Inteligencia artificial y violencia machista
Retos y oportunidades

Intel·ligència artificial i violència masclista
Reptes i oportunitats

Mª Ángeles Calero Fernández (ed.)

Inteligencia artificial y violencia machista
Retos y oportunidades

Intel·ligència artificial i violència masclista
Reptes i oportunitas

Mª Ángeles Calero Fernández (ed.)

Edicions de la Universitat de Lleida
Lleida, 2025

Finançat per: Regidoria de Polítiques Feministes de l'Ajuntament de Lleida i Departament de Filologia i Comunicació de la Universitat de Lleida

Coordinadora del projecte: Regidoria de Polítiques Feministes. Ajuntament de Lleida

ISBN 978-84-9144-602-6
Maquetació: Edicions i Publicacions de la UdL
Imatges de portada: Rosa Maria Gil Iranzo, *Sexist Violence: Challenges and Opportunities*, creat amb Midjourney 7.0

Edicions de la Universitat de Lleida, 2025

Nota: Els diferents capítols que conformen aquest llibre han passat per un procediment de revisió per parells.

Índice

Pròleg

Carme Valls Llaràs
4a Tinenta d'Alcaldia i Regidora
de Bon Govern i Polítiques Feministes
Ajuntament de Lleida

Aquesta publicació neix de la iniciativa i el compromís de l'Ajuntament de Lleida de buscar aliances amb el teixit institucional, acadèmic, empresarial, cultural i social del municipi de Lleida per tal d'avançar en la construcció d'una ciutat lliure de violències masclistes. Tracta sobre les oportunitats que ens ofereix la Intel·ligència Artificial (IA) en l'àmbit del feminisme i dins el marc del *Pacte Local contra les violències masclistes*.

En aquest sentit, durant l'any 2024 l'Ajuntament va liderar diferents actuacions amb l'objectiu de buscar sinèrgies entre persones expertes de l'àmbit tecnològic i de l'àmbit social per tal d'aprofundir en les oportunitats que la IA ens ofereix en l'erradicació de les violències masclistes.

A través de la creació d'un grup d'expertes, es va dissenyar un projecte en diferents fases per tal de detectar i afrontar els biaixos de gènere que ja s'estan denunciant. Partint d'un seminari amb el mateix títol, es va definir de manera col·laborativa l'elaboració de tres productes: a) recerca amb IA centrada en els consums culturals de l'alumnat universitari i l'anàlisi d'estereotips de gènere i de violències masclistes; b) recull de reflexions sobre la IA i les violències masclistes des de mirades diverses del món acadèmic; c) elaboració de propostes tecnològiques amb IA per resoldre problemes relacionats amb les violències masclistes. Arran de les reflexions d'aquest grup d'expertes i amb la participació d'alumnat de les escoles municipals, també es va dur a terme una *hackató* per trobar solucions tecnològiques a problemes socials en relació amb les violències masclistes i, al

mateix temps, sensibilitzar el col·lectiu jove perquè sigui conscient de les oportunitats que ofereix la IA en la prevenció de la violència.

El llibre resultant d'aquest projecte presenta una anàlisi interdisciplinària i un estat de la qüestió, incloent investigacions i projectes que posen sobre la taula tant les possibilitats reals de la IA com els seus reptes actuals.

Relaciona aspectes que tenen a veure amb els biaixos de gènere, les violències masclistes i les accions que s'aborden transversalment des de l'àmbit local, així com les oportunitats que s'obren envers la prevenció de les violències, la protecció i seguretat de les dones, el potencial com a agent crític d'igualtat... Es posa en relleu la necessitat d'un abordatge transversal, legislatiu i de drets.

Es tracta d'oferir una anàlisi que vinculi drets de les dones, ciutadania i tecnologia i com tot això es pot abordar des l'agenda feminista, amb perspectiva de gènere, des de l'ètica i el coneixement compartit.

Presentación

Mª Ángeles Calero Fernández
Universidad de Lleida

Los cambios tecnológicos han tenido, a lo largo de la historia, un fuerte impacto en los seres humanos y en las sociedades que estos constituyen. La creación y el desarrollo de la Inteligencia Artificial (IA) no se escapa de esta circunstancia, por lo que es, más que recomendable, imprescindible plantearse no solo cuáles son los efectos que la IA está produciendo o producirá en nuestra vida y en nuestras relaciones interpersonales, sino también qué tipo de realidad es la que está transmitiendo. En un mundo patriarcal como este en el que vivimos, las posibilidades de que la IA genere y transmita sesgos de género o que, incluso, sea un instrumento más del patriarcado para perpetuarse, son elevadas. En estas páginas se revisa esta cuestión de tanto calado a través de siete breves ensayos y un glosario terminológico elaborados por especialistas en comunicación, en IA y en estudios de género, que intentan explicar, de manera divulgativa, los peligros, las oportunidades y los retos de esta nueva tecnología en el ámbito de la violencia de género.

Previo a este prólogo, habrán podido leer una presentación institucional de la concejala de Políticas Feministas del Ayuntamiento de Lleida, toda vez que este libro es resultado de una de las acciones que se llevaron a cabo en 2024 dentro de un proyecto financiado por el Departamento de Igualdad y Feminismos de la Generalitat de Cataluña y que se detalla en el texto que presenta la Unidad Técnica de Políticas de Igualdad siguiendo este prólogo, tras contextualizar dicho proyecto tanto socialmente como dentro de la política del Ayuntamiento de Lleida.

El primer ensayo que encontramos es el de Iolanda Tortajada y Cilia Willem, titulado *La IA y las mujeres: una relación no tan inteligente*. En él, las autoras plantean cómo la IA ha introducido un nivel de violencia hacia las mujeres mayor del que ha existido hasta

ahora, ya que permite, por ejemplo, crear y difundir, sin su consentimiento, imágenes falsas de mujeres reales desnudando sus cuerpos o colocándolas en posturas obscenas. Reflexionan sobre cómo una práctica como esta encaja en una sociedad patriarcal que observa el mundo con ojos masculinos y que pretende mantener a las mujeres sometidas, lo que significa que lo que genera y lo que provoca la violencia machista en el mundo real se traslada inalterable al mundo virtual y se amplifica en él. Asimismo, las autoras exponen el hecho de que, mientras los contenidos que circulen por internet y por las redes sociales sean androcéntricos, sexistas o claramente machistas, la IA, que se alimenta de la información que circula por ese mundo virtual, contribuirá a la opresión de las mujeres y las obligará a mantenerse al margen o a huir de él. Por otra parte, si el sector tecnológico sigue siendo un espacio masculinizado, resultará difícil revertir esta situación. Por ello, Tortajada y Willem divulgan diferentes acciones institucionales y sociales que intentan reconducir la IA. Además proponen apropiarse colectivamente de ella y utilizarla para proteger a las mujeres, promoviendo un cambio cultural y contribuyendo a señalar quién se beneficia de cada comportamiento y a costa de quién.

El segundo trabajo es de Leonor M. Cantera y lleva por título *La Inteligencia artificial y sus implicaciones éticas, sociales y políticas*. La autora advierte de la imperiosa necesidad de aplicar la perspectiva de género en el diseño de la IA para que sea posible una representación de las mujeres ajustada a la realidad, se evite reforzar las desigualdades y la discriminación, y se promueva la justicia y la equidad. Explica algunas aplicaciones de IA para luchar contra la violencia de género y para asistir a las víctimas, pero reconoce que no es suficiente y que es esencial revisar y auditar de manera recurrente los algoritmos que utiliza la IA y la calidad y diversidad de las bases de datos sobre las que esta trabaja. Asimismo, es imprescindible que las decisiones que se tomen para regular esta tecnología se adopten de manera democrática y se establezcan mecanismos para que sea accesible a todo el mundo. Cantera considera que las ciencias sociales y las humanidades representan un papel crucial en el reto que supone el desarrollo, la implementación y la vigilancia de la IA para asegurar que se utilice en beneficio de toda la humanidad.

El tercer ensayo, de Rosa María Gil Iranzo, cuyo título es *Inteligencia Artificial y sesgos*, comienza describiendo algunos ámbitos en los que la IA puede generar sesgos de mayor impacto, como son el reconocimiento facial, los simuladores conversacionales (*chatbots*), los algoritmos de contratación de personal o las aplicaciones multimedia con sistema de recomendación. Continúa presentando diversos desafíos que deben afrontarse tales como la falta de transparencia de empresas generadoras de aplicaciones de IA, cuyos datos y algoritmos no desean compartir; o la uniformidad interna en los equipos de desarrollo y la eventual parcialidad en los resultados que ello puede conllevar; o los sesgos preexistentes en los datos que se utilizan para entrenar los modelos de IA; o, finalmente, el propio desconocimiento de las usuarias y usuarios sobre los riesgos de unilateralidad que puede tener y sobre la forma de identificarlos. Para acabar, Gil Iranzo comenta una serie de aplicaciones de IA, tanto gratuitas como de pago, tanto públicas como privadas,

que ponen en marcha estrategias para evaluar y/o reducir sesgos, advirtiendo de incorrecciones en las que a veces incurren algoritmos que pretenden ser más inclusivos. En lo referente a la violencia de género, la autora menciona efectos negativos y positivos de la IA, y reflexiona sobre el caso particular de los metadatos asociados a imágenes de agresores.

Cynthia Gálvez López es la autora del cuarto ensayo, al que titula *Entornos tecnológicos y violencias machistas: la IA en la era digital*. En él plantea cómo los entornos digitales se han convertido en un nuevo canal por el que se vehicula la violencia machista, que se ve favorecida por el anonimato que posibilita internet y sus herramientas. La autora reflexiona sobre el especial peligro que corren las niñas y adolescentes, más vulnerables y sin capacidad suficiente para protegerse de este tipo de violencia, y recuerda varios casos de agresión digital a mujeres en la comunidad que juega a videojuegos y en plataformas de realidad virtual. No obstante, la autora considera la IA como una oportunidad para luchar contra la violencia de género, a pesar de que una gestión inadecuada de esta tecnología puede reproducir sesgos de género preexistentes y perpetuar estereotipos machistas, de los que presenta algunos ejemplos como *chatbots* que generan mensajes sexistas, herramientas de reclutamiento que marginan a las mujeres, algoritmos de reconocimiento facial que fallan identificando rostros femeninos, o voces simuladas de mujeres en los asistentes virtuales que refuerzan una imagen de sumisión, entre otras situaciones. Seguidamente, Gálvez expone diversas herramientas que ofrece la IA para prevenir y detectar la violencia de género, para asistir a las víctimas, para la protección y seguridad digitales, para la intervención y eliminación de contenido nocivo, así como para investigar y recoger pruebas. Finalmente, recomienda transparencia, colaboración y privacidad como condiciones imprescindibles para que sea efectiva la lucha contra la violencia machista desde la IA, convirtiendo los entornos digitales en espacios más seguros para las mujeres.

