textos para detectar comportamientos sexistas en el lugar de trabajo. Sistemas de IA que permitan reducir los sesgos en los procesos de selección del personal.

En este sentido Vinnova ha financiado el proyecto Ceretai en el que se desarrollan herramientas automatizadas y diseñadas para la detección de normas, patrones y estereotipos discriminatorios presentes en la cultura popular. De esta forma, se busca cuantificar y visibilizar comportamientos discriminatorios que sufren las mujeres y que, con frecuencia, pasan inadvertidos. Estas tecnologías también se utilizan para asesorar a personas dentro de la industria del cine y así evitar la reproducción de estereotipos discriminatorios.

⁶⁰ Miner, A. S. (2016) Smartphone-Based Conversational Agents and Responses to Questions about Mental Health, Interpersonal Violence and Physical Health. JAMA Internal Medicine 176, no 5 (2016) 619-25.

Vinnova también financia NoBias, un proyecto que busca desarrollar soluciones para ayudar a las compañías y los/as empleados/as sin experiencia o conocimientos previos en cuestiones de género, a expresarse a través del uso de un lenguaje inclusivo. Estas soluciones basadas en el machine learning eliminan los prejuicios y las normas a través de explicaciones que muestran usos no inclusivos y discriminatorios del lenguaje.

- **Igualdad educativa:** Análisis del lenguaje no verbal en las clases a través de imágenes de vídeo. Análisis de audios para estudiar las diferencias en las intervenciones entre chicos y chicas, si existen patrones, si el profesor pregunta con más frecuencia a unos o a otras, cómo es el tono y la forma de las preguntas formuladas. Usar la IA para evaluar tareas y exámenes de modo que se eviten sesgos por género. IA para detectar casos de bullying desde momentos tempranos. Usar la IA para evitar sesgos de género en los procesos de admisión en las universidades.

En educación, Suecia ha desarrollado diferentes proyectos como Gradecam, enfocado en el análisis de texto para evaluar exámenes y tareas, o Cognii's Virtual Learning, un chat que sirve para poner en contacto a estudiantes con IA que actúa como "profesores" y les ayuda en sus tareas.

- **Distribución igualitaria de los cuidados y el trabajo no remunerados:** Analizar imágenes de vídeo procedentes de las cámaras de los supermercados para determinar el género y la edad de las personas que compran y las diferencias entre los productos que se llevan. Analizar datos públicos para detectar tendencias y actitudes en la distribución de los cuidados y el trabajo no remunerado.
- **Igualdad económica:** A través del análisis de audio y de discursos se pueden detectar diferencias en las preguntas que se le realizan a directivos de distinto sexo y así determinar cómo estas formas de tratar a unos y otras tienen un impacto en los procesos de tomas de decisiones. Analizar bases de datos de las empresas para identificar desigualdades en el salario entre hombres y mujeres.

Rikare II es un proyecto financiado por Vinnova que busca desarrollar un algoritmo que tome decisiones respecto a la concesión de financiación para proyectos emprendedores y de negocios. El objetivo es reducir los sesgos de género en la concesión de financiación para proyectos de áreas tradicionalmente masculinas.

- **Acabar con la violencia de género:** Análisis de imágenes de vídeo para detectar comportamientos agresivos o violentos hacia las mujeres. Analizar las llamadas de emergencia realizadas por mujeres con base en el tono de su voz, el lenguaje utilizado, así como otros patrones auditivos y discursivos. Identificar publicaciones en redes sociales que contengan contenido de acoso a las mujeres, fomento o promoción de la prostitución y/o prácticas similares, etc. Determinar la probabilidad de que un hombre que se encuentra en prisión por delitos de violencia de género reincida.

Nibye es una organización sueca que está desarrollando un reloj para detectar una agresión a una mujer en tiempo real, de tal manera que sea notificada automáticamente a un amigo o familiar. Además, la alarma aporta información sobre la localización de la víctima y si la alarma ha sido accionada de forma manual o automática.

- **Acceso y disfrute igualitario a la salud:** Análisis de imágenes para detectar el cáncer de mama en fase temprana. Usar audios para analizar las interacciones médico-paciente y así detectar posibles diferencias de trato y diagnóstico con base en el género. Analizar publicaciones en redes sociales para estudiar el tratamiento público que se hace a la salud que atañe a cuestiones puramente femeninas. Desarrollo de aplicaciones que versen específicamente sobre salud femenina y que permita responder preguntas a las usuarias.

Para mejorar la salud de las mujeres, Innova ha financiado el proyecto Grace Health que tiene como objetivo el desarrollo de un chatbot para ayudar a las usuarias a entender mejor su ciclo menstrual, días de fertilidad, resolver dudas sobre métodos anticonceptivos y otras cuestiones relacionadas con la higiene íntima femenina. Otro caso de IA aplicado a la salud femenina, también desarrollado en Suecia, es Bonzun, una aplicación, que, a través de la figura de una mujer de mediana edad, acompaña a las

mujeres durante su embarazo y comparte con ellas información relevante. Lleva la cuenta de sus síntomas y visitas al médico, y explica los resultados de los tests y progresos en el desarrollo del feto a través de su movimiento, crecimiento y la presión sanguínea de la madre.

Para que estos usos de la IA sean posibles cuanto antes, Suecia también ha propuesto una serie de factores habilitadores de la IA con perspectiva de género. Entre ellos se encuentran: la generación de sinergias de cooperación entre aquellas personas que sufren el problema, emprendedores/as, programadores/as de IA y expertos/as en cuestiones de género; la creación de equipos de desarrollo de IA que sean competentes y diversos; usar estructuras económicas basadas en incentivos; incluir una perspectiva crítica en todos los procesos de desarrollo. Otros habilitadores de carácter más general, pero que también importan en la cuestión del género son: el acceso a datos relevantes y la regulación de los procesos de innovación.

5. España, Inteligencia Artificial e igualdad de género

5.1 España: una IA con sello europeo

España ha seguido muy de cerca los lineamientos europeos en el desarrollo de su enfoque sobre IA. Después de firmar la Declaración de Cooperación en Inteligencia Artificial de la UE, en 2018, España creó un Grupo de Trabajo en IA, conformado por especialistas procedentes tanto del sector público como del privado. Este grupo desarrolló la Estrategia Española I+D+I en Inteligencia Artificial, publicada por el Ministerio de Ciencia, Innovación y Universidades en 2019, y que ha servido como punto de partida para los documentos oficiales que han surgido después. De esta manera, España cumplió con el compromiso establecido en el Plan Coordinado de la IA de la UE, que recomendaba a los Estados Miembros desarrollar sus estrategias en esta materia hasta mediados de 2019⁶¹.

En este documento se hace evidente la preocupación de España por desarrollar una IA en línea con los valores europeos y con el propósito de acelerar el cumplimiento de los Objetivos de Desarrollo Sostenible que forman parte de la Agenda 2030. En este sentido, se establece que el desarrollo tecnológico, las infraestructuras nacionales y los mecanismos de colaboración entre las entidades debe acompañarse de un compromiso con la ética, los valores humanos y los derechos fundamentales, como toda “IA hecha en Europa”. Se propone que el Comité Español de Ética de la Investigación tenga un rol fundamental, y que se desarrolle un trabajo de cooperación internacional para definir un Código Ético de la IA, en línea con los desarrollos europeos.

Durante el año 2020, en plena crisis por la pandemia de la COVID-19, se publicaron dos documentos importantes para el desarrollo de la IA en España: en julio, la Estrategia Digital España 2025, y en noviembre, la Estrategia Nacional de Inteligencia Artificial. En el primer caso, se trata de una política de Estado que sirve de marco general para el avance de la transformación digital en España, en línea con la estrategia europea establecida en la comunicación Dar forma al futuro digital de Europa⁶². En este sentido, España reconoce la

⁶¹ Hay que señalar, sin embargo, que este documento se consideró una primera aproximación y un punto de partida para la elaboración de la Estrategia Nacional de Inteligencia Artificial, publicada a finales de 2020.

⁶² https://ec.europa.eu/commission/presscorner/detail/es/ip_20_273

importancia de una “transformación digital que redunde en beneficio de todos, y refleje lo mejor de Europa: abierta, justa, diversa, democrática y digna de confianza”⁶³. Y, en esta misma senda, el documento español considera que la transformación digital debe servir de palanca para el crecimiento económico, la reducción de la desigualdad y el aprovechamiento de las oportunidades, sin perder el compromiso con la Constitución y los valores europeos.