La quinta aportación, titulada *Impacte de la intel·ligència artificial en la violència masclista i l'Agenda del Feminista*, ha sido escrita por Lourdes Muñoz Santamaría. La autora considera que la IA implica una nueva revolución dentro de la actual sociedad digital en la que vivimos y está teniendo un efecto directo en las oportunidades de las mujeres, razón por la cual el feminismo de datos propone una agenda de intervención y de regulación de los derechos digitales. Expone tres retos que esta agenda debe afrontar: 1) disponer de datos que visibilicen la realidad y las necesidades de las mujeres, 2) aplicar la perspectiva de género en todo el proceso de construcción de aplicaciones de IA desde el diseño del algoritmo hasta el impacto diverso que eventualmente puede tener en mujeres y varones, y 3) reducir la segregación horizontal de género en los equipos que desarrollan estas aplicaciones en consonancia con las responsabilidades éticas que les corresponden. Seguidamente, Muñoz se centra en la violencia de género en las redes sociales y analiza algunos casos en los que se observa la capacidad de multiplicación de las ofensas y agresiones a mujeres y el consiguiente efecto amplificado sobre las víctimas. Declara

que, para mantener una conducta ética, estas nuevas tecnologías deben desarrollar mecanismos que aseguren la privacidad y la seguridad de las mujeres, dado que constituyen un espacio de socialización y de exposición de ideas y opiniones. Asimismo, debe tenerse en cuenta que la brecha digital intergéneros coloca a las mujeres en una situación de desventaja y aumenta su vulnerabilidad en las situaciones de violencia machista. Para acabar, comenta estrategias para diseñar y utilizar la IA con el objetivo de proteger los derechos de las mujeres y participar en la prevención, tratamiento y erradicación de la violencia de género.

El siguiente trabajo es el de María Marta Elustondo y Miguel Ángel Blanco, *Indicadores de estereotipos en inteligencia artificial. Test para su detección*. Se trata de una aportación práctica —resultado de la larga experiencia obtenida en el mundo de la ingeniería y la docencia universitaria— que está destinada al reconocimiento del problema por parte de la gente común, aunque puede servir para cualquiera —especialista o no— que quiera aproximarse a los sesgos que esconde la IA. Elustondo y Blanco reconocen la fuerza transformadora de la IA y sus beneficios en diversos ámbitos, pero advierten de que esta tecnología puede esconder riesgos, como es la posible perpetuación de la violencia simbólica y estructural contra colectivos históricamente marginados, como son las mujeres, a causa de la forma en que se diseñan, desarrollan e implementan los sistemas de IA. Consideran trascendental que las personas que desarrollan esta tecnología sean conscientes del problema y que intervengan de manera proactiva para reducirlo; pero también, que el público usuario debe observar los resultados de la IA de manera crítica. Atendiendo a todo esto, plantean indicadores de reconocimiento de prejuicios vehiculados por las aplicaciones de IA y proponen un test que contiene 15 preguntas repartidas en cinco categorías dirigido a descubrir cuáles son y localizar dónde están los estereotipos cuando se usa la IA, de modo que pueda evaluarse el grado de equidad que manifiestan dichas aplicaciones. Incorporan una clave para interpretar los datos resultantes del test, una vez cumplimentado este. Se trata de una herramienta que han diseñado Elustondo y Blanco y para la que proponen, asimismo, un método de validación.

En el texto *La mujer en la era de la Ingeniería y la Tecnología: un camino hacia la igualdad*, Marta Rosa Poiasina García reflexiona sobre la relación de las mujeres con la IA, los efectos que esta tiene en el mundo académico y empresarial, y su implicación cada vez mayor en los cambios tecnológicos. Habla de la necesidad urgente de aumentar la presencia de las mujeres en todos los estamentos universitarios, en la ciencia y en las empresas de tecnología para diversificar el pensamiento, encontrar soluciones innovadoras y desarrollar herramientas que respondan a las distintas necesidades de una sociedad que es heterogénea. También considera imprescindible que se incremente el número de mujeres en cargos de liderazgo en el ámbito tecnológico y que se visibilicen sus logros para que puedan servir de referente y para retener a más mujeres en este sector. Asimismo, reflexiona sobre la importancia de desarrollar programas de corresponsabilidad entre varones y mujeres y de conciliación entre la vida personal, familiar y profe-

sional, así como de la trascendencia que tiene que las mujeres, en su incorporación a los ámbitos de la ciencia, la ingeniería y la tecnología, sobrepasen los objetivos de mercado planteando buenas prácticas y activando valores éticos. Por ello, anima a las ingenieras a ser agentes de cambio.

El libro acaba con un glosario en el que Mª Ángeles Calero intenta definir de manera sencilla todos los términos y expresiones técnicas relacionados con la IA, en particular, y con la tecnología, en general, que se han utilizado a lo largo de estas páginas, de modo que la lectura sea comprensible y que este vocabulario pueda acabar incorporándose en el acervo léxico común.

Esperamos que este libro sirva para abrir las mentes, alertar de los riesgos que pueden esconderse tras los beneficios más evidentes de las aplicaciones actuales de la IA, dar a conocer los efectos negativos que tiene en las mujeres una IA basada en una visión patriarcal del mundo, y para promover la ocupación y el aprovechamiento de los espacios virtuales, así como la neutralización de los sesgos de género, de la violencia simbólica y de la violencia real contra las mujeres que hoy existen y circulan campantes en dichos espacios.

Agradecemos a la Concejalía de Políticas Feministas del Ayuntamiento de Lleida, a Sílvia Puertas Novau —entonces jefa de la Unidad Técnica de Políticas de Igualdad— y a su equipo la iniciativa de organizar el Seminario *Inteligencia Artificial y violencias machistas. Retos y oportunidades*, en julio de 2024, que permitió un espacio de transmisión de conocimiento y de debate que fue el germen del libro que presentamos aquí como una de las actuaciones del *Pacto Local Contra las Violencias Machistas en Lleida*. También agradecemos a la Universidad de Lleida que haya apoyado la publicación, en papel y en abierto, de este libro.

Violències masclistes i intel·ligència artificial. Reptes i oportunitats

Unitat Tècnica de Polítiques d'Igualtat
Ajuntament de Lleida

1. Antecedents i context

1.1. Pacte Local contra les violències masclistes

El 20 de desembre de 2019, el Ple de l'Ajuntament de Lleida va aprovar la Moció dels Grups Municipals d'ERC-AM, PSC, JxCat Lleida i Comú de Lleida, per declarar Lleida Municipi Feminista i es va comprometre a aplicar el decàleg que acompanyava aquella declaració que concretava les deu mesures a dur a terme per efectiva la transformació feminista de la ciutat.

D'aleshores ençà, el decàleg s'ha anat implementant amb el desplegament de diferents actuacions en els àmbits de l'educació, l'urbanisme, la participació, el treball, els serveis d'atenció, entre altres. Amb l'inici del nou mandat (2023-2027) es va impulsar una de les accions del decàleg amb més impacte comunitari: el *Pacte Local Contra les Violències Masclistes*.

Amb data 13 d'octubre de 2023, el Departament d'Igualtat i Feminismes de la Generalitat de Catalunya va obrir la convocatòria de subvencions de concurrència competitiva per a Ens locals per finançar les despeses derivades d'accions per a la prevenció de les violències masclistes per als exercicis 2023 i 2024 (RESOLUCIÓ IFE/3431/2023, de 10 d'octubre). L'Ajuntament de Lleida va veure l'oportunitat de concórrer-hi presentant un projecte amb el títol "Pacte Local Contra les Violències Masclistes a Lleida", un projecte ambiciós amb una gran diversitat d'actuacions, inclosa l'acció que es descriu a

continuació, que dona títol a aquesta publicació i a les dues altres activitats que la complementen: *Intel·ligència Artificial i violències masclistes. Reptes i Oportunitats*.

1.2. Intel·ligència artificial (IA) i feminisme

La IA forma part de les nostres vides professionals i personals. La nostra quotidianitat està envoltada d'aplicacions digitals que ens connecten a internet per revisar un text, per saber la previsió meteorològica o dir-li a Alexa o a Siri que ens posin una música apropiada per al moment.

Les tecnologies estan al servei de les persones i les comunitats i, com el feminisme i la ciència feminista han posat en evidència de manera reiterada, no hi ha res que sigui neutre pel que fa a l'impacte diferencial que aquestes puguin tenir sobre els homes i les dones.

L'Instituto de las Mujeres va publicar, l'any 2023, l'informe realitzat per Lorena Jaume-Palasí amb el títol *Informe preliminar con perspectiva interseccional sobre sesgos de género en la Inteligencia Artificial*. L'autora posa de manifest els riscos dels anomenats sistemes automatitzats de generació de coneixement, el que popularment anomenem com IA, així com la necessitat de treballar per posar aquests sistemes al servei de la igualtat de gènere i de la transformació feminista. Autores i autors dels articles que recull aquesta publicació, parlen més extensament d'aquests riscos, des de diferents mirades, així com de les oportunitats de la IA per a la igualtat de gènere i per a l'impuls de les polítiques feministes.

La celeritat amb la qual la IA s'escampa com una taca d'oli entrant en els diferents àmbits de la vida personal, laboral i comunitària, demana un debat reposat per pensar en quina direcció ha de caminar l'avenc tecnològic. Un debat que eludeixi la mirada neutra que invisibilitza les dones i també els homes i dones dels col·lectius minoritaris (en allò real o en allò simbòlic) i que posi en evidencia les petjades androcèntriques dels continguts de la xarxa i dels algoritmes que els estructuren en forma de resposta.

Cal un debat seré, promogut des de les institucions públiques, per no deixar només en les mans privades de les grans empreses tecnològiques multinacionals i dels interessos del mercat, unes tecnologies decisives per a la vida de les persones i de les societats, a escala global, en el moment actual i en el futur més immediat.

1.3. Violències masclistes i IA

Les violències masclistes tenen dos territoris interconnectats: el real i el virtual. L'impacte de la violència que pateixen les dones en tots els àmbits de la vida personal, familiar i comunitària s'expandeix a les xarxes socials, generant violències que es perllonguen en

el temps i en l'espai. Aquestes violències trasbalsen la vida diària individual i col·lectiva i són, al mateix temps, el fet que mostra la persistència de les discriminacions de gènere vers les dones i el principal obstacle per a l'avenç en la igualtat de tracte i d'oportunitats.

Les aplicacions que ens ofereixen les tecnologies de la informació i la comunicació amb la IA obren noves oportunitats per fer prevenció de les violències masclistes, per informar i atendre les persones que les pateixen i també per establir lligams entre serveis i organitzacions. Posar al servei de la lluita contra les violències masclistes aquestes tecnologies és una oportunitat i una necessitat que cal explorar.

2. IA i violències masclistes. Reptes i oportunitats. Disseny i implementació d'aquest projecte

L'octubre de 2023 es va fer un primer esbós de la idea del projecte amb aquest mateix títol i es va incorporar com a una de les activitats que integraven el projecte *Pacte Local contra les Violències Masclistes a Lleida*, amb l'objectiu que pogués comptar amb finançament extern per a la seva execució. El disseny inicial preveia realitzar tres activitats: a) una recerca pilot sobre les violències digitals utilitzant la IA, contextualitzada en el municipi de Lleida; b) crear un espai de debat des d'una visió polièdrica sobre els reptes i oportunitats de la IA en relació amb les violències masclistes i elaborar un document per compartir les reflexions i c) elaborar un recull de solucions tecnològiques amb IA a problemes socials relacionats amb les violències masclistes.

El punt de partida va ser posar en relació gent experta del món tecnològic i de l'àmbit social, amb la intenció de crear sinèrgies i apropar col·lectius professionals, acadèmics i socials que solen treballar en àmbits separats, així com per integrar la mirada de les dones del territori i de les persones amb responsabilitat política i tècnica municipal a aquests debats.