Si bien el conjunto de 10 medidas⁶⁴ para la transformación digital influirá en la implementación de la IA, una de ellas (la número 9) se refiere explícitamente a la necesidad de favorecer la economía del dato, al tiempo que se garantiza la seguridad y la privacidad, y se aprovechan los sistemas de IA. La meta, en este sentido, es que para 2025 exista un 25% de empresas usando IA y big data. El documento señala que España apoya la iniciativa europea de crear un ecosistema de excelencia para la economía del dato y la IA, en el que no solo estén presentes los Estados Miembros, sino también el sector privado, con la intención de desarrollar e implementar sistemas de IA centrados en los seres humanos. Para lograr la meta, se propusieron cuatro medidas: a) La publicación de la Estrategia Nacional de IA; b) La creación de la Oficina del Dato y el Chief Data Officer, que ya están conformados⁶⁵; c) La creación del Consejo Asesor de IA, ya constituido⁶⁶; y d) La participación en los espacios compartidos de la European Cloud Federation, que permite el desarrollo conjunto de una infraestructura europea de datos.

La Estrategia Nacional de Inteligencia Artificial, publicada en noviembre de 2020, ofrece una aproximación más concreta del enfoque propuesto en los documentos anteriores, con medidas

⁶³ La traducción oficial al español utiliza el término “confianza en sí misma”, pero aquí empleamos “digna de confianza”, que consideramos que expresa mejor el significado de trustworthy: https://ec.europa.eu/commission/presscorner/detail/es/ip_20_273

⁶⁴ De forma sintética, las 10 medidas que se proponen en el documento España Digital 2025 son: a) Garantizar la conectividad digital del 100% de la población; b) Seguir liderando en Europa el despliegue de la tecnología 5G; c) Reforzar las competencias digitales de los trabajadores y la ciudadanía en su conjunto; d) Reforzar la capacidad española en ciberseguridad; e) Impulsar la digitalización de las Administraciones Públicas; e) Acelerar la digitalización de las empresas; f) Acelerar la digitalización del modelo productivo con proyectos de transformación sectorial; g) Mejorar el atractivo de España como plataforma de negocio, trabajo, e inversión en el ámbito audiovisual; h) Favorecer el tránsito hacia la economía del dato; i) Garantizar los derechos de la ciudadanía en el entorno digital.

⁶⁵ Orden ETD/803/2020, de 31 de julio, por la que se crea la División Oficina del Dato y la División de Planificación y Ejecución de Programas en la Secretaría de Estado de Digitalización e Inteligencia Artificial. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-10008

⁶⁶ <https://www.mineco.gob.es/portal/site/mineco/menuitem.ac30f9268750bd56a0b0240e026041a0/?vgnnextoid=51884ba89bc63710VgnVCM1000001d04140aRCRD>

para el diseño y la implementación de una IA que esté al servicio de la ciudadanía y que sea acorde a los valores europeos. Hay medidas para impulsar la investigación científica y la innovación en IA, promover las capacidades y los talentos, desarrollar infraestructuras, transformar la economía, garantizar el uso de la IA en las administraciones públicas, y desarrollar un marco ético y normativo que permita proteger los derechos y garantizar el bienestar social. Si bien el documento ofrece una lista de objetivos concretos, podríamos destacar los siguientes elementos que guían toda la estrategia:

- El liderazgo de los poderes públicos: En las páginas iniciales del documento se destaca la importancia que deben tener los poderes públicos en el diseño y la implementación de la IA. Aquí pueden identificarse dos aspectos. En primer lugar, está el rol de las instituciones públicas para establecer las rutas generales que permitan la creación de sistemas de IA en línea con la ética, los derechos humanos y los valores europeos. En este sentido, el Estado tiene un rol vertebrador de las políticas y la colaboración entre el sector público y el sector privado, para que la IA no solo se implemente de forma efectiva, sino que se enmarque en los principios éticos definidos por la UE y España. En segundo lugar, la IA tiene un impacto importante en la digitalización de las propias administraciones públicas. El documento señala que estas tecnologías tienen el potencial de aumentar la productividad de los servicios públicos, las interfaces de relación con las personas, la gestión interna, el proceso de toma de decisiones, la transparencia y, entre otras cosas, la definición y puesta en marcha de políticas públicas. “España busca que la IA contribuya a consolidar nuestro estado de bienestar, aportando a su vez los datos y activos necesarios para impulsar la innovación y un desarrollo tecnológico por y para la sociedad, en un círculo virtuoso entre la tecnología y nuestro sistema político, social, económico e industrial”⁶⁷.
- IA para transformar la economía y la sociedad: De acuerdo con el documento de la estrategia, el Gobierno español no considera a la IA como una herramienta aislada, que solo tiene efectos en ciertos sectores como la investigación científica, la industria y la

⁶⁷ Ministerio de Asuntos Económicos y Transformación Digital (2020b). Estrategia Nacional de Inteligencia Artificial. [Última visita 20.03.2021]. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>, p. 9.

innovación. Como señala el texto, “existe una gran oportunidad para aplicar la IA como un elemento de transformación de la economía y la sociedad”⁶⁸. Es decir, no solo se asume el compromiso de proteger los derechos de la ciudadanía con una actitud vigilante hacia posibles sesgos y daños que se deriven de la implementación de la IA, sino que se entiende que estos sistemas pueden funcionar de forma activa para afrontar los desafíos de la sociedad, entre los que se cuentan la brecha de género en el empleo y liderazgo en IA, la transición ecológica, la cohesión y organización del territorio, y la reducción de la brecha digital. Esta perspectiva se enmarca con la promoción de los valores humanistas en la IA, que se incluye como uno de los objetivos de la estrategia, y que consiste en el impulso del debate sobre el desarrollo tecnológico con un enfoque en los seres humanos.

Otro aspecto destacable de este documento, que aporta un sello nacional a la estrategia, es que propone la proyección de la lengua española en la IA. Se plantea la necesidad de impulsar el liderazgo del país en el desarrollo de sistemas de IA en español, sobre todo considerando la conexión con Latinoamérica. El documento menciona el Plan de Impulso de las Tecnologías del Lenguaje, que fue publicado en 2015⁶⁹, y que promueve el desarrollo del procesamiento de lenguaje natural, la traducción automática y los sistemas de conversación en español y en las lenguas cooficiales. Más recientemente, en el año 2019, la Real Academia Española anunció el lanzamiento del proyecto Lengua Española e Inteligencia Artificial (LEIA), desarrollado en conjunto con el sector privado, que busca fomentar el buen uso del español en las aplicaciones de IA⁷⁰.

⁶⁸ Ibid., p.11

⁶⁹ <https://plantl.mineco.gob.es/tecnologias-lenguaje/PTL/Paginas/plan-impulso-tecnologias-lenguaje.aspx>

⁷⁰ <https://www.rae.es/noticia/la-rae-presenta-el-proyecto-lengua-espanola-e-inteligencia-artificial-leia-en-el-xvi>



Fuente: Elaboración propia

4.2. Género e Inteligencia Artificial en España

Ahora, ¿los documentos oficiales españoles incluyen la perspectiva de género? Y, si es así, ¿de qué manera? Para responder a estas preguntas, se han seguido los mismos criterios definidos para la evaluación de la estrategia europea; es decir, se ha realizado un análisis documental para identificar las referencias implícitas y explícitas al género, siguiendo el planteamiento de la UNESCO⁷¹.

⁷¹ UNESCO. (2020). Artificial Intelligence and Gender Equality. <https://unesdoc.unesco.org/ark:/48223/pf0000374174>

En primer lugar, los documentos oficiales están alineados con el enfoque europeo en la definición de la IA, que puede relacionarse con la perspectiva de género de una forma implícita. Tanto la Estrategia Española I+D+I en IA como la Estrategia Nacional de IA señalan de una forma clara el compromiso de desarrollar sistemas de IA con una visión socio-técnica. Más concretamente, se menciona la importancia de la co-construcción en la relación entre humanos y tecnología: “El éxito de la IA dependerá de cómo las personas y las máquinas colaboren para ofrecer mejores servicios —transparentes, razonables y éticos— a los usuarios potenciales, en un mundo donde cada vez seremos más exigentes con los servicios prestados”⁷². No se hace referencia de manera explícita a la perspectiva de género en estas definiciones, pero se considera que esta perspectiva centrada en los seres humanos y en los valores éticos, implícitamente lleva aparejada una relación con la promoción de los derechos humanos, incluyendo la igualdad de género. Se trata de un marco general que sirve de paraguas para la promoción de los principios de igualdad y no discriminación que, entre otras cosas, considera los esfuerzos para alcanzar una igualdad efectiva entre hombres y mujeres.

En efecto, los documentos hacen referencia a la promoción de la igualdad, la inclusión y la no discriminación en otros apartados, más centrados en las estrategias y medidas concretas para diseñar e implementar la IA. Por ejemplo, se menciona el uso de la IA y otras tecnologías para promover la igualdad y proteger a las personas en riesgo de exclusión, para garantizar la inclusión y la diversidad, para reducir las desigualdades, y para contribuir con la protección de los derechos humanos y promover el bienestar social. De manera implícita, estos términos incluyen la perspectiva de género, una parte fundamental de la mayoría de los tratados internacionales de derechos humanos⁷³.