Un cop confirmat el finançament del projecte, la primera actuació va consistir en crear el grup de treball integrat per Rosa Mª Gil, professora agregada del Departament Enginyeria Informàtica i Disseny Digital i coordinadora del Grau en *Dissenys Digitals* de la Universitat de Lleida; Mª Ángeles Calero, catedràtica del Departament de Filologia i Comunicació i coordinadora del Màster en *Estudis de Gènere i Gestió de Polítiques d'Igualtat* de la Universitat de Lleida; i Rosa Prats, professora d'Ensenyament Secundari, coordinadora del projecte *Technovation Girls* a Lleida i membre de l'equip directiu de l'Entitat Espiral. La coordinació del grup i de tot el projecte ha anat a càrrec del personal tècnic de la Regidoria de Polítiques Feministes, liderat en aquell moment per Sílvia Puertas. Aquest equip ha estat el grup motor que ha impulsat les activitats del projecte.

La metodologia emprada per definir la realització de les tres activitats (recerca, debat, recull de solucions tecnològiques) va ser mitjançant la realització d'un seminari amb la participació de persones expertes de l'àmbit tecnològic i social, personal tècnic i polític

municipal i conselleres integrants del Consell Municipal de les Dones. La participació va ser tancada, per invitació expressa, amb un nombre màxim de 30 persones, el nom de les quals es pot consultar a l'Annex 1.

El seminari es va dur a terme el dia 15 de juliol de 2024 a Lleida, al Parador Nacional *El Roser*, de les 9:30h. a les 14:30h, en format presencial i amb la participació online de tres persones expertes que es trobaven a Argentina i Canadà.

La primera part va ser en sessió plenària. Un cop donada la benvinguda institucional a les persones participants, es va procedir a:

- presentar el projecte: objectius, metodologia, accions, resultats;
- explicar l'aplicació de l'*Enfocament de Gènere i basat en Drets Humans* (EGiBDH)[1] i la metodologia "Dones i ciutat"[2]; i
- detallar el funcionament de la jornada, amb la distribució de les persones assistents en grups, així com les tasques a realitzar i els resultats esperats de la sessió plenària i dels grups.

---

1. L'*Enfocament de Gènere i Basat en Drets Humans* (EGiBDH) és un plantejament teòric, metodològic i polític que sorgeix a l'entorn de la Cooperació Internacional. Suposa un canvi de paradigma que ofereix una gran capacitat d'abordatge de les desigualtats. La seva aplicació resultat molt encertada en aquelles accions que tenen per objectiu lluitar contra les violències masclistes i les desigualtats de gènere. La guia editada per l'Agencia Catalana de Cooperació al Desenvolupament / Generalitat de Catalunya, amb el títol *EGiBDH Enfocament de gènere i basat en Drets Humans* recull en la pàgina 3 que "l'EGiBDH és també un marc conceptual i d'anàlisi, una estratègia, una metodologia i alhora una praxis que cerca analitzar i eradicar les causes estructurals que provoquen vulneració dels Drets Humans, desigualtats i discriminació envers les dones, en tots els àmbits (econòmic, laboral, polític, social, cultural".

2. *Dones i ciutat* va ser un projecte europeu amb el mateix títol, desenvolupat entre juliol de 1996 i juny de 1998, cofinançat per la Comissió Europea dins del *IV Programa d'Acció Comunitària a Mig Termini par a la Igualtat d'Oportunitats entre Dones i Hombres*. La direcció del projecte va anar a càrrec de la Fundació Maria Aurèlia Capmany, amb la participació de la Diputació de Barcelona, els ajuntaments de Cardona, Cerdanyola, Esplugues de Llobregat, Lleida, Sant Feliu de Llobregat, Terrassa i Vilafranca del Penedès, Barcelona, Consell de Dones de la Comunitat de Madrid, Donostia, el VES Emancipatieubureau Zuid (Holanda) i Arbeit und Leben, de Halle (Sajonia-Anhalt). En el marc d'aquest projecte es van crear el fòrums "Dones i ciutat".

En el marc d'aquest projecte els ajuntaments participants, van crear als seus territoris els fòrums locals *Dones i ciutat*, espais de participació per a les dones, amb la finalitat de saber la forma en què les dones viuen, ocupen, transiten, desitgen i pensen els espais públics (Bofill, Dumenjó i Segura, 1998). Els fòrums locals havien d'estar integrats per dones amb rols i perfils diversos per garantir la participació de dones amb responsabilitat política i tècnica, dones del teixit associatiu i dones no vinculades a organitzacions que volguessin compartir les seves vivències sobre els usos de l'espai públic. Aquesta metodologia participativa plantejava igualment que calia posar al mateix nivell tots els discursos i posar en valor totes les aportacions (tant les compartides des de l'expertesa tècnica o política, com des de la vivència quotidiana). Aquest enfocament demanava, finalment, trobar punts de consens (sobre la diagnosi, les prioritats a atendre, les possibles solucions, entre altres). La riquesa d'aquest enfocament i dels seus resultats va ser la que va orientar la vetlla perquè el Seminari *IA i violències masclistes. Reptes i oportunitat*s s'estructurés en tres grups de treball i perquè cada grup de treball —tot i que mixt, amb predomini de dones— comptés amb dones que responguessin als perfils diversos que contempla aquesta metodologia.

La segona part de la jornada va consistir en el treball en grups. Les persones participants es van dividir en tres grups, segons una distribució prèvia definida per la coordinació tècnica de l'activitat tenint en compte la idoneïtat de cada persona amb la temàtica del grup assignat i amb la finalitat de garantir una presència equilibrada de persones d'expertes del món TIC i del món social a cada grup, així com la participació de persones amb responsabilitat política municipal i dones en representació del moviment associatiu local.

De manera paral·lela Guadalupe Amongero Noriega i Judit Sanmartí Dea, il·lustradores de *Niu Gremi Associació*, van elaborar els resums visuals (que apareixen a l'Annex 2), a mode de conclusions gràfiques del seminari, sobre el conjunt general de la jornada i sobre el treball de cada grup.

Els objectius del seminari van ser:

- **General:** Debatre a l'entorn de les oportunitats de la IA per generar nou coneixement sobre les dones i, en partícula, sobre les violències masclistes, fent èmfasi en les violències digitals, evidenciant els reptes i les oportunitats des de la perspectiva feminista i de la realitat municipal.
- **Específics:**
  - Grup 1, coordinat per Rosa Mª Gil Iranzo:
    Dissenyar i encarregar una investigació amb IA sobre la realitat local, en relació amb les violències masclistes, especialment en l'àmbit digital.
    Resultat: definició dels objectius (generals i específics) i de la metodologia; concreció de metodologies i cronograma: identificació dels resultats esperats, i proposta de la forma de presentar les dades obtingudes.
  - Grup 2, coordinat per Mª Ángeles Calero Fernández:
    Dissenyar i coordinar una publicació que reculli la mirada multidisciplinar sobre la IA i els seus reptes i oportunitats, amb l'objectiu de millorar el coneixement sobre les violències masclistes d'una manera global i d'aplicar aquests coneixements a nivell local, aprofitant aquesta tecnologia.
    Resultat: document amb la definició del tipus de publicació, continguts, estructura i format.
  - Grup 3, coordinat per Rosa Prats Novau:
    Elaborar propostes sobre com es poden donar solucions tecnològiques als problemes socials relacionats amb les violències masclistes a nivell local.
    Resultat: informe amb el recull de les aportacions obtingudes i propostes concretes d'activitats per desenvolupar a nivell local.

Les conclusions del Seminari van donar lloc a la feina posterior realitzada per cada coordinadora de grup, concretant la línia de treball per avançar en cada encàrrec específic i assolir els resultats finals.

A mode de resum, els resultats han estat els següents:

Grup 1. Recerca

En el grup de treball es van plantejar diferents propostes:

1. Identificar potencials situacions d'agressions al municipi de Lleida a partir de l'anàlisi de continguts de la xarxa.
2. Desenvolupar una eina d'ajuda 24h per informar i atendre dones del municipi de Lleida en situació de violència masclista.
3. Detectar estereotips i percepció de la violència entre el jovent del municipi de Lleida.

Aquestes tres idees es van compartir amb el grup motor i es va veure la possibilitat recórrer a eines com el *Test de Bechdel*, utilitzat per analitzar els estereotips de gènere en el cinema, i adaptar-lo per detectar les violències masclistes en els consums audiovisuals del jovent, a partir d'una experiència pilot de recerca, amb gent jove de Lleida. La recerca es va concretar amb:

- el disseny del qüestionari, realitzat de manera col·laborativa per Rosa Mª Gil i Iolanda Tortajada (autores, ambdues, de sengles articles d'aquest llibre), que incloïa una descripció de si mateix/a que havia de fer cada informant;
- la concreció del jovent que havia de respondre el test: alumnat del Grau de *Dissenys Digitals* de la Universitat de Lleida;
- el tractament dels resultats amb IA i l'anàlisi amb perspectiva de gènere, quantitativa i qualitativa:
  - Qüestionaris: presentació de resultats per a l'anàlisi quantitativa i qualitativa amb perspectiva de gènere sobre la percepció de la presència de violències masclistes en els continguts audiovisuals de les persones enquestades.
  - Presentació de quatre imatges generades per la IA de cada alumne/a a partir de les seves auto-descripcions (*prompts*) i anàlisi qualitativa amb perspectiva de gènere, tant de les auto-descripcions, com de les imatges obtingudes amb IA.

L'article de Rosa Mª Gil Iranzo inclòs en aquesta publicació contextualitza i aprofundeix sobre el desenvolupament i descripció de la investigació.

Aquesta recerca s'ha plantejat com una prova pilot que adapta una eina àmpliament coneguda en el món de la comunicació audiovisual (el *test de Bechdel*, com hem comentat anteriorment) i que utilitza la IA per conèixer i contextualitzar les respostes. Permet valorar el grau de percepció de la violència masclista en els consums culturals audiovisuals realitzats amb ordinadors, mòbils, tablets, TV. Aquests consums els realitzen persones que s'auto-defineixen com a homes o dones, majoritàriament. Les auto-definicions permeten analitzar quins són els models de feminitat i masculinitat predominants, i veure si estan més a prop dels valors més coherents amb la igualtat de gènere i d'oportunitats per a dones i homes o be en valors més propers als rols de gènere tradicionals i estereotipats.

La voluntat final d'aquest treball de recerca era doble:

- que es pogués replicar altres universitats i equips d'investigació, amb la finalitat d'ampliar i contrastar resultats, millorar en allò que sigui possible les eines metodològiques creades i facilitar orientacions des del món acadèmic a altres administracions públiques per lluitar contra les violències masclistes; i
- crear sinèrgies entre experteses tecnològiques i socials, aprofitant les eines de la IA per avançar en la transformació feminista.

Grup 2. Debat

El plantejament d'aquest grup de treball va ser generar diàleg entre les persones assistents sobre els reptes i oportunitats de la IA en relació amb les violències masclistes i sobre com obrir l'interès d'un públic ampli per aquest tema, més enllà del món acadèmic i de l'expertesa. El grup va consensuar el disseny d'una publicació de caire divulgatiu, integrada per articles d'autors i autores d'àmbits de coneixement diversos, amb reflexions personals i evidència científica des de la mirada de l'enginyeria informàtica, de la comunicació audiovisual, de l'educació, de l'ètica i la psicologia, el disseny gràfic o el treball social.