Pero, además, los documentos españoles se refieren de manera explícita a la promoción de la igualdad entre hombres y mujeres. Por ejemplo, la Estrategia Española I+D+I en IA señala que se deben diseñar algoritmos que eviten los sesgos de género: “Es una condición en el desarrollo de las tecnologías y aplicaciones de la IA ligado a esta Estrategia de I+D+I evitar el sesgo

⁷² Ministerio de Ciencia, Innovación y Universidades (2019). Estrategia Española de I+D+I en Inteligencia Artificial. [Última visita 20.03.2021]. https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf, p. 40.

⁷³ <https://www.ohchr.org/sp/issues/women/wrgs/pages/wrgsindex.aspx>

negativo y los prejuicios de los que adolece nuestra sociedad, como el género, raza, u otras formas de discriminación, y que deberán evitar los sistemas de soporte a la toma de decisiones”⁷⁴.

Los documentos también hacen alusión al uso activo de las nuevas tecnologías para abordar los desafíos de la sociedad, entre los que se incluye la reducción de la brecha de género. De hecho, la Estrategia Nacional de IA señala que, en vista de los retos que aún quedan en relación con esta materia, la promoción de la igualdad de género es un objetivo transversal de toda la estrategia. En este sentido, los sistemas de IA no solo deben ser “inclusivos y sostenibles”, como indica el documento, sino que también deben usarse como herramientas para reducir la desigualdad de género en distintos ámbitos.

A continuación, se destacan dos áreas relacionadas con la igualdad de género, en las que las estrategias españolas se han enfocado:

- Promoción de competencias digitales y formación de talento: Los documentos españoles reconocen la brecha de género que existe en las competencias digitales básicas y en el campo de la tecnología y las carreras STEAM (Science, Technology, Engineering, Arts and Mathematics). El Plan España 2025, por ejemplo, señala como una de sus metas que el 80% de las personas cuenten con competencias digitales básicas para el año 2025, y que de ese porcentaje 50% sean mujeres. En cuanto a las habilidades más avanzadas, se plantea formar a 250.000 personas para el diseño, desarrollo o explotación de sistemas digitales (entre los que se incluye la IA), con una “especial atención a la composición por género de estos especialistas”⁷⁵. La Estrategia Nacional de IA destaca la importancia de la formación digital a lo largo de toda la vida y en todos los ámbitos, “con una perspectiva ética, humanista y de género”⁷⁶. En

⁷⁴ Ministerio de Ciencia, Innovación y Universidades (2019). Estrategia Española de I+D+I en Inteligencia Artificial. [Última visita 20.03.2021]. https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf, p. 40.

⁷⁵ Ministerio de Asuntos Económicos y Transformación Digital (2020a). Plan España Digital 2025. [Última visita 20.03.2021]. https://portal.mineco.gob.es/ca-es/ministerio/estrategias/Pagines/00_Espana_Digital_2025.aspx, p. 30.

⁷⁶ Ministerio de Asuntos Económicos y Transformación Digital (2020b). Estrategia Nacional de Inteligencia Artificial. [Última visita 20.03.2021]. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>, p. 31.

concreto, la medida número 8 propone el desarrollo del Plan Nacional de Competencias Digitales, que ya está publicado⁷⁷, y que presenta la lucha contra la brecha digital de género como una de sus líneas de actuación. En cuanto a la formación en habilidades más avanzadas, el documento señala que se desarrollará un plan de acción para reducir la discriminación de género y promover la igualdad en el campo científico, tanto en la educación como en el acceso a puestos de trabajo relacionados con la IA.

- Apoyo al liderazgo de las mujeres en el ámbito de la IA: En relación con el punto anterior, una mayor capacitación de las mujeres puede influir también en la reducción de la brecha de género en el ámbito del emprendimiento digital. Como señala el Plan España Digital 2025, solo el 15,6% de las startups del país fueron fundadas por mujeres en 2018. También se consideran estrategias para atender esta situación en concreto. Por ejemplo, la Estrategia Nacional de IA incluye medidas como el Fondo NextTech, con capital de riesgo público y privado para desarrollar el emprendimiento digital y el crecimiento de empresas que trabajen con IA. Aquí se entiende que el emprendimiento digital es clave para el desarrollo de la IA y la transformación de la sociedad, con un foco claro en la promoción del talento y la igualdad entre hombres y mujeres.

El programa SpAIIn Talent Hub, propuesto también en la Estrategia Nacional de IA, se centrará, en coordinación con ICEX Invest in Spain, en la atracción de talentos académicos y profesionales y de inversión extranjera, e incluye la incorporación de mujeres emprendedoras. En el ámbito de la gobernanza para la ejecución de la estrategia, también se presta atención a la presencia de las mujeres en puestos de liderazgo. El documento destaca la creación, en julio de 2020, del Consejo Asesor para la Inteligencia Artificial⁷⁸, conformado por especialistas de las áreas científicas, económicas y sociales, “procurando la presencia equilibrada de mujeres y hombres”⁷⁹.

⁷⁷ https://portal.mineco.gob.es/RecursosNoticia/mineco/prensa/noticias/2021/210127_np_digital.pdf

⁷⁸ https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-8302

⁷⁹ Ministerio de Asuntos Económicos y Transformación Digital (2020b). Estrategia Nacional de Inteligencia Artificial. [Última visita 20.03.2021]. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>, p. 76.

La incorporación de la perspectiva de género en los documentos españoles sobre IA, tanto de forma implícita como de forma explícita, puede considerarse un paso importante para promover la igualdad efectiva entre hombres y mujeres. Como se ha señalado, se identifican dos rutas para alcanzar este fin: el uso de la IA como una herramienta para desarrollar proyectos que contribuyan con la reducción de las brechas de género, y la promoción de la igualdad de género en los datos y en el diseño de la propia tecnología, así como en las instituciones.

El discurso político y los documentos de las estrategias son fundamentales para sentar las bases sobre las que pueda avanzarse en políticas y proyectos más concretos. En función de este marco general, se deben seguir avanzando en propuestas que permitan abordar la igualdad de género con profundidad y considerando todas las aristas (por ejemplo, tomando en cuenta, entre otros aspectos, la interseccionalidad de género). Del mismo modo, será necesario evaluar en el futuro si estos discursos se llevan efectivamente a la práctica y de qué manera.

6. Propuestas para promover la igualdad de género en el uso de la IA en España

Como hemos visto hasta ahora, España cuenta tanto con una Estrategia Nacional de IA como con un plan de digitalización de la administración pública (Plan España Digital 2025). Aunque las propuestas incluidas en ambos documentos son muy variadas, la mayor parte de ellas, en lo que se refiere usar la IA para reducir la brecha de género, están orientadas hacia la creación de equipos paritarios y representativos, la formación de las mujeres en carreras STEM y la presencia de mujeres en posiciones de liderazgo que tengan que ver con la IA. Desde luego, estas propuestas son necesarias y están orientadas hacia la reducción de los sesgos de género. Sin embargo, a falta de fechas y elaboraciones más detalladas y concretas, estas propuestas aún parecen ser demasiado generales y no suelen ir más allá de los compromisos y propuestas recomendados por la UE. Además, no existen, o al menos no están adecuadamente documentados, los proyectos concretos que hacen uso de la IA para resolver problemas relacionados con la desigualdad de género. En el caso de España, creemos que además de centrarse en propuestas de carácter social como las anteriormente mencionadas, es necesario desarrollar propuestas de IA de carácter técnico, que sirvan para reducir la brecha de género existente en nuestro país. Consideramos que España podría desarrollar proyectos que usen tecnologías de IA para luchar contra la discriminación de las mujeres atendiendo a dos frentes.

Primero, usar la IA para rastrear, mapear y cuantificar las desigualdades existentes. Uno de los primeros problemas que nos encontramos cuando se intenta acabar con la desigualdad entre hombres y mujeres suele ser, por un lado, el negacionismo de la desigualdad por parte de ciertos sectores de la sociedad. En este sentido, la IA, mediante sus capacidades de monitorización, reconocimiento de patrones, voz, texto, etc., podría usarse para, al igual que en el caso del crimen, la seguridad, o la educación, detectar desigualdades (anomalías presentes en los datos) en distintos ámbitos de la sociedad. Por ejemplo, al igual que en el caso de Suecia, España podría proponer y fomentar la utilización de IA para cuantificar el tiempo que hablan hombres y mujeres en reuniones, el número de interrupciones, el tono y la forma de las intervenciones, etc. También se podrían usar sistemas de IA para cuantificar y medir el tipo de búsquedas de naturaleza médica que hacen las mujeres, los lugares más frecuentes desde donde suelen llamar a emergencias, los motivos de las llamadas, etc. Todo ello permitiría

aportar datos cuantificables a problemas reales que sufren las mujeres y que aún hoy en día son negados por muchos, por referirse a ellos como cuestiones subjetivas o de naturaleza marginal, así como descubrir otras potenciales discriminaciones que pueden sufrir las mujeres y de las que aún no somos conscientes. Segundo, se debería usar la IA para solucionar problemas y discriminaciones que sufren las mujeres por el simple hecho de ser mujeres. Como en el caso de Suecia, proponemos que la IA podría utilizarse, entre otras cosas, para proteger los derechos de las mujeres en algunas de las áreas prioritarias como, por ejemplo, la desigualdad salarial entre hombres y mujeres, la violencia de género, y la salud femenina.