Les conclusions d'aquest grup de treball van ser recollides i elaborades per la coordinadora, Mª Ángeles Calero Fernández, que s'ha ocupat, un cop finalitzat el projecte, de donar-li continuïtat, convidant les autores i autors a participar en aquesta publicació. Això ha permès assolir un text coherent que permet al públic lector familiaritzar-se amb els principals debats sobre els perills i les oportunitats que planteja la IA, des d'una perspectiva de gènere, així com sobre l'aprofitament d'aquesta tecnologia per avançar en l'eradicació de les violències masclistes. Finalment, el resultat del GRUP 2 és aquesta publicació, que convida a conèixer més extensament la globalitat del projecte, en les seves tres vessants, i el seu encaix amb el Pacte Local contra les Violències masclistes a Lleida.

Grup 3. Recull de solucions tecnològiques

La coordinadora del grup Rosa Prats Novau va plantejar la sessió a partir de dues preguntes: la primera, per contextualitzar la problemàtica de les violències masclistes en l'àmbit local i en el moment actual; i la següent, per generar idees sobre possibles solucions tecnològiques als problemes identificats a partir del debat suscitat amb la primera pregunta.

PREGUNTA 1: Com hem arribat a la situació actual? Quins són els factors desencadenants de la violència masclista? Selecció de respostes/debats:

- El rol assignat a les dones i els factors relacionats amb la història i circumstàncies personals: interrelació entre context social i context personal (edat, classe, raça, salut...)

- Involució o avenç en la igualtat. Percepció de relaxació amb els comportaments del jovent. Les noies toleren comportaments masclistes que semblaven superats.
- Perills i oportunitats de les xarxes i d'*influencers*.
  - Perills: agressions, manipulació, control.
  - Oportunitats:
    - Ajuden a treure a la llum abusos sexuals en l'àmbit familiar, laboral, esportiu, de lleure.
    - Faciliten informació de recursos, contacte amb grups de suport, alerten de possibles situacions de risc.
- Posar en valor les lluites feministes de les generacions de dones anteriors i connectar amb el jovent i el moviment feminista actual.

PREGUNTA 2: Un cop fet un repàs a l'escenari actual ens preguntem: a) Quins són els col·lectius a qui dirigir-se?, b) Quines serien les nostres eines actuals?, i c) Quines podrien ser les propostes, els productes, la planificació?

La resposta a aquesta Pregunta 2 es va recollir en una graella que organitza les propostes del grup a partir de quatre descriptors: 1) tipus de proposta, 2) públic destinatari, 3) objectiu de la proposta i 4) recursos necessaris per desenvolupar-la.

Les solucions tecnològiques proposades, tenen en compte la diversitat de situacions i necessitats de les persones i dels col·lectius, com per exemple l'edat, les competències digitals, la diversitat funcional, les barreres d'idioma, entre altres. D'altra banda, també tenen en compte si s'adrecen a un públic mixt o bé només a homes o només a dones de manera específica, al jovent, a gent adulta o gran o bé a un públic ampli.

Es van proposar, entre altres solucions amb IA:

- Aplicacions formatives sobre masculinitats igualitàries.
- Testimoniatge d'*influencers* que:
  - ajudin a sortir de la violència a altres persones, serveixin de model de referència,
  - contrarestin continguts masclistes a les xarxes.
- Aplicacions que facilitin posar una denúncia, conèixer els serveis locals d'atenció i suport a persones que estan en situació de violència masclista.

El grup va proposar, així mateix, que les eines a desenvolupar hagin de tenir en compte un conjunt de consideracions generals, entre aquestes:

- Tenir i utilitzar indicadors per fer el seguiment i l'avaluació dels resultats (quantitatius i qualitatius).
- Partir de la informació local, dels recursos i contacte del territori, amb els agents que presten serveis a la població del municipi de Lleida (mapa de recursos).
- Aplicar l'*Enfocament de la quíntuple hèlix*, pel fet que aquest representa una interacció col·lectiva i comunitària dels cinc subsistemes o hèlixs: 1) el sistema polític; 2) el sistema educatiu; 3) el sistema econòmic; 4) l'entorn natural (sostenibilitat); 5) el públic, basat en els mitjans de comunicació i en la cultura i/o societat civil.

- Treballar des de la coeducació, la igualtat de gènere i la coresponsabilitat en l'àrea de la cura.
- Buscar referents masculins de valor. Posar de moda el #bontracte de manera que ser masclista no estigui ben vist. I que els homes no se sentin atacats pel discurs.
- Evitar les escletxes digitals per motius d'edat, gènere, situació econòmica, diversitat funcional, competències lingüístiques.
- Tenir en compte la formació prèvia necessària per identificar les situacions de violència i facilitar que es puguin denunciar i demanar ajuda.

En el debat final es va recollir la idea que s'ha d'anar més enllà de la declaració d'intencions i que, per avançar, seria molt oportú elaborar un Pla Estratègic de lluita contra les violències masclistes.

Com a conclusió, el GRUP 3 va aportar 16 propostes concretes i va suggerir idees de com desenvolupar-les, així com també referències sobre solucions similars ja existents que poden inspirar les que es puguin dur a terme a Lleida.

La tasca de coordinadora de Rosa Prats Novau va continuar amb la coordinació i organització de la HACKATÓ[33] *Solucions tecnològiques a problemes socials relacionats amb les violències masclistes*, una activitat adreçada a joves estudiants de les escoles municipals. La *hackató* es va realitzar mitjançat l'*Associació Espiral, Educació i Tecnologia*, amb el seu equip de dinamització, coordinat per Josep Marés, el 25 de setembre de 2024, al Campus Universitari de l'Escola Politècnica Superior de la Universitat de Lleida, de 9h. a 14h.

A partir d'una metodologia de treball que utilitza l'empatia com a eix generador d'idees, 50 alumnes organitzats en petits grups, nois i noies de l'Escola d'Art Municipal *Leandre Cristòfol* i de l'Institut Municipal d'Ocupació (de La Casa d'Oficis i del mòdul formatiu d'Educació de Lleure) i amb l'acompanyament del seu professorat, van disposar de cinc hores per pensar i proposar solucions tecnològiques als problemes plantejats.

L'activitat va finalitzar amb la presentació plenària de les 9 propostes elaborades, una per grup, i va permetre visualitzar el talent del jovent lleidatà que, tot i cursar estudis que res tenen a veure amb el treball social ni amb la enginyeria informàtica, van saber connectar amb el què són i què signifiquen les violències masclistes i com contribuir a eradicar-les. Aquesta experiència va buscar el doble objectiu de sensibilitzar el jovent contra aquest tipus de violència i fer-lo protagonista de les solucions per evitar-la. La *hackató* del 25 de setembre de 2024 va posar el punt i final a l'execució d'aquest projecte obrint, juntament amb la resta d'activitats, noves línies d'intervenció, sinèrgies i treball en xarxa.

---

3. L'Associació Espiral recull en la guia de l'activitat que se li va encomanar que "el terme *hackató* prové de les comunitats de *hackers* i programadors, referint-se a les trobades que organitzen per desenvolupar aplicacions i noves solucions a problemes o reptes". De manera anàloga, utilitza el terme *hackató* per realitzar tallers amb persones no expertes, principalment població adolescent i jove, per tal que, a partir de dinàmiques de participació i de generació d'idees, elaborin solucions tecnològiques a reptes socials.

El seminari *IA i violències masclistes. Reptes i oportunitats* va facilitar l'assoliment dels reptes plantejats inicialment en el projecte. Això no hauria estat possible sense la complicitat i l'entusiasme de totes les persones que han col·laborat conduint-lo a bon port. A totes elles, un agraïment molt especial.

3. Referencias bibliográficas

Agencia Catalana de Cooperació al Desenvolupament / Generalitat de Catalunya (s/a). *EGiBDH Enfocament de gènere i basat en Drets Humans*. Disponible en: https://cooperaciocatalana.gencat.cat/ca/com-ho-fem/egibdh/

Bofill, Anna; Dumenjó, Rosa María i Segura, Isabel (1998). *Las mujeres y la Ciudad. Manual de recomendaciones para una concepción del entrono habitado desde el punto de vista del género*. Barcelona: Fundació Maria Aurèlia Capmany.

Jaume-Palasí, Lorena (2023). *Informe preliminar con perspectiva interseccional sobre los sesgos de género en la Inteligencia Artificial*. Madrid: Instituto de las Mujeres - Ministerio de Igualdad.

Annex 1. Participants al Seminari Intel·ligència Artificial i violències masclistes. Reptes i oportunitats

Grup 1: recerca

Coordinadora: Rosa María Gil Iranzo

Ponents:

Miguel Ángel Blanco, especialista en Criptografia i Seguretat Teleinformàtica.

Cintia Guerrero Tapia, Iniciativa Barcelona Open Data (IBOD).

Iolanda Tortajada, professora titular de Comunicació Audiovisual a la Universitat Rovira i Virgili (URV).

Graciela Atencio, Associació La Sur – Feminicidio.Net.

Ramon Arnó, professor associat de Dret de la Universitat de Lleida i consultor.

Manel Marin, OAC, Policia Mossos d'Esquadra, Generalitat de Catalunya.

Montse Robles López, tècnica del Departament d'Igualtat i Feminismes de la Generalitat de Catalunya.

Grup 2: publicació

Coordinadora: Mª Àngeles Calero Fernández

Ponents:

Marta Rosa Poiasina, matemàtica, antiga professora de la Universidad Tecnológica Nacional (Buenos Aires).

Alaitxz Sáez Suárez, Dones en Xarxa (DEX).

Verònica Martínez, comissionada d'alcaldia de transversalitat de les polítiques feministes.

Víctor Merino, professor agregat del Departament de Dret Públic de la Universitat Rovira i Virgili (URV).

Ana Florista Izquierdo, portaveu del grup municipal Partit Popular de l'Ajuntament de Lleida.

Tania Puyol Castet, tècnica del Departament d'Igualtat i Feminismes de la Generalitat de Catalunya.

Grup 3: recull posibilitats tecnològiques

Coordinadora: Rosa Mª Prats Novau

Ponents:

Cynthia Gálvez, experta en TIC, doble grau Enginyera i Disseny Gràfic.

Leonor Cantera, professora titular de Psicologia de la Universitat Autònoma de Barcelona.

Aranzazu Nieto, infermera del Centre Jove de Salut Sexual, Regidoria de Salut Pública de l'Ajuntament de Lleida.

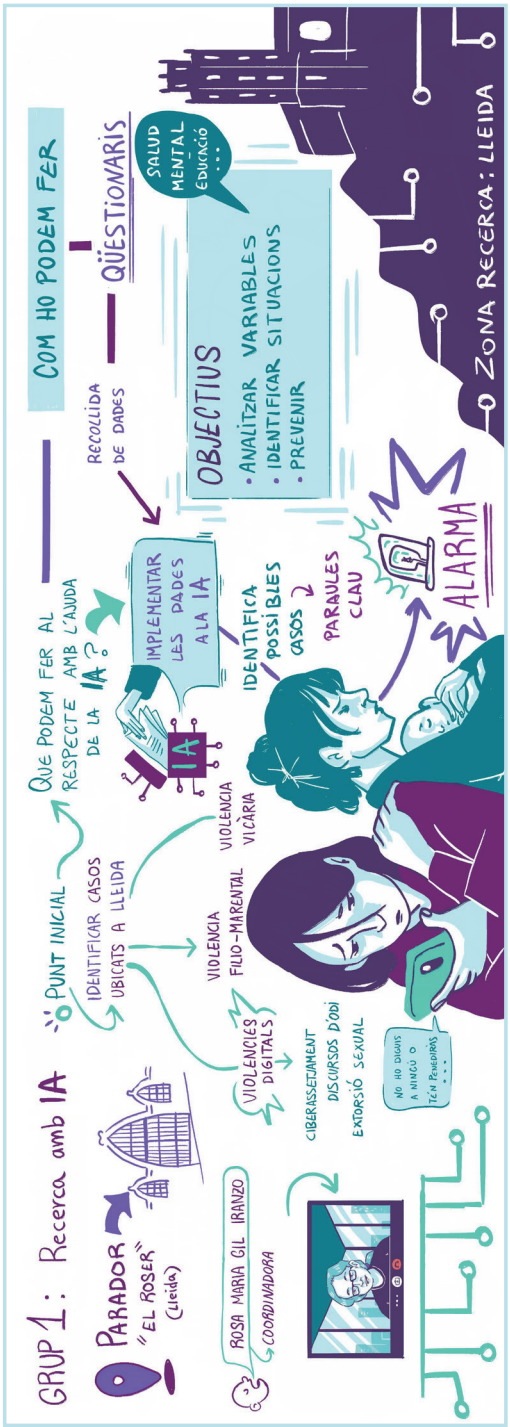
Maria Marta Eleustondo, experta en prevenció i intervenció en violències (Buenos Aires).