Con el fin de orientar la incorporación de la IA en el sector público español hacia la reducción de la desigualdad entre hombres y mujeres, a continuación, presentamos una serie de propuestas concretas que consideramos podrían ser de utilidad. Las propuestas presentadas se han dividido en dos: transversales y concretas. Las propuestas transversales son de carácter técnico y social, orientadas a rastrear, mapear y reducir la desigualdad de género presente en todos los ámbitos de la sociedad. Las propuestas concretas están enfocadas a la resolución o mitigación de problemas específicos de gran relevancia y presentes en la sociedad española.

6.1. Propuestas transversales

En esta sección presentamos dos propuestas transversales que, con base en las políticas e iniciativas desarrolladas y probadas en otros países de nuestro entorno, consideramos que podrían contribuir a la reducción de la brecha de género: la redacción de guías con perspectiva de género para la implementación de la IA en las administraciones públicas españolas; la elaboración de informes anuales o bianuales sobre IA y género. Una de las dos propuestas es de carácter social, no necesitan de grandes presupuestos para su elaboración y podrían ponerse en marcha ya mismo.

Guías para la implementación de la IA en las AAPP con una perspectiva de género

Cuando hablamos de Inteligencia Artificial en el sector público estamos hablando, principalmente, de Inteligencia Artificial en las Administraciones Públicas. Las AAPP son las organizaciones públicas que configuran y dan forma a la organización política y territorial del Estado español. Su buen funcionamiento interno, así como la confianza ciudadana en estas

instituciones, es fundamental para disfrutar de democracias plenas y sanas. Sin embargo, la confianza en el gobierno y, por extensión, en buena parte de las AAPP ha ido descendiendo durante los últimos años. Como añadido a la pérdida de la confianza en el sistema político, las AAPP son en su mayor parte percibidas por la ciudadanía española como grandes estructuras burocráticas, poco abiertas a la innovación, y con grandes dificultades para modernizarse. Esto hace que la percepción de la eficiencia de la AAPP por parte de la ciudadanía no sea la más deseable. Estos y otros problemas compartidos por buena parte de nuestros vecinos europeos del sur, también de tradición napoleónica en lo que se refiere al sistema burocrático-administrativo, ha hecho que el Gobierno de España apueste por la transformación digital de la Administración Pública, con el protagonismo de la IA como uno de sus principales ejes de actuación en los próximos años.

La implementación de la IA en las AAPP es una oportunidad sin precedentes para que España transforme sus organizaciones públicas, de modo que no solo las haga más eficaces y eficientes, sino también más justas e igualitarias. Tradicionalmente tecnologías ampliamente implementadas en el sector público como es el caso de los sistemas expertos, y las decisiones tomadas por ellas, han sido percibidas como más neutrales, objetivas y racionales que aquellas tomadas por personas en tareas de dificultad media y/o baja. Esta visión de la tecnología como más objetiva y neutral puede ayudar a solucionar algunos de los problemas más importantes en los que se encuentra sumido el sector público, entre ellos el más destacado es la cada vez más frecuente idea de que las AAPP están altamente politizadas y, por ello, son poco objetivas y fiables. En este sentido, la adopción de la IA en las AAPP podría ayudar a restaurar los niveles de confianza en el sector público a través de la neutralidad y la objetividad que aportan las máquinas. Sin embargo, la incorporación de la IA al sector público también conlleva importantes riesgos (ver capítulo 1). La mayor parte de ellos tienen que ver o bien con problemas técnicos de la IA, o bien con problemas éticos derivados del uso de la IA: privacidad, transparencia, sesgos, rendición de cuentas, etc.

Para resolver el segundo tipo de problemas es fundamental diseñar un plan de implementación de la IA para las AAPP que sea pragmático, realista, adaptado a la realidad española, ético y, por supuesto, que incluya la perspectiva de género. Aunque la decisión de adoptar la IA en las AAPP es en su mayor parte una decisión política, no es la clase política la encargada de convertir

estas propuestas en una realidad. Los principales responsables de la implementación de la IA en las administraciones públicas españolas, a todos los niveles (nacional, regional y local), son los conocidos como Chief Information Officers (CIOs) de las propias administraciones. En este sentido, el principal reto que tiene asumir a día de hoy España es la de formar a estos profesionales públicos, en su mayor parte hombres con formación técnica (informática e ingeniería), de tal modo que estén preparados y capacitados para asumir la tarea de implementar la IA en las administraciones en su totalidad, es decir, tanto desde un punto de vista técnico como social y ético.

Al mismo tiempo, la introducción de la tecnología en las AAPP y otro tipo de organizaciones públicas no se ha llevado a cabo siguiendo una estrategia de implementación clara y eficaz. En este sentido, es fundamental que esta vez la implementación de la IA se haga con base en estudios científicos sobre teoría organizativa de las AAPP y su relación con la tecnología, en los que se enfatiza no solo los aspectos formales de las organizaciones (los que más se han tenido en cuenta hasta ahora), sino también en los aspectos informales (normas, prejuicios, ideas preconcebidas, cultura, etc.). Conocer los aspectos y normas informales de las AAPP es imprescindible para implementar la IA de forma que esta no solo sea recibida adecuadamente por parte de funcionarios y ciudadanía, sino también para que se pueda hacer un uso completo de todas sus potencialidades. Por ello, y siguiendo el ejemplo de otros países como el Reino Unido⁸⁰, proponemos que la recientemente creada Oficina del Dato, junto a la Secretaría de Estado de Digitalización e IA, realice una guía para que tanto los CIOs como los funcionarios TICs de todas las AAPP sepan cómo implementar la IA de forma adecuada.

Respecto al desarrollo y creación de la guía, sugerimos que sea desarrollada por personas expertas en el sector público e IA, procedentes de distintos ámbitos: académicos, investigadores, funcionarios públicos (TIC y no TIC), personas del sector privado, etc. Estos/as expertos/as no solo deben ser personas del área de la ingeniería y la informática, sino, y principalmente, politólogos, filósofos, sociólogos, etc., es decir, especialistas formados/as en disciplinas sociales y humanas, y con conocimiento, entre otras cosas, en teoría de género.

⁸⁰<https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector#examples-of-artificial-intelligence-use>

Además, consideramos oportuno que la propuesta resultante sea sometida a escrutinio público durante un periodo de recepción de comentarios, críticas y evaluaciones para su mejora.

Respecto al contenido de la propuesta, creemos que debe constar de tres partes. La primera parte del documento debería incluir una breve introducción al concepto de IA, sus potenciales beneficios, pero, sobre todo, sus riesgos y los problemas éticos que se derivan de una mala implementación y uso. Esta primera parte es fundamental para concienciar a funcionarios TIC y CIOs del carácter humano de la tecnología y su realidad como objeto socialmente construido y determinado. De esta forma, los encargados de implementar la IA en las AAPP podrán ser conscientes de que un mal desarrollo de su trabajo puede tener consecuencias muy graves, no solo a nivel técnico, sino también ético y social. En este sentido, es fundamental remarcar los efectos que una mala implementación puede tener sobre las mujeres como elemento amplificador de su situación desigual en la sociedad.

La segunda parte del documento debería estar compuesta por una estrategia completa en la que se expliquen los pasos de los que se compone el proceso de implementación de la IA, así como planes de acción y procesos de evaluación disponibles para asegurar que la adopción se ha realizado de forma correcta. Esta parte debería constar a su vez de distintas secciones.

- Evaluación: A la hora de implementar una tecnología de IA en una AAPP, el primer paso debería ser evaluar si este tipo de tecnología es la herramienta adecuada para solucionar el problema concreto. En el caso de que la respuesta sea afirmativa, se debe decidir el tipo de IA específico. También hay que evaluar si se dispone del personal suficiente y adecuado para esta labor de implementación (siguiendo criterios de paridad). A continuación, estudiar si se dispone de la información y los datos suficientes (tanto en cantidad como en calidad-representatividad) para que el sistema de IA funcione de forma adecuada. Hay que decidir si la IA la va a proporcionar una empresa cuyos servicios se contratan o si su elaboración va a depender del gobierno. Finalmente, hay que realizar una distribución de responsabilidades para que, en el caso de que algo salga mal, los/as usuarios/as de la IA reciban las explicaciones que se merecen y se puedan administrar las correspondientes sanciones.