Neus Caufapé, regidora del grup municipal JuntsxCat.

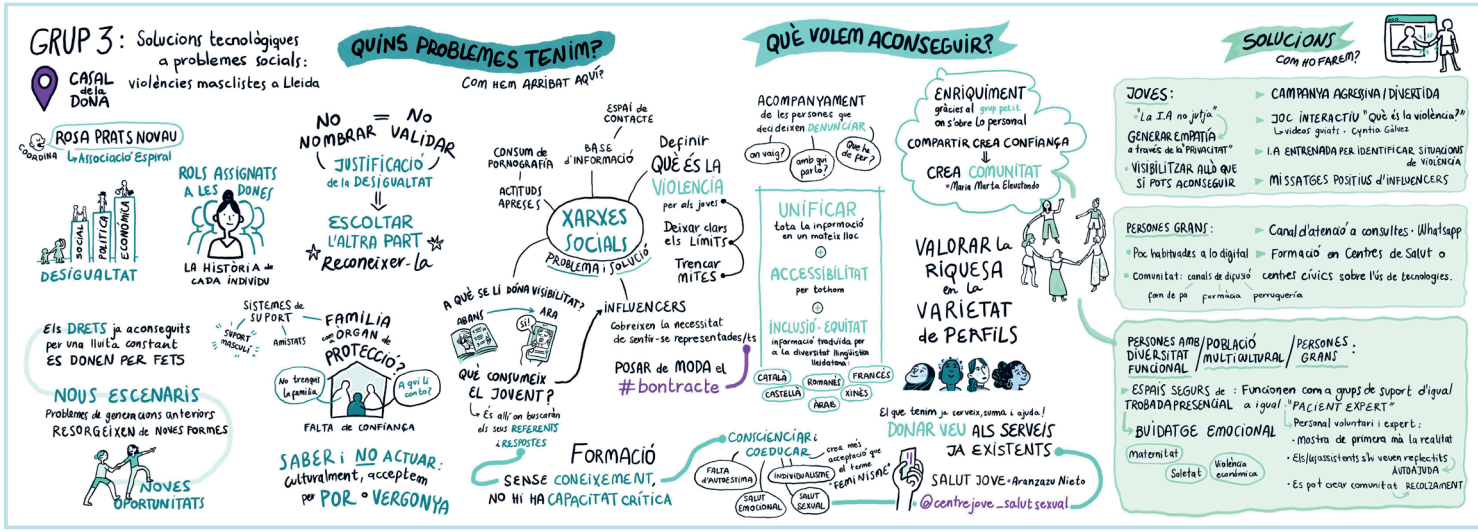
Claudia Lara Castellón, tècnica del Departament Igualtat i Feminismes de la Generalitat de Catalunya.

M. Antònia Roca, consellera del Consell Municipal de les Dones - Comissió Permanent.

Annex 2. Resum visual del treball dels grups a la Jornada




GRUP 2 : Dissenyar i elaborar una PUBLICACIÓ
PARADOR "EL ROSER" (Lleida)
QUE SERÀ?
Publicació tipus llibre per a DIVULGAR com fer bon ús de la IA
MARIA ANGELES CALERO FERNÁNDEZ
Coordinadora
¡EL PROBLEMA NO ÉS LA HERRAMIENTA, SINO CÓMO SE USA!
-MARTA POIASINA
iA
Pàg 3 / 5
LA IA COM UNA EINA DE PREVENCIÓ I AJUDA
¡IMATGE IA? ¡ILUSTRAT
OBJECTIUS
EDUCAR
INFORMAR
PREVENIR
DIVULGAR
PÚBLIC OBJECTIU
• JOVENT, ADOLESCENTS
• COL·LECTIUS FEMINISTES
• GRAN PÚBLIC
CONTINGUT
• RELATS de víctimes o gent propera...
• PAUTES per a prevenir mals usos
DEEP FACE PORNOGRÀFIC


GRUP 3: Solucions tecnològiques a problemes socials: violències masclistes a Lleida
CASAL de la DONA
QUINS PROBLEMES TENIM?
COM HEM ARRIBAT AQUÍ?
ROSA PRATS NOVAU
COORDINA Associació Espiral
DESIGUALTAT
SOCIAL
POLÍTICA
ECONÒMICA
ROLS ASSIGNATS A LES DONES
LA HISTÒRIA de CADA INDIVIDU
NO NOMBRAR = NO VALIDAR
JUSTIFICACIÓ de la DESIGUALTAT
ESCOLTAR L'ALTRA PART
Reconeixer-la
XARXES SOCIALS
PROBLEMA i SOLUCIÓ
CONSUM de PORNOGRAFIA
ACTITUDS APRESES
BASE d'INFORMACIÓ
ESPAI de CONTACTE
Definir QUÈ ÉS LA VIOLÈNCIA per als joves
Deixar clars els LÍMITS
Trencar MITES
A QUÈ SE LI DONA VISIBILITAT?
AVANS
ARA
QUÈ CONSUMEIX EL JOVENT?
És allà on busquen els seus REFERENTS i RESPOSTES
INFLUENCERS
Cobreixen la necessitat de sentir-se representades/ts
POSAR de MODA el #bontracte
SISTEMES de SUPORT
SUPORT MASCULÍ
AMISTATS
FAMÍLIA com ÒRGAN de PROTECCIÓ?
No trenquis la família
A qui li conta?
FALTA de CONFIANÇA
Els DRETS ja aconseguits per una lluita constant ES DONEN PER FETS
NOUS ESCENARIS
Problemes de generacions anteriors RESORGEIXEN de NOVES FORMES
NOVES OPORTUNITATS
SABER i NO ACTUAR: culturalment, acceptem per POR o VERGONYA
FORMACIÓ SENSE CONEIXEMENT, NO HI HA CAPACITAT CRÍTICA
CONSCIENCIAR i COEDUCAR
FALTA d'AUTOESTIMA
INDIVIDUALISME
SALUT EMOCIONAL
SALUT SEXUAL
crea més acceptació que el terme FEMINISME
QUÈ VOLEM ACONSEGUIR?
ACOMPANYAMENT de les persones que decideixen DENUNCIAR
on vaig?
amb qui parlo?
Que he de fer?
UNIFICAR tota la informació en un mateix lloc
ACCESSIBILITAT per tothom
INCLUSIÓ · EQUITAT informació traduïda per a la diversitat lingüística ciutadana
CATALÀ
CASTELLÀ
ROMANÈS
ÀRAB
FRANCÈS
XINÈS
ENRIQUIMENT gràcies al grup petit on s'obre lo personal
COMPARTIR CREA CONFIANÇA
CREA COMUNITAT
•Maria Marta Eleustondo
VALORAR la RIQUESA en la VARIETAT de PERFILS
El que tenim ja serveix, suma i ajuda!
DONAR VEU ALS SERVEIS JA EXISTENTS
SALUT JOVE · Aranzazu Nieto
@centrejove_salutsexual
SOLUCIONS
COM HO FAREM?
JOVES:
"La I.A no juja"
GENERAR EMPATIA a través de la "PRIVACITAT"
VISIBILITZAR ALLÒ QUE SÍ POTS ACONSEGUIR
CAMPANYA AGRESSIVA / DIVERTIDA
JOC INTERACTIU "Què és la violència?"
↳ videos guiats · Cynthia Gálvez
I.A ENTRENADA PER IDENTIFICAR SITUACIONS de VIOLÈNCIA
MISSATGES POSITIUS d'INFLUENCERS
PERSONES GRANS:
Poc habituades a lo digital
Comunitat: canals de difusió
fan de pa
farmàcia
perruqueria
Canal d'atenció a consultes · Whatsapp
Formació en Centres de Salut o centres cívics sobre l'ús de tecnologies.
PERSONES AMB DIVERSITAT FUNCIONAL / POBLACIÓ MULTICULTURAL / PERSONES GRANS:
ESPAIS SEGURS de: Funcionen com a grups de suport d'igual
TROBADA PRESENCIAL a igual "PACIENT EXPERT"
BUIDATGE EMOCIONAL
Maternitat
Soletat
Violència econòmica
Personal voluntari i expert:
· Mostra de primera mà la realitat
· Els/les assistents s'hi veuen reflectits
AUTOAJUDA
· Es pot crear comunitat RECOLZAMENT

La IA y las mujeres: una relación no tan inteligente

Iolanda Tortajada
Cilia Willem
Universitat Rovira i Virgili

1. El porno *deepfake* y el caso Almendralejo

A finales del año 2023 comenzó a circular una gran cantidad de fotos de chicas desnudas en grupos de WhatsApp de la localidad de Almendralejo (Badajoz). Los archivos no habían salido de los ordenadores ni de los teléfonos de las protagonistas de las imágenes. Ni siquiera eran reales: habían sido creadas con los rostros de las menores superpuestos sobre cuerpos que no eran los suyos y que aparecían sin ropa y/o en posturas pornográficas generadas por una inteligencia artificial (IA). Se trataba de un caso del llamado porno *deepfake*. Fueron más de veinte chicas las afectadas por aquella distribución ilícita de imágenes en diferentes grupos de mensajería instantánea de cuatro centros escolares.

Las *apps* y páginas web que ofrecen la posibilidad de crear este tipo de imágenes pornográficas se publicitan abiertamente con frases como "desviste a la chica que tú quieras" o "desnuda a cualquier mujer gratis", lo que pone en evidencia no solo que el objetivo principal son las mujeres, sino también que la intención de sus desarrolladores es presentar y vender su producto como un juego inocente, una diversión, algo lúdico. No es ninguna novedad que el cuerpo desnudo de una mujer —real o fantaseado— sea objeto de las miradas compartidas masculinas. Las *majas* vestida y desnuda, cuadros pintados por Francisco de Goya a principios del siglo XIX por encargo de su amigo Manuel Godoy, son un ejemplo —de manual— de la pornificación de los cuerpos femeninos —con o sin su consentimiento— para el placer compartido de un grupo de hombres.