- Planificación: Al proceso de evaluación le debe seguir un proceso de planificación del proyecto que debe ser realizado por expertos/as en IA. La planificación incluye la elección del tipo de IA, la selección del equipo (paridad), preparar los datos (representatividad), hacer modelos de prueba (sensibles a las discriminaciones por género), etc.
- Gestión: Finalmente, es necesario enmarcar los resultados del proyecto dentro de la ley y siguiendo unas directrices éticas. Para ello, es fundamental la constante supervisión humana de la herramienta de IA y las decisiones tomadas por estas tecnologías, siempre teniendo en cuenta, entre otras cosas, la perspectiva de género.

Finalmente, la guía debería incluir un apartado en el que se recogiesen otros casos y ejemplos de usos de IA en el sector público, tanto español como de otros países, de modo que pudiera servir de ejemplo y orientar a las personas encargadas de implementar la IA. Los casos escogidos deberían ser siempre ejemplos de un uso ético de la IA en los que se haya tenido en cuenta la perspectiva de género. También sería deseable que alguno de los ejemplos tratase sobre el uso de la IA para resolver un problema que afecta específicamente a las mujeres.

Informe sobre IA y género

Otra de las propuestas transversales orientadas a reducir la brecha de género es la elaboración de un informe anual o bianual sobre IA y género. Aunque los informes en sí mismos no ayudan a reducir de manera directa la brecha de género, sí pueden tener efectos positivos de manera indirecta. El objetivo de este informe sería aportar toda la información disponible cada año, o cada dos años, sobre el estado y los avances de la IA en relación con la igualdad de género. Esta tarea serviría como ejercicio de mapeo y rastreo sobre el desarrollo y la evolución de la Estrategia Nacional de IA enfocada a la reducción de la brecha de género. Aunque para rastrear el estado de la IA y las capacidades tecnológicas inteligentes en España actualmente existe la herramienta Mapa IA⁸¹, esta no está exclusivamente enfocada a las cuestiones de género y se encuentra infrautilizada, con un nivel de actualización muy bajo y con un escaso aporte de información relevante⁸². Respecto al desarrollo y creación del informe, creemos que debería

⁸¹ <https://mapa.estrategiaia.es/mapa>

⁸² En este sentido también animaríamos a la mejora de la herramienta Mapa IA para que sea más dinámica, fácil de usar, se actualice con más frecuencia y aporte más información relevante.

ser similar al propuesto para el caso de la guía para la implementación de IA en las AAPP. En este caso los/as expertos/as, tanto técnicos/as como no técnicos/as, deberían estar formados en su mayoría en cuestiones de género.

En lo que se refiere al contenido del informe, debería contar con una primera parte en la que se aporten estadísticas en las que se mida: el porcentaje de iniciativas de IA enfocadas a reducir la desigualdad de género; el número de mujeres que estudian carreras STEAM; el porcentaje de mujeres trabajando en cuestiones relacionadas con IA; y el número de mujeres liderando empresas tecnológicas o proyectos de investigación sobre IA. En este último caso, se debería atender tanto a equipos de investigación de las áreas de ingeniería y ciencias naturales, como ciencias sociales y humanidades. En una segunda parte, el informe podría dividirse siguiendo un criterio territorial, es decir, creando una sección para cada Comunidad Autónoma. A continuación, para cada comunidad deberían incluirse los siguientes temas. Primero, al igual que ya han hecho algunas comunidades autónomas como la Comunidad Valenciana, el resto de territorios españoles irán progresivamente desarrollando y publicando sus planes y/o estrategias territoriales para el desarrollo y la implementación de la IA. Por tanto, para cada territorio debería incluirse un resumen breve de la estrategia de IA de cada comunidad y en caso de que no exista, publicar información sobre fechas y proyectos previstos. Además, debería hacerse una comparativa para ver en qué medida las estrategias autonómicas siguen los criterios marcados por la estrategia nacional y la UE. A continuación, deberían incluirse los casos de IA de esa comunidad que están enfocados a resolver problemas específicos de las mujeres. Para cada caso se debería incluir: el nombre del sistema de IA que ha utilizado; el tipo de IA del que se hace uso; el estado del proyecto (planificado, en fase de prueba, implementado o retirado); si se trata de una iniciativa privada o pública (en el caso de que sea pública, si procede del gobierno central, la comunidad autónoma, el gobierno local, etc.); el sector al que pertenece la iniciativa (salud, educación, medioambiente, economía, etc.) dado que la desigualdad entre hombres y mujeres se manifiesta en todas las áreas; el número de mujeres que han formado parte del desarrollo del proyecto (tanto técnico como de otro tipo); el tipo de datos que se han utilizado; la fuente de los datos; una descripción del proyecto y, en el caso de que fuese retirado, los motivos por los que sucedió (para que así otras personas no comentan los mismos errores). Como referencia para la elaboración del informe podría servir el documento emitido por el AI Watch de la UE sobre IA en el sector público, Artificial

Intelligence in public services y el documento de la UNESCO sobre IA e igualdad de género, Artificial Intelligence and Gender Equality, ambos publicados en 2020.

Finalmente, el informe podría contar con una pequeña sección en la que se planteasen los principales problemas y obstáculos identificados en la implementación de este tipo de propuestas, futuras líneas de trabajo e investigación, así como enlaces a fuentes relevantes donde aquellas personas interesadas pudieran ampliar conocimientos. A continuación, mostramos de forma esquemática los puntos y la estructura que podría seguir un posible informe sobre IA y género.

I Informe sobre IA e igualdad de género

<p>1 Estadísticas</p> <ul style="list-style-type: none">  Proyectos IA enfocados a la igualdad de género  Mujeres trabajando en IA  Mujeres liderando iniciativas de IA  Mujeres estudiando carreras STEAM 	<p>2 Comunidades Autónomas</p> <p><u>Comunidad de Madrid</u></p> <p>Estrategia de IA: NO Fecha prevista: XX/XX/XXXX Otras iniciativas: Madrid. Innovación y Ciudad Inteligente</p> <p><u>Caso: Reconocimiento de voz violencia de género</u></p> <p>Nombre: VioMad Tipo de IA: Análisis de Audio Estado: Retirado Financiación: público (Comunidad de Madrid) - privada (nombre empresa) Sector: Seguridad, salud Datos: semiestructurados Fuente: CAM</p>  <p>Mujeres en puestos técnicos: XX% Mujeres en puestos de liderazgo: XX%</p>
<p>3 Propuestas de mejora</p> <ul style="list-style-type: none"> - La mayor parte de los casos presentan problemas de dudosa compatibilidad con la ley de protección de datos - Mejorar representatividad de los datos - Invertir en investigación sobre teoría de género  <p>FUENTE: SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E IA </p>	
<p>Descripción del proyecto: Sistema inteligente de reconocimiento de voz para identificar la gravedad de las llamadas al emergencias y el 016 y así detectar la necesidad actuación inmediata.</p> <p>Motivo de la retirada: Problemas de privacidad para las denunciadas que no solo dificultaban el uso del servicio, sino que iban en contra de la ley de protección de datos</p>	

Fuente: Elaboración propia

6.2. Propuestas concretas

Como hemos señalado, no solo se debe incluir la perspectiva de género en el diseño y la implementación de la IA de una forma transversal, sino que también estas tecnologías disruptivas pueden contribuir de manera activa con la promoción de la igualdad de género. Se han seleccionado tres áreas de acción, que podrían considerarse prioritarias para promover los derechos de las mujeres en España, de acuerdo con los principales retos identificados por el Plan Estratégico de Igualdad de Oportunidades de España 2014-2016⁸³.

IA para reducir la violencia de género

La violencia contra las mujeres es un gran problema en muchos países del mundo y España no es una excepción. De acuerdo con la Macroencuesta de Violencia contra la Mujer de 2019, 1 de cada 2 mujeres de 16 o más años, residentes en España, ha sufrido algún tipo de violencia por el hecho de ser mujer, desde miradas lascivas hasta violencia física y violación. A lo largo de su vida, el 13,7% de las mujeres ha sufrido violencia sexual, contando los casos en los que los perpetradores son parejas o personas con las que no se ha mantenido relación de pareja. En concreto, el 6,5% lo ha sufrido fuera del ámbito de la pareja. Considerando estos últimos casos, solo el 8% ha hecho una denuncia, y el porcentaje sube al 11% si se incluyen las denuncias realizadas a través de otras personas o de instituciones⁸⁴.