Aquellas dos pinturas estaban colgadas una delante de la otra de modo que, mediante un ingenioso mecanismo de poleas, la maja desnuda apareciera detrás de la maja vestida, como un artilugio erótico del gabinete secreto del entonces primer ministro de Carlos IV. Hoy en día estos artilugios son digitales —y, por tanto, más fáciles y rápidos de compartir con muchos más hombres— pero, igual que antes, dan pie a prácticas de las cuales suelen sacar provecho los mirones masculinos mientras que ellas —las observadas— más bien sufren las consecuencias. Beverly Skeggs (2004) ya demostró el valor de cambio que estas prácticas pueden tener para los chicos que las practican: el intercambio de fotos "sexis" de sus compañeras de clase puede reportarles una mayor acumulación de valor social y una subida en el ranking de prestigio entre los "colegas".

En definitiva, en el régimen de representación patriarcal, la mirada masculina sobre el cuerpo femenino siempre ha sido compartida, competitiva, comparativa. Sin embargo, como el caso de Almendralejo ha puesto de manifiesto, la generación automática de cuerpos desnudos con rostros reales añade un nivel de violencia contra las mujeres difícil de gestionar ante la dificultad de establecer la autoría de dichas imágenes y su rápida y descontrolada extensión.

Los datos de los estudios que analizan este tipo de prácticas (por ejemplo, los del organismo regulador británico Ofcom) muestran que la gran mayoría de *deepfakes* sexualmente explícitos —también llamados *deepnudes*— son protagonizados por mujeres, muchas de las cuales sufren trastornos de estrés postraumático o ansiedad como resultado de ser el objetivo de tales imágenes. Los *deepnudes,* aunque no sean "reales", pueden causar graves daños muy auténticos a las personas. La *Ley Orgánica de Libertad Sexual* en España establece que el daño que pueden producir en el honor, la intimidad personal y la autoestima de la víctima la creación y la posterior difusión de imágenes realistas generadas por IA sin su consentimiento es similar al que se produce con una fotografía "real"; de ahí que esto último se considere delito. En el caso de Almendralejo, además, las afectadas son menores y, por ello, estamos ante un caso de pornografía infantil.

Algunas mujeres con poder simbólico y/o con una gran exposición pública, como Rosalía, Emma Watson, Ana de Armas o Inés Arrimadas, también han sufrido este tipo de violencia. Las expertas interpretan estas prácticas denigrantes contra figuras públicas como una forma de "disciplinar" a las mujeres exitosas y con poder de convocatoria, y de devolverlas a su posición de inferioridad.

Es importante tener en cuenta que, aunque a veces las concibamos como algo separado, las prácticas *online* y *offline* están interconectadas. La vida en espacios virtuales no deja de ser vida real. Las causas y las consecuencias de las violencias machistas que se producen en cualquiera de estos espacios son las mismas, tanto para personas anónimas como para personajes públicos. Es importante, además, analizar las conexiones entre ambos entornos: ¿qué usos y apropiaciones de la IA promueven la violencia machista?, ¿por parte de quién?, ¿cómo y en qué circunstancias?, ¿qué intereses (comerciales, políticos, ideológicos…) hay detrás? Son preguntas que pueden ayudarnos a comprender

unas prácticas como estas, que se generan con suma rapidez y ponen en circulación contenido que causa daño o perjudica gravemente a las mujeres, tanto anónimas como públicas.

2. La IA y los sesgos de género

La IA es una herramienta que puede provocar reacciones bastante contrastadas: hay quienes la abrazan entusiastas y hacen pública exaltación de sus posibilidades, pero también hay quienes la rechazan o la denuestan. Estos extremos nos impiden, a menudo, tener en cuenta de forma simultánea tanto los riesgos que entraña como las oportunidades que ofrece. La rapidez con la que se han generado discursos sobre la IA nos induce a pensar que llegamos tarde y esto favorece un cierto simplismo, ya sea de tintes optimistas, ya sea con cierto sesgo apocalíptico (Eco, 1995). La realidad es más compleja: la tecnología forma parte integral de nuestras vidas para bien y para mal, también en el caso de la socialización del género. Los estudios sobre redes sociales realizados hasta el momento aportan pruebas significativas sobre la importancia de las prácticas *online* para la construcción de nuestra identidad de género, sobre todo en la adolescencia. Las feminidades y las masculinidades que encarnamos y exhibimos no son ajenas a lo que da 'popularidad' o viralidad en las redes. Nos corresponde, pues, identificar qué elementos considerados originales, divertidos o auténticos por los propagadores del contenido sexista contribuyen a dicha propagación. A fin de cuentas, los modelos generativos de la IA transforman contenido (y han sido entrenados con datos) existente(s) en internet, y no tienen aún capacidad para distinguir entre lo real y la ficción, lo apropiado y lo que no lo es (Peirano, 2023).

Hoy existe ya un volumen considerable de investigaciones académicas relevantes sobre internet y las redes sociales en relación al género, pero aún necesitamos generar mucho conocimiento sobre los sesgos en el diseño y los usos de la IA. Es preciso que nos preguntemos por quién la está definiendo, por las formas de apropiación machista de la herramienta y por la razón de que estas hayan aparecido tan rápido, pero también por las prácticas no machistas o transformadoras, entre otras cuestiones. Y el panorama analizado no debería ser únicamente el occidental, ya que el fenómeno es global y todos los saberes que podamos generar cuentan para combatir las formas nuevas (o redefinidas) de opresión sexista que genera la IA. La violencia de género facilitada por la tecnología (denominada TFGBV por la ONU por sus siglas en inglés)[1] se halla en evolución cons-

1. TFGBV: Technology-facilitated Gender-Based Violence, definida por UN Women: https://www.unwomen.org/en/digital-library/publications/2024/03/placing-gender-equality-at-the-heart-of-the-global-digital-compact

tante, lo que fuerza a mujeres y niñas de todo el mundo a autocensurarse, a abandonar las redes sociales o a reducir su interacción/participación en los espacios *online*.

La IA en teoría está pensada para solucionar grandes problemas del mundo (Peirano, 2023). Es evidente que tiene un enorme potencial también para las mujeres y las niñas, pues puede transformar el acceso al empleo, a los servicios públicos, a la educación y a la sanidad. Se está empezando a usar, por ejemplo, en ensayos clínicos para diagnosticar enfermedades y desarrollar estrategias terapéuticas específicas para el cuerpo femenino, que históricamente ha sido ignorado o infrarrepresentado en la ciencia. Sin embargo, estos procesos requieren de una participación igualitaria de las mujeres en el sector tecnológico —y de un papel clave de estas en su configuración desde la raíz— si queremos evitar que la IA perpetúe los prejuicios, la discriminación y la desigualdad existentes. Un estudio reciente de la ONU advierte que las mujeres representan solo el 30% de los profesionales tecnológicos a nivel mundial. Ese porcentaje se reduce al 22% en el campo de la IA (UN Women Expert Group, 2023). El grupo de expertas advierte que en la actual arquitectura de la IA, los beneficios y los riesgos no están distribuidos equitativamente, pues el poder se concentra en manos de unas pocas empresas que controlan el talento, los datos y los recursos informáticos. Hay que reconsiderar los métodos de selección de profesionales de la industria de la IA para que esta incorpore una mayor diversidad de puntos de vista. Si las mujeres no podemos dar forma a las tecnologías en pie de igualdad, entonces la IA se construirá sobre estereotipos y formas de violencia del pasado. Parafraseando a la científica británica Karen Spärck Jones, la IA es demasiado importante para dejarla en manos de los hombres.

Además, por el camino hemos perdido el control público sobre estas grandes empresas tecnológicas privadas que desarrollan y explotan comercialmente la IA. No son pocos los casos en que las *big tech* (o sus directivos) se muestran sospechosamente neutrales ante la evolución de la extrema derecha, el racismo o la violencia machista, y, en ocasiones, criminalizan o directamente eliminan las alternativas. Ejemplos como el caso de Almendralejo nos dejan perplejas, indignadas y sin capacidad de reacción (inmediata, al menos). El caso está ya en manos de la Justicia y, muy probablemente, los involucrados en el caso Almendralejo no quedarán impunes, pero el daño para las niñas afectadas ya está hecho.

3. ¿Hay alternativas?

Ni las autoridades públicas, ni la academia, ni la sociedad civil podemos quedarnos cruzados de brazos ante este problema, yendo siempre a la zaga de la realidad. Debemos plantar cara recurriendo a la soberanía tecnológica y a la acción proactiva. Una propuesta reciente a nivel internacional es la herramienta de supervisión "Gender Social Media Monitoring" del Programa de las Naciones Unidas para el Desarrollo, un

proyecto piloto que utiliza la IA para detectar contenido dañino y discursos de odio contra las mujeres y niñas en más de cien idiomas. Estos datos pueden, por ejemplo, capacitar a las autoridades públicas para actuar contra la violencia de género propiciada (o facilitada) por la IA. En España, el colectivo *Donestech*, que profundiza en estrategias prácticas y políticas de autodefensa digital feminista, nos ofrece herramientas concretas para luchar por una IA al servicio de las mujeres y las minorías. En el mismo sentido, nosotras proponemos apropiarnos de esta tecnología desde lo colectivo y usar la propia IA para detectar, rastrear y eliminar más fácilmente las imágenes nocivas creadas artificialmente.

Una alternativa al modelo actual de total desregulación, sería otro de mayor transparencia; deberíamos poder obligar por ley a la IA a informar o incluso a intervenir si alguien le solicita generar imágenes que sean vejatorias, inciten al odio o se puedan considerar pornografía infantil. Esta transparencia también implicaría que los algoritmos de IA estuvieran abiertos para la inspección de las autoridades civiles responsables de la implementación de políticas públicas. Esta rendición de cuentas podría proceder, por ejemplo, por medio de auditorías o informes periódicos que alerten a los usuarios y las usuarias del sesgo residual de la IA (sea este de género o de otro tipo) en las herramientas que utilizan. La nueva *Ley Europea de Inteligencia Artificial* directamente prohíbe el uso de la IA generativa en función del peligro que suponga para las personas e identifica sistemas de alto riesgo que solo se podrán poner en el mercado si se demuestra que respetan los derechos fundamentales. Visto que las empresas dedicadas a la comercialización de herramientas de IA ya se han mostrado preocupadas por esta ley comunitaria, argumentando que es "demasiado estricta", quizás estemos yendo en la buena dirección.

Estas medidas legales ayudan, por supuesto, a controlar el mal uso de una herramienta o la disponibilidad de productos potencialmente peligrosos para la salud pública. De hecho, igual que en el pasado se han regulado sectores como el farmacéutico, el alimentario o el tabaquero, en los últimos años se está empezando a legislar de manera contundente el ámbito de las *big tech* para atajar la pandemia de depresiones y suicidios entre la juventud causada por el abuso de las redes sociales. Pero la clave no solo está en regular, legislar o perseguir. Dado que la IA está aquí para quedarse y nos va a estar acompañando en la vida diaria, donde realmente se necesita un cambio es en la cultura: en la cultura de la violación, de la explotación de los cuerpos de las mujeres, de la misoginia y del antifeminismo. Hay que poder detectar rápidamente los intereses detrás de un comportamiento de una IA en un espacio o un momento dado, y señalar a los que se benefician de ella a costa de otras personas. Si no conseguimos este cambio cultural y no transformamos esa mirada que ve el cuerpo de las mujeres como un divertimento lúdico sin consecuencias, la IA seguirá produciendo las mismas formas de violencia que existen en la sociedad.

4. Referencias bibliográficas

Eco, Umberto (1995). *Apocalípticos e integrados*. Barcelona: Lumen.

Peirano, Marta (2023). "La doble vida de la inteligencia artificial". En CCCB, *IA: Inteligencia Artificial*. Barcelona: CCCB-Diputación de Barcelona, 54-61.