En España existe una iniciativa que utiliza la IA para ayudar a prevenir la violencia de género. Se trata del Sistema VioGén, que forma parte de la Secretaría de Estado de Seguridad del Ministerio del Interior y que se creó en 2007⁸⁵. El sistema utiliza la información de las denuncias y otros indicadores para valorar el riesgo de que una mujer vuelva a sufrir una agresión, y ofrece información para tomar decisiones policiales que permitan garantizar la protección de las

⁸³ Disponible en este enlace: <https://www.inmujer.gob.es/areasTematicas/AreaPlanificacionEvaluacion/docs/PlanesEstrategicos/PEIO2014-2016.pdf> Hay que señalar que la actualización de este plan estratégico está pendiente desde 2016 y que, de acuerdo con el Instituto de las Mujeres, está en proceso el nuevo documento, correspondiente a 2021-2025: <https://www.inmujer.gob.es/actualidad/noticias/2021/Marzo/comisionigualdad.htm>

⁸⁴ La Macroencuesta de Violencia contra la Mujer se realiza cada 4 años desde 1999. Se puede consultar la edición más reciente en este enlace: https://violenciagenero.igualdad.gob.es/violenciaEnCifras/macroencuesta2015/pdf/Resumen_Ejecutivo_Macroencuesta2019.pdf

⁸⁵ <http://www.interior.gob.es/web/servicios-al-ciudadano/violencia-contra-la-mujer/sistema-viogen>

víctimas. Recientemente se anunció la incorporación de una plataforma desarrollada por la empresa SAS Iberia para hacer estos análisis de una forma mucho más rápida⁸⁶.

En este sentido, planteamos algunas propuestas para avanzar hacia la prevención de la violencia contra las mujeres con el uso de la IA:

- Mantener la colaboración interinstitucional en todos los ámbitos de la administración pública para garantizar el correcto funcionamiento del sistema VioGén en el futuro. Es importante también continuar y profundizar la colaboración con el sector privado para que el sistema pueda beneficiarse de los últimos avances en el ámbito de la IA y el big data.
- Diseñar un chatbot que facilite la denuncia de casos de violencia contra las mujeres. Un ejemplo es el chatbot Sis Bot, desarrollado en Tailandia para ofrecer información a las mujeres que han sufrido violencia y guiarlas en el proceso de denuncia y en el acceso a los servicios de protección. Se trata de una aplicación que se descarga en el teléfono y que permite a las mujeres mantener una conversación para recibir la ayuda que necesitan en tiempo real⁸⁷. También se puede desarrollar un sistema por SMS, como el australiano Hello Cass⁸⁸. La iniciativa española puede inspirarse también en el chatbot elaborado por la Universidad de Maastricht, en Holanda, que permite a las mujeres contar su historia con libertad y ofrece al final consejos del lugar al que se debe acudir en función de cada caso (como la comisaría, el hospital, un psicólogo o un refugio)⁸⁹.

En España, este sistema podría integrarse a la app LIBRES, presentada en el año 2013 por el entonces Ministerio de Sanidad, Servicios Sociales e Igualdad, pero que actualmente no está disponible. Esta app quedaba camuflada en el menú principal del teléfono, para garantizar la seguridad, y ofrecía información para ayudar a las mujeres

⁸⁶ <https://noticias.juridicas.com/actualidad/noticias/15848-interior-recurre-a-la-tecnologia-de-inteligencia-artificial-para-mejorar-la-valoracion-policial-de-riesgo-en-casos-de-violencia-de-genero/>

⁸⁷ <https://www.unwomen.org/en/news/stories/2019/5/feature-using-ai-in-accessing-justice-for-survivors-of-violence>

⁸⁸ <https://hellocass.com.au/>

⁸⁹ https://elpais.com/tecnologia/2019/09/19/actualidad/1568904947_797374.html

a saber si estaban siendo víctimas de maltrato y conocer los servicios y recursos de ayuda⁹⁰. Se desconocen las razones por las que la app dejó de estar en funcionamiento, pero podría ser una buena plataforma para alojar el chatbot que permita a las mujeres contar su historia de forma confidencial y recibir la información que necesitan de una forma más rápida y efectiva. El resto de los recursos de la app pueden seguir siendo útiles para obtener información más general y concienciar a la población sobre este problema.

- Usar la IA para los casos de denuncias de violencia económica, que también es una forma de violencia de género. Si bien es cierto que se trata de un número minoritario de casos en España (3% del total de 601.416 victimizaciones de violencia contra la mujer entre 2015 y 2019⁹¹), sigue siendo una modalidad que afecta a las mujeres, principalmente por el impago de prestaciones económicas y el abandono de la familia. También se incluyen casos de hurto, daños en vehículos y robo con fuerza⁹². Se puede incluir información sobre este tipo de violencia en la app y el chatbot que se comentó en el apartado anterior, y también se pueden realizar otros desarrollos en el ámbito de la justicia. Un ejemplo sobre el uso de la IA en estos casos es el del Ministerio Público Fiscal de Buenos Aires, Argentina, que con el apoyo del laboratorio IALAB implementó el sistema Prometea para optimizar los procesos judiciales en los casos de incumplimiento del deber de asistencia familiar. Se utilizó la IA para generar automáticamente los documentos necesarios para el proceso y reducir la solicitud de informes a las víctimas, lo que hacía que el procedimiento fuera más lento⁹³.

En España ya se está avanzando en el análisis del uso de la IA en el Poder Judicial. Uno de los ámbitos de aplicación, precisamente, es en el cálculo de las pensiones compensatorias alimenticias⁹⁴. Una sugerencia en este sentido es inspirarse en otros

⁹⁰https://violenciagero.igualdad.gob.es/laDelegacionInforma/pdfs/DGVG_Informa_ACCESIBILIDAD_APP_LIBRE_S_2016.pdf

⁹¹ Informe de Violencia contra la Mujer de la Secretaría de Estado de Seguridad, 2015-2019: http://www.interior.gob.es/documents/642317/13041442/Informe_sobre_violencia_contra_la_mujer_+2015-2019_126210076.pdf/8b4a2cc8-530e-4927-b79f-d1f2ad004d36

⁹² Ibid.

⁹³ <https://ialab.com.ar/iaconperspectivadegenero/>

⁹⁴ <https://elpais.com/tecnologia/2021-02-20/la-inteligencia-artificial-se-abre-paso-en-la-justicia-espanola.html>

casos, como el argentino, para mantener el foco en las víctimas y contribuir con la reducción de la burocracia, especialmente en casos de violencia de género. Es decir, que más allá de esos cálculos concretos, la IA sirva para optimizar los propios procesos de la justicia, considerando la perspectiva de género. Por supuesto, es indispensable ponderar adecuadamente los usos de estos sistemas para que las decisiones estén apegadas al derecho. La IA se puede constituir en una herramienta que ayude a la labor de los seres humanos en este tipo de decisiones.

IA para reducir la brecha salarial entre hombres y mujeres

En España se ha avanzado en los últimos años en la reducción de la brecha salarial de género, pero todavía queda trabajo pendiente. Un estudio de la Confederación Española de Organizaciones Empresariales y PwC, publicado en 2019, señala que las mujeres cobran 14% menos que los hombres por hora (una brecha más baja que la media europea de 17,6%). Los factores que pueden influir en la existencia de estas diferencias son variados, desde la antigüedad en la empresa, el sector económico, el cargo que se ocupa, el tipo de contrato, la dedicación a los cuidados familiares, etc. Pero cuando se toma en cuenta la brecha salarial ajustada, una medida en la que se aíslan esas variables para determinar solo las diferencias por razones de género, se encuentra que las mujeres cobran 12,2% menos que los hombres. Este dato se puede explicar por la masculinización de algunos sectores que tienen mejores salarios o por compensaciones por horas extras y horarios nocturnos, por ejemplo. También se ha encontrado que la brecha salarial es más alta (16,6%) en los niveles con mayores salarios⁹⁵.

Los sistemas de IA pueden ayudar a obtener información detallada sobre la brecha salarial de género dentro de las empresas. La compañía británica Gapsquare, por ejemplo, desarrolla softwares que permiten vincular los sistemas de nóminas y de recursos humanos para saber si dentro de una organización existen diferencias salariales entre hombres y mujeres, para entender las razones detrás de estas diferencias (con datos de investigaciones y estudios), y para hacer predicciones y ofrecer información que ayude a tomar decisiones que mejoren la

⁹⁵ El informe Análisis de la brecha salarial de género en España. Identificando las causas para encontrar las soluciones, de la CEOE y PwC, se puede descargar en este enlace: https://contenidos.ceoe.es/CEOE/var/pool/pdf/publications_docs-file-597-analisis-de-la-brecha-salarial-de-genero-en-espanav2.pdf

situación⁹⁶. También se puede usar esta información y otros datos para crear modelos que permitan personalizar los paquetes de compensación, en función de las necesidades y expectativas de los/as trabajadores/as y para aumentar la transparencia dentro de las organizaciones⁹⁷.

En España, estos sistemas de IA pueden ser útiles para facilitar el cumplimiento de las normas en materia de transparencia retributiva, que obliga a las empresas a identificar y publicar las diferencias salariales entre hombres y mujeres, y determinar si se deben a discriminaciones⁹⁸. A diferencia de los sistemas informáticos que se utilizan normalmente para las tareas contables dentro de las empresas, los sistemas de IA podrían ayudar a realizar este trabajo de una forma más rápida y ofrecer datos que permitan la toma de decisiones informada, a través de análisis y predicciones.

En este sentido, ofrecemos las siguientes propuestas:

- Impulsar y apoyar a las empresas en el diseño de sistemas de IA para garantizar la transparencia retributiva y la igualdad de género dentro del sector privado, a través de subvenciones y otros mecanismos de financiación.
- Diseñar e implementar estos sistemas de IA para la transparencia retributiva dentro del propio sector público, donde existe una brecha salarial de 10,3⁹⁹, y una brecha salarial ajustada de 6,7%¹⁰⁰. Si bien se trata de un porcentaje menor al de la media general, valdría la pena usar la IA para entender cómo y por qué ocurre esta situación en las administraciones públicas.

⁹⁶ <https://thenextweb.com/work2030/2019/04/24/how-ai-can-help-close-the-gender-pay-gap-and-eliminate-bias/>

⁹⁷ <https://www.hrtechnologist.com/articles/digital-transformation/how-ai-in-hr-will-close-the-gender-pay-gap/>

⁹⁸ Real Decreto 902/2020, de 13 de octubre, de igualdad retributiva entre mujeres y hombres: <https://www.boe.es/boe/dias/2020/10/14/pdfs/BOE-A-2020-12215.pdf>

⁹⁹ Datos de la Encuesta Anual de Estructura Salarial 2018: https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259925408327&p=1254735110672&pagename=ProductosYServicios%2FPYSLayout¶m1=PYSDetalle¶m3=1259924822888

¹⁰⁰ De acuerdo con el informe de CEOE y PwC, citado más arriba.

- Crear mecanismos para la colaboración entre el sector público y el privado, para hacer seguimiento constante del funcionamiento de estos sistemas de IA, compartir experiencias y desarrollar soluciones en conjunto.

IA para la salud de las mujeres

Las mujeres tienen necesidades específicas de salud a las que se debe prestar atención de manera especial. Más allá del embarazo y la salud sexual, hay datos que demuestran algunos problemas específicos que sufren las mujeres en España. Por ejemplo, las mujeres tienen una tasa más alta de discapacidad que los hombres (106,3 por cada 1.000 habitantes, comparado con 72,6 por cada 1.000 habitantes)¹⁰¹; 90% de las personas que sufren Trastornos de la Conducta Alimentaria son mujeres¹⁰²; los hombres perciben que tienen un buen estado de salud en mayor medida que las mujeres¹⁰³; y los problemas de salud crónicos que experimentan con mayor frecuencia las mujeres y los hombres son diferentes¹⁰⁴.

En este ámbito hay mucho potencial para desarrollar sistemas de IA que contribuyan con el diagnóstico y el tratamiento de enfermedades. Investigadores del MIT, por ejemplo, crearon un algoritmo de deep learning para analizar mamografías¹⁰⁵, y en la Universidad de Valencia y la Universidad Politécnica de Valencia se desarrolló un método basado en redes neuronales y algoritmos predictivos para ayudar en la detección del cáncer de mama, como resultado del proyecto mundial Digital Mammography DREAM Challenges¹⁰⁶. Más recientemente, se anunció la implementación de sistemas de IA para estudiar historias clínicas de pacientes con coronavirus en el Hospital Clínic de Barcelona, como parte del plan España Digital 2025¹⁰⁷. En

¹⁰¹ Datos de la Encuesta de Discapacidad, Autonomía personal y situaciones de Dependencia 2008, del INE: https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259926668516&p=1254735110672&pagename=ProductosYServicios%2FPYSLayout¶m1=PYSDetalle¶m3=1259924822888

¹⁰² Datos de la Fundación Fita y la Asociación Española para el Estudio de los Trastornos de la Conducta Alimentaria, publicados en 2019: <https://www.lavanguardia.com/vida/20190110/454064488085/400000-personas-la-mayoria-jovenes-sufren-un-trastorno-de-alimentacion.html>

¹⁰³ Datos de la Encuesta Nacional de Salud 2017: https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259926692949&p=1254735110672&pagename=ProductosYServicios%2FPYSLayout¶m1=PYSDetalle¶m3=1259924822888

¹⁰⁴ Datos de la Encuesta Nacional de Salud 2017: https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259926692949&p=1254735110672&pagename=ProductosYServicios%2FPYSLayout¶m1=PYSDetalle¶m3=1259924822888

¹⁰⁵ <https://swhr.org/the-promise-and-peril-of-ai-in-womens-health/>

¹⁰⁶ <https://www.uv.es/uvweb/universidad/es/ficha-persona/desarrollan-nuevo-sistema-ayuda-al-diagnostico-del-cancer-mama-basado-inteligencia-artificial-1285950309813/Novetat.html?id=1286030737530>

¹⁰⁷ https://www.elespanol.com/invertia/empresas/tecnologia/20200909/inteligencia-artificial-hospitales-luchar-covid-digitalizacion-sanitaria/519199248_0.html

total, España cuenta con 168 entidades, entre empresas, administraciones públicas, centros de enseñanza e instituciones privadas sin fines de lucro, que desarrollan servicios o tecnología de IA en el ámbito de la salud¹⁰⁸.

En esta línea, presentamos las siguientes recomendaciones:

- Incorporar la perspectiva de género en todas las iniciativas de IA en el ámbito de la salud, para garantizar que se consideran de forma diferenciada las necesidades de hombres y mujeres. Del mismo modo, se debe prestar especial atención a los sesgos de género que pueden contener los datos que usen estos sistemas, para garantizar un análisis adecuado y reducir las posibles discriminaciones de género.
- Considerando iniciativas desarrolladas en otros países y en el sector privado, como Grace Health (mencionada en apartados anteriores) y Emi Bot (desarrollado por la consultora española Acuam HealthCare)¹⁰⁹, el Ministerio de Sanidad de España podría diseñar un chatbot que interactúe con las usuarias para ofrecer información sobre asuntos de salud específicos de las mujeres, basándose en evidencias científicas y datos publicados por las organizaciones públicas. También se puede incluir información sobre los servicios a los que se puede acudir para buscar ayuda en caso de que sea necesario.

¹⁰⁸ De acuerdo con el Mapa de Capacidades de tecnologías de IA: <https://mapa.estrategiaia.es/mapa>

¹⁰⁹ <https://www.acuam.com/portfolio/emi-bot/>

7. Conclusiones

Por todo lo tratado en el presente documento, dos cosas deben quedar claras. Primero, que la Inteligencia Artificial puede traer cambios sin precedentes en el sector público. Debido a su gran capacidad de almacenamiento, manejo y procesamiento de datos, junto al uso de nuevas técnicas como el machine learning, el deep learning, etc., el uso de la IA en el sector público permitirá que estas tecnologías realicen tareas hasta ahora imposibles de atribuir a las máquinas y a una velocidad sin precedentes. Todos estos cambios transformarán el paradigma administrativo, no solo español, sino también de todos aquellos países que sustituyan sus estructuras burocráticas por estructuras algorítmicas (algoritmización), haciendo realidad el nuevo paradigma de la gobernanza algorítmica. Sin embargo, si queremos asegurar que este nuevo paradigma sea sinónimo de progreso y no de retroceso, primero hay que tener en cuenta y resolver diferentes cuestiones, tanto técnicas como sociales.

Como hemos visto, entre las ventajas de la adopción de la IA en el sector público se encuentra el aumento de la eficacia y la eficiencia de los servicios públicos, la transformación del ciclo de políticas públicas, la personalización de los servicios, etc. En contraposición, entre las desventajas más importantes de una mala adopción de la IA hemos destacado: la perpetuación de sesgos existentes (en algunos casos amplificación), la opacidad algorítmica, la falta de transparencia en los datos, problemas de privacidad, pérdida de empleo, etc. Todos estos problemas presentes en la IA son especialmente peligrosos en el caso de las mujeres. En este sentido, hemos señalado las dos principales fuentes de la discriminación de las mujeres cuando se usa la IA: los problemas y sesgos en los datos, y los problemas con el diseño de los algoritmos. En el caso de los datos, la falta de representatividad y los sesgos presentes en ellos son los dos principales problemas que se deben atender. En lo que se refiere al diseño de los algoritmos, es fundamental incluir la perspectiva de género para así sesgar lo sesgado, es decir, corregir los sesgos presentes en los datos.