Skeggs, Beverly (2004). "Exchange, Value and Affect: Bourdieu and 'The Self'". *The Sociological Review*, 52(2), 75-95.

Un Women Expert Group (2023). "Innovation and technological change, and education in the digital age for achieving gender equality and the empowerment of all women and girls", *Expert guidance and substantive inputs to preparations for the 67th Session of the Commission on the Status of Women*. Disponible en: https://www.unwomen.org/sites/default/files/2023-02/230213%20BLS22613%20UNW%20CSW67.v04%20%282%29.pdf

La Inteligencia Artificial y sus implicaciones éticas, sociales y políticas

Leonor M. Cantera
Universitat Autònoma de Barcelona

1. Introducción

A lo largo de la historia, se han introducido cambios tecnológicos significativos, como el telégrafo, la electricidad, los antibióticos, el avión y el tren. Estos avances han mejorado la calidad de vida, pero también han afectado las interrelaciones personales, lo social y el medio ambiente. La inteligencia artificial (IA) no es la excepción. Esta tecnología emergente tiene el potencial de revolucionar múltiples aspectos de la vida humana, pero su implementación y desarrollo plantean interrogantes éticos, sociales y políticos que deben ser abordados con cuidado, especialmente en problemáticas tan sensibles como la violencia de género. ¿Cómo se aplicará? ¿Quién la supervisará? ¿Cómo asegurar la seguridad de las personas a nivel personal y social? ¿Qué precio estamos dispuestos/as a pagar para asegurar su utilización?

Este escrito explora estas cuestiones, especialmente en su relación con la violencia de género, una problemática social que muestra múltiples desigualdades, prejuicios, estereotipos y respuestas sesgadas.

2. Implicaciones y retos de la IA en relación con la violencia de género

La IA puede procesar y analizar grandes cantidades de datos para generar predicciones y decisiones. Sin embargo, lo que muestre dependerá de cómo se estructuren los sistemas

que la hacen posible. Es crucial diseñarla desde una perspectiva de género para reflejar una representación que, al tener en consideración aspectos históricos, relacionales y sociales, presente de manera más justa y precisa la realidad, evitando sesgos que perpetúen desigualdades y discriminación. En el contexto de la violencia de género, sistemas como VioGén[1] deben ser transparentes y explicables, asegurando que las decisiones tomadas sean justas y basadas en datos precisos (González Álvarez, López Ossorio y Muñoz Rivas, 2018).

Recientemente, Vargas (2024) señala diversas herramientas de IA para la lucha contra la violencia de género. Por ejemplo, el Instituto de Estudios de Género de la Universidad Carlos III de Madrid ha desarrollado el dispositivo llamado *Bindi*, que pretende evitar la violencia de género y sexual, siendo capaz de detectar emociones y situaciones de riesgo. Existe el *chatbot* AINO con el objetivo de proporcionar asesoramiento inmediato y detectar casos de violencia de género. Esta herramienta, que según la autora pretende ayudar al 80% de víctimas que no denuncian sus agresiones, es capaz de simular la conversación humana, evaluar el riesgo y orientar sobre recursos disponibles. En América Latina y el Caribe encontramos el *chatbot* SARA, que, según su presentación, es capaz de interactuar y desarrollar aprendizaje automático. Cabe el señalar que la IA no aprende, sino que se alimenta de la retroalimentación y, por tanto, utiliza algoritmos para ampliar la información que ofrece.

A pesar de la presentación de herramientas con carácter ventajoso como las previamente señaladas, la directora ejecutiva adjunta de ONU Mujeres, Sarah Hendriks, ha destacado la necesidad de regular la IA para prevenir actos de violencia contra las mujeres, subrayando que ella ha sido utilizada para crear pornografía falsa y rastrear y acosar a mujeres (Triay, 2023). Es crucial que la IA no se convierta en una herramienta que domine otras tecnologías o sistemas sin una supervisión adecuada. Debe promover la equidad y la justicia, no reforzar las desigualdades existentes (Presno Linera, 2023). Para evitar estos posibles sesgos es determinante, de manera constante, revisar y auditar los algoritmos utilizados en la creación de IA, asegurando que no perpetúen estereotipos de género y discriminación.

La IA puede estar enmarcada bajo diversas lógicas políticas, económicas y sociales. En un contexto capitalista neoliberal, podría ser utilizada para maximizar la eficiencia y la rentabilidad, a menudo a expensas de la equidad y la justicia social, lo que podría llevar a una mayor concentración de poder y recursos en manos de unas pocas personas, exacerbando las desigualdades existentes (Arribas Macho, 2018). Por otro lado, en un marco de economía social y solidaria, la IA podría promover el bienestar colectivo, me-

1. En julio de 2007 se puso en marcha el *Sistema policial de seguimiento integral en los casos de violencia de género* (Sistema VioGén) centralizado en el Ministerio del Interior para las evaluaciones de riesgo de la mujer víctima de violencia de género.

jorar los servicios públicos y reducir las desigualdades. La clave está en cómo se diseñan y regulan estos sistemas, y en asegurar que las decisiones sobre su uso se tomen de manera democrática y participativa (Cotino y otros, 2021).

Las ciencias sociales y humanistas juegan un papel crucial en el desarrollo, implementación y vigilancia de la IA, asegurando que sea justa y equitativa. Estas disciplinas proporcionan una comprensión más profunda de las implicaciones sociales y éticas de la IA, ayudando a diseñar sistemas que respeten los derechos humanos y promuevan la justicia social (Roa Avella, Sanabria-Moyano y Peña Piñeros, 2023). Por ejemplo, asegurando que no perpetúen estereotipos de género y discriminación. Como señala Alonso Betanzos (2023), es fundamental vigilar que la IA se desarrolle de manera ética y responsable.

En la actualidad, la IA es presentada como una tecnología que ofrece ventajas significativas en la lucha contra la violencia de género, con la capacidad de analizar grandes volúmenes de datos para identificar patrones de comportamiento y predecir situaciones de riesgo. Además, herramientas como *chatbots* pueden proporcionar apoyo inmediato y recursos a las víctimas, incluso en situaciones donde el acceso a ayuda humana es limitado. Sin embargo, estos programas también presentan retos significativos. A nivel social, es relevante reconocer que no todas las personas tienen acceso a estos servicios debido a limitaciones económicas o tecnológicas, lo que subraya la necesidad de políticas inclusivas que garanticen el uso equitativo de estas herramientas innovadoras. A nivel técnico, la precisión de los algoritmos depende de la calidad y diversidad de los datos utilizados, lo que puede llevar a sesgos si no se manejan adecuadamente. Además, la implementación de estas tecnologías debe estar acompañada de marcos regulatorios y éticos robustos para asegurar que se utilicen de manera justa y equitativa, sin perpetuar las desigualdades existentes (O'Neil, 2016; Noble, 2018).

3. Conclusión

La inteligencia artificial tiene el potencial de transformar profundamente nuestras sociedades, pero su desarrollo y uso plantean una serie de interrogantes éticos, sociales y políticos que deben ser abordados con cuidado. Es fundamental que se diseñe con perspectiva de género y se implemente de manera transparente, equitativa y responsable, asegurando que promueva la justicia social y respete los derechos humanos; garantizando que la IA se utilice para el beneficio de toda la humanidad y no como una herramienta que perpetúe las desigualdades (Brynjolfsson & McAfee, 2014; Crawford & Calo, 2016).

4. Referencias bibliográficas

Alonso Betanzos, Amparo (2023). "Artificial intelligence and gender bias". *Revista Gender on Digital*, 1, 11-32. Disponible en: https://revistas.uvigo.es/index.php/GOD/article/view/5060

Arribas Macho, José María (2018). "Cathy O'Neil: Armas de destrucción matemática. Como el Big Data aumenta la desigualdad y amenaza la democracia. Madrid: Capitán Swing, 2017, 269 pp.". *EMPIRIA. Revista de Metodología de Ciencias Sociales*, 41, 199-202. Disponible en: https://www.redalyc.org/journal/2971/297165396011/html/

Brynjolfsson, Erik y Mcafee, Andrew (2014). *The second machine age*. *Work, progress, and prosperity in a time of brilliant technologies*. New York: WW Norton & Company.

Cotino Hueso, Lorenzo; Castillo, José Antonio; Salazar, Idoia; Benjamins, Richard; Cumbreras, María y Esteban, Adaya María (2021). "Un análisis crítico constructivo de la propuesta de reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial Artificial Intelligence Act.". *Diario La Ley* (2 de julio de 2021). Disponible en: https://investigacionusp.ceu.es/es/ipublic/item/9770411

Crawford, Kate y Calo, Ryan (2016). "There is a blind spot in IA research". *Nature*, 538, 311-313. https://doi.org/10.1038/538311a

González Álvarez, José Luis; López Ossorio, Juan José y Muñoz Rivas, Marina (2018). *La valoración policial del riesgo de violencia contra la mujer pareja en España. Sistema de seguimiento integral en los casos de violencia de género. Vio Gén.* Madrid: Ministerio del Interior. Gobierno de España. Disponible en: https://bit.ly/3B5lS8y

Noble, Safiya Umoja (2018). *Algorithms of Oppression: How search engines reinforce racism*. New York: New York University Press.

O'Neil, Cathy (2016). *Weapons of math destruction: How big Data increases inequality and threatens democracy*. New York: Crown Publishing.

Presno Linera, Miguel Ángel (2023). "Inteligencia Artificial, policía predictiva y prevención de la violencia de género". *Revista de Vitimologia e Justiça Restaurativa,* 1(2), 85-118. https://doi.org/10.58725/rivjr.v1i2.39

Roa Avella, Marcela del Pilar; Sanabria-Moyano, Jesús Eduardo y Peña Piñeros, Andrea Catalina (2023). "Los estándares internacionales de protección de la violencia basada en género de las mujeres aplicados a la inteligencia artificial predictiva. Herramientas de predicción de violencia basada en género y feminicidio mediante la Inteligencia Artificial". *Justicia,* 28(43), 43-56. https://doi.org/10.17081/just.28.43.6161

Triay, Emilia (2023). "ONU Mujeres pide regular la inteligencia artificial con el objetivo de prevenir la violencia de género". *Business and Economics* (24 de noviembre de 2023). Disponible en: https://es.scribd.com/document/718796086/ONU-mujeres-regular-IA-VdG

Vargas, Laura (2024). "Inteligencia artificial y violencia de género: transformando la prevención y detección". *AlasTensas* (21 de junio de 2024). En línea: https://bit.ly/4euTmM2

Inteligencia artificial (IA) y sesgos

Rosa Maria Gil Iranzo
Departamento de Informática y Diseño Digital
Universitat de Lleida

1. Aplicaciones de la IA y sesgos

Hacer un análisis exhaustivo y definitivo de las aplicaciones de IA conocidas en relación con los estudios de sesgo es una tarea muy compleja por la rapidez con la que evoluciona el campo y la gran cantidad de variables involucradas. Sin embargo, podemos analizar algunos casos de estudio y tendencias generales. Primero, vamos a describir algunos ámbitos donde la IA y los posibles sesgos asociados tienen mayor impacto.