Respecto al estado de desarrollo e implementación de la IA en Europa, hemos visto que la IA es una de las prioridades de la UE desde hace varios años. En este sentido, la UE ha creado varios comités, grupos de expertos y observatorios para abordar la adopción y uso de la IA en Europa desde una perspectiva ética, que ponga en el centro a los seres humanos y que sea digna de confianza. A día de hoy, la mayor parte de países de la Unión han seguido las recomendaciones de la Comisión Europea y ya han publicado sus estrategias nacionales de IA. De entre todas ellas, en este documento nos hemos enfocado en las estrategias nacionales de Suecia y España.

Hemos visto que Suecia se trata de un país muy adelantado en temas de IA en general, pero concretamente en proyectos de IA enfocados a reducir la desigualdad de género. Muestra de ello es la elaboración de informes y documentos propios enfocados en abordar este problema, el elevado número de proyectos destinados a esta labor, así como el alto número de mujeres presentes en el ámbito de la IA respecto a otros países de su entorno. En el caso de España, su Estrategia Nacional de IA también hace menciones a la igualdad de género y distintas afirmaciones mediante las cuales se compromete a utilizar la IA de forma ética, es decir, de modo que no ahonde en la brecha de género. Aunque los documentos españoles incluyen propuestas para evitar que la IA repercuta negativamente en las mujeres como, por ejemplo, la formación de mujeres en carreras STEAM, la revisión y transparencia de los datos, el liderazgo femenino en la IA, etc.; no hemos encontrado apenas propuestas proactivas dirigidas a reducir la desigualdad entre hombres y mujeres. Es decir, que aunque existen propuestas, no hemos encontrado entre los documentos oficiales propuestas concretas en las que se mencionen proyectos (en fase desarrollo, prueba o ya implementados) en los que se usa la IA para acabar con problemas específicos de las mujeres. Al contrario que Suecia, España se encuentra casi exclusivamente centrada en evitar los efectos negativos del uso de la IA en las mujeres, y prácticamente nada en el uso de la IA para, de manera activa, luchar contra la desigualdad.

Por ello, creemos que, aunque España transita dentro de la senda adecuada en la lucha a favor de la igualdad de género a través del uso de IA, todavía se encuentra lejos de otros países como Suecia. Aunque esta no es una tarea sencilla, España es una de las principales economías de la zona euro y cuenta con todos los recursos disponibles para impulsar proyectos e iniciativas similares a las de otros países de su entorno. La lucha contra la desigualdad de género, incluso

cuando en ella está involucrada una tecnología como la IA, es siempre política y, por tanto, en nuestras manos está trabajar en ella. En este informe hemos presentado cinco propuestas que creemos que podrían servir como primeros pasos hacia un uso más feminista de la IA, tanto a nivel transversal como concreto en áreas y problemas específicos de las mujeres: IA para reducir la violencia de género, IA para acabar con la desigualdad salarial, IA para mejorar la salud de las mujeres, desarrollo de una guía para una implementación feminista de la IA en las AAPP y la elaboración de un informe sobre IA e igualdad de género en España. De este modo, pretendemos que la IA y las propuestas relacionadas con ella dejen de enfocarse casi exclusivamente de manera negativa, en el sentido de propuestas dirigidas a solucionar problemas relacionados con la IA en su origen, y que la IA se vea, además, como una oportunidad, un puente, para alcanzar la igualdad de género.

Referencias

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain. <https://papers.nips.cc/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf>

Buchanan, B. G. (2006). A (Very) Brief History of Artificial Intelligence. *AI Magazine*, 26(4).

Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.

Buolamwini, J., y Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, 1-15.

Collett, C. y Dillon, S. (2019). AI and gender. Four Proposals for Future Research. [Última visita 20.03.2021]. <https://www.repository.cam.ac.uk/handle/1810/294360>

Comisión Europea. (2018). Artificial Intelligence for Europe. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>

Faggella, D. y D'Ambra, L. (2020). Women in Artificial Intelligence – A Visual Study of Leadership Across Industries. [Última visita 20.03.2021]. <https://emerj.com/ai-market-research/women-in-artificial-intelligence-visual-study-leaderships-across-industries/>

Feine, J., Gnewuch, U., Morana, S., y Maedche, A. (2020). Gender Bias in Chatbot Design. En Følstad, A., Araujo, T., Papadopoulos, S., Lai-Chong Law, E., Granmo, O., Luger, E., Bae Brandtzaeg, P. (Eds.) *Chatbot Research and Design* (11970:79-93). *Lecture Notes in Computer Science*. Cham: Springer International Publishing, https://doi.org/10.1007/978-3-030-39540-7_6.

Gobierno de Suecia. (2018). National Approach for Artificial Intelligence. [Última visita 16.03.2021] <https://www.government.se/4a7451/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf>

HLEG. (2019a). A definition of Artificial Intelligence: main capabilities and scientific disciplines.

<https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>

HLEG. (2019b). Ethics guidelines for trustworthy AI. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

HLEG. (2019c). Policy and Investment Recommendations for Trustworthy AI. <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>

HLEG. (2020a). The assessment list for Trustworthy Artificial Intelligence (ALTAI). <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

HLEG. (2020b). Sectoral Considerations on the Policy and Investment. <https://futurium.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-sectoral-considerations-policy-and-investment-recommendations-trustworthy-ai>

Innova. (2018). Artificial intelligence in Swedish business and society. Analysis of development and potential. https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf

Instituto Europeo para la Igualdad de Género. (2020). Gender Equality Index 2020. <https://eige.europa.eu/gender-equality-index/2020>

Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. In *Nature* (Vol. 568, Issue 7751, pp. 163–165). Nature Publishing Group. <https://doi.org/10.1038/d41586-019-01099-5>

Martínez-Cantos, J.L. (2017). Digital skills gaps: A pending subject for gender digital inclusion in the European Union. *European Journal of Communication*, 32(5), 419-438. <https://doi.org/10.1177/0267323117718464>

McDonnell, M., & Baxter, D. (2019). Chatbots and gender stereotyping. *Interacting with Computers*, 31(2), 116-121.

Miner, A. S. (2016). Smartphone-Based Conversational Agents and Responses to Questions about Mental Health, Interpersonal Violence and Physical Health. *JAMA Internal Medicine* 176, no 5 (2016) 619-25.

Ministerio de Asuntos Económicos y Transformación Digital (2020a). Plan España Digital 2025. [Última visita 20.03.2021]. https://portal.mineco.gob.es/ca-es/ministerio/estrategias/Pagines/00_Espana_Digital_2025.aspx

Ministerio de Asuntos Económicos y Transformación Digital (2020b). Estrategia Nacional de Inteligencia Artificial. [Última visita 20.03.2021]. <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>

Ministerio de Ciencia, Innovación y Universidades (2019). Estrategia Española de I+D+I en Inteligencia Artificial. [Última visita 20.03.2021]. https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf

Misuraca, G. and Van Noordt, C., AI Watch - Artificial Intelligence in public services, EUR 30255 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19540-5 (online), doi:10.2760/039619 (online), JRC120399.

Prates, M. O., Avelar, P. H., & Lamb, L. C. (2019). Assessing gender bias in machine translation: a case study with Google Translate. *Neural Computing and Applications*, 1-19.

Ridgeway, C. (2009). Framed Before We Know It: How Gender Shapes Social Relations. *Gender & Society*, 23(2), 145-160. <https://doi.org/10.1177/0891243208330313>

Sáinz, M., Arroyo, L., y Castaño, C. (2020). Mujeres y digitalización. De las brechas a los algoritmos. Instituto de la Mujer y para la Igualdad de Oportunidades. [Última visita 18.03.2021]. https://www.inmujer.gob.es/disenov/novedades/M_MUJERES_Y_DIGITALIZACION_DE_LAS_BRECHAS_A_LOS_ALGORITMOS_04.pdf

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, 361, 751–752. <https://doi.org/10.1126/science.aat5991>

UIT (2020). Measuring digital development. Facts and figures. [Última visita 17.03.2021]. <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf>

UNESCO. (2020). Artificial Intelligence and Gender Equality. <https://unesdoc.unesco.org/ark:/48223/pf0000374174>

West, M., Kraut, R., y Ei Chew, H. (2019). I'd blush if I could. Closing gender divides in digital skills through education. [Última visita 17.03.2021]. <https://en.unesco.org/ld-blush-if-i-could>

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (First Edit). PublicAffairs. <https://doi.org/10.1386/jdmp.10.2.229>