En el ámbito del reconocimiento facial, encontramos que los sesgos raciales y de género son numerosos pues los sistemas de reconocimiento facial tienen dificultades para identificar correctamente a personas de color y a mujeres, lo que implica un problema en el ámbito de seguridad, control de acceso o en aplicaciones de identificación. Los *chatbots* y asistentes virtuales no se escapan a los sesgos de género y estereotipos, pues los *chatbots* a menudo los refuerzan y pueden ser discriminatorios hacia ciertos grupos. Baste recordar el caso de Tay, un *chatbot* basado en IA que se puso en funcionamiento el 23 de marzo en Twitter (la red social que conocemos actualmente como 'X') y, en el transcurso de un día, se había convertido en un *hater*: devolvía comentarios racistas y xenófobos (Vincent, 2016). Las aplicaciones donde los *chatbots* basados en IA se utilizan son numerosas destacando, sobre todo, los servicios telemáticos para clientes, y en el ámbito de la salud y la educación.

Conseguir un puesto de trabajo también es un campo donde más se está utilizando la IA mediante la incorporación de algoritmos de contratación. El reclutamiento, o tam-

bién la llamada selección personal, no escapa a los sesgos de género y raza. Existen muchos casos de empresas donde ha habido una discriminación. Un ejemplo es el de una empresa de cuidado infantil, donde el algoritmo nunca escogía a hombres para poder realizar tal función, como se puede apreciar en Cheng *et al.* (2023).

Cualquier aplicación multimedia de ocio/redes sociales que tenga un sistema de recomendación tampoco escapa al sesgo. Los llamados sesgos de *filtro burbuja* pueden reforzar creencias existentes y limitar la exposición a nuevas ideas. Así pues, las plataformas de redes sociales o los servicios de *streaming* no escapan a ello (Kaluža, 2023).

2. Tendencias generales y desafíos

Respecto a las tendencias generales y a los posibles desafíos, nos encontramos que la falta de transparencia es un reto que se debe afrontar ya en el presente, pues muchas empresas son reticentes a compartir los datos y algoritmos utilizados en sus aplicaciones de IA, lo que dificulta el análisis de sesgos.

Idealmente, la diversidad en los equipos de desarrollo es un hándicap, ya que la falta de esta en los equipos de desarrollo puede conducir a la creación de productos que reflejan los sesgos de sus creadores.

Una de las peores pesadillas son los datos de entrenamiento, que a menudo ya están sesgados de por sí, dado que los datos utilizados para entrenar los modelos de IA con frecuencia contienen sesgos inherentes, que se reflejan en los resultados. No siempre existen las herramientas y métricas adecuadas para evaluar el rendimiento de los modelos de IA, puesto que pueden no ser suficientes para detectar todos los tipos de sesgo.

El panorama deseable para abordar estos retos pasa por que las empresas: 1) sean más transparentes en cuanto a los datos y algoritmos que utilizan; 2) aumenten la diversidad en los equipos de desarrollo; 3) incorporen auditorías independientes de los sistemas de IA para identificar y mitigar los sesgos, implementando regulaciones y estándares éticos para el desarrollo de la IA, y 4) eduquen a los usuarios y usuarias sobre los riesgos del sesgo en la IA y cómo identificarlos, algo importante en el proceso.

3. Tipos de aplicaciones

Algunas de las aplicaciones gratuitas —nos referimos a aquellas que no exigen un pago directo monetario— que han implementado estrategias para mitigar el sesgo en sus modelos de IA vienen de la mano de Google e IBM, principalmente. Los primeros promueven prácticas responsables de IA, incluyendo la evaluación continua de modelos para detectar y corregir sesgos (Wang *et al.*, 2020). Pese a estos esfuerzos Google ha enfrentado críticas por sesgos en sus algoritmos de reconocimiento facial y de lenguaje,

como comenta Robinson (2020). En el caso de IBM, tenemos concretamente que esta empresa ha desarrollado herramientas como el AI *Fairness 360 Toolkit*, que ofrece algoritmos para detectar y mitigar sesgos en los datos y modelos de IA según Bellamy (2020). IBM también promueve la transparencia y la responsabilidad en el desarrollo de IA, con iniciativas como los *AI FactSheets y Watson OpenScale*, tal y como explica Dominique (2023). Aunque IBM ha avanzado significativamente en la mitigación de sesgos, los modelos aún pueden reflejar sesgos inherentes en los datos de entrenamiento, como se puede ver en Dolata (2023).

En lo que se refiere a las aplicaciones de pago, *Microsoft Azure AI* utiliza el paquete de código abierto *Fairlearn* para atenuar problemas de equidad en los modelos de *Machine Learning (ML)*, como describe Luzt (2023). Además, Microsoft ha desarrollado un conjunto de herramientas dentro del *Responsible AI Toolbox* para identificar, diagnosticar y mitigar problemas de sesgo antes de desplegar los sistemas de ML (Kiraly, 2021). El gran desafío, en este caso, es el costo de estas herramientas ya que pueden ser una barrera para pequeñas empresas o individuos. Por su parte, Amazon Web Services (AWS) AI ofrece *SageMaker Clarify*, una herramienta que detecta sesgos en los datos y modelos de ML, y explica las predicciones de los modelos (Hardt, 2021). AWS también proporciona cursos gratuitos sobre IA responsable a través de su *Machine Learning University*. Por contra, no pueden asegurar que no existan sesgos en los datos utilizados para entrenar sus modelos.

Respecto a las aplicaciones públicas, encontramos que *Hugging Face* ha desarrollado herramientas y métricas para evaluar y mitigar sesgos en modelos de lenguaje, como se muestra en su biblioteca *Evaluate* (Kathicar, 2023). También han investigado cómo las arquitecturas de redes neuronales pueden influir en los sesgos y han propuesto soluciones para mejorar la equidad en el reconocimiento facial. De nuevo, debemos ser conscientes que la calidad y el sesgo de los modelos pueden variar dependiendo de los contribuyentes y la diversidad de los datos de entrenamiento.

Respecto a las aplicaciones privadas, encontramos que la mundialmente conocida *OpenAI* ha implementado técnicas para reducir el sesgo en sus modelos, como en DALL·E 2, donde se han aplicado métodos para generar imágenes que reflejen mejor la diversidad de la población mundial (ver Choudhry, 2023). OpenAI también ha publicado documentos sobre los riesgos y limitaciones de sus modelos para abordar los sesgos inherentes, como señala Alfano (2024). Sin embargo, a pesar de los esfuerzos, los modelos de *OpenAI* aún pueden reflejar sesgos presentes en los datos de entrenamiento, y la mitigación de estos sesgos sigue siendo un desafío continuo.

Así pues, vemos que en todos los campos hay unas preocupaciones comunes. Las herramientas que se pueden utilizar de IA en el campo de la violencia de género son muy diversas, y van desde la preocupación por la perpetuación de los estereotipos, o la banalización de la violencia, hasta el tratamiento de las bases de datos de casos de violencia de género, pasando por la atención automática por parte de la IA a las víctimas.

La semántica de la información en las aplicaciones de IA todavía no está muy presente, tal y como la entendemos los humanos. Actualmente se generan imágenes escribiendo los conocidos *prompts*, pero las palabras y su semántica no son iguales para los humanos que para las máquinas, en ningún sentido todavía. Por eso, cuando los algoritmos tratan de ser más inclusivos, comenten el error, por ejemplo, de poner soldados de raza negra como soldados de Alemania en el período nazi.

Las IA que generan imágenes también las pueden describir con palabras, así que me gustaría acabar con una reflexión sobre estos metadatos asociados a las imágenes. A tenor de que muchas veces no hay suficientes evidencias para asociar facciones físicas a comportamientos anómalos, como los de los agresores, ¿deben existir bases de datos que relacionen las imágenes de los agresores con ciertas etiquetas (o metadatos)?, y, si es así, ¿debe ser para toda la vida del agresor o solo mientras tenga una condena?, ¿se deberían etiquetar los individuos que han sido reiteradamente denunciados?, ¿cualquier ciudadano/a debería poder informarse sobre estas listas?

Deberíamos encontrar, cada vez más, espacios de discusión para afrontar los retos que la tecnología nos presenta ahora y seguirá haciéndolo en el futuro.

4. Referencias bibliográficas

Alfano, Mark; Abedin, Ehsan; Reimann, Ritsaart; Ferreira, Marinus y Cheong, Marc (2024). "Now you see me, now you don't: an exploration of religious exnomination". *DALL-E. Ethics and Information Technology,* 26(2). https://doi.org/10.1007/s10676-024-09760-y

Bellamy, Rachel K.E.; Dey, Kuntal; Hind, Michael; Hoffman, Samuel C.; Houde, Stephanie; Kannan, Kalapriya; Lohia, Pranay; Martino, Jacquelyn; Mehta, Sameep; Mojsilovic, Aleksandra; Nagar, Seema; Ramamurthy, Karthikeyan Natesan; Richards, John; Saha, Diptikalyan; Sattigeri, Prasanna; Singh, Moninder; Varshney, Kush R. y Zhang, Yunfeng (2018). "AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias", *ArXiv*, 1810.01943. En línea: https://arxiv.org/abs/1810.01943

Cheng, Myra; De-Arteaga, María; Mackey, Lester; Kalai, Adam Tauman, *et al.* (2023). "Social norm bias: residual harms of fairness-aware algorithms", *Data Min Knowl Disc*, 37, 1858–1884. https://doi.org/10.1007/s10618-022-00910-8

Dolata, Mateusz y Crowston, Kevin (2023). "Making sense of AI systems development", *IEEE Transactions on Software Engineering*, 50(1), 123-140. https://doi.org/10.1109/tse.2023.3338857

Dominique, Brandon; Maghraoui, Kaoutar El; Piorkowski, David y Herger, Lorraine (2023). "FactSheets for Hardware-Aware AI Models: A Case Study of Analog In Me-

mory Computing AI Models", *2023. IEEE International Conference on Software Services Engineering* (SSE), 148-158. https://doi.org/10.1109/SSE60056.2023.00029

Hardt, Michaela *et al.* (2021). "Amazon SageMaker Clarify: Machine Learning Bias Detection and Explainability in the Cloud". En Feida Zhu y Beng Chin Ooi (eds.), *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 2974-2983. https://doi.org/10.1145/3447548.3467177

Kaluža, Jernej (2023). "Far-reaching effects of the filter bubble, the most notorious metaphor in media studies", *AI & Society*, 38, 1391-1393. https://doi.org/10.1007/s00146-022-01399-x

Kathikar, Anirudh; Nair, Anupam; Lazarine, Brandon; Sachdeva, Ankit y Samtani, Shamik (2023). "Assessing the Vulnerabilities of the Open-Source Artificial Intelligence (AI) Landscape: A Large-Scale Analysis of the Hugging Face Platform", *2023. IEEE International Conference on Intelligence and Security Informatics (ISI),* 1-6. https://doi.org/10.1109/ISI58743.2023.10297271

Király, Franz; Löning, Markus; Blaom, Amos; Guecioueur, Alexis y Sonabend, Robin (2021). "Designing Machine Learning Toolboxes: Concepts, Principles and Patterns", *ArXiv,* 2101.04938. En línea: https://arxiv.org/abs/2101.04938

Lutz, Richard (2023). "Fairlearn: Assessing and Improving Fairness of AI Systems", *J. Mach. Learn. Res.,* 24, 257:1-257:8. https://doi.org/10.48550/arXiv.2303.16626

Robinson, Joshua; Livitz, Gregory; Henon, Yann; Qin, Cheng; Fu, Yuchen y Timoner, Samuel (2020). "Face Recognition: Too Bias, or Not Too Bias?", *2020. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 1-10. https://doi.org/10.1109/CVPRW50498.2020.00008

Vincent, James (2016). "Twitter Taught Microsoft's AI Chatbot to Be a Racist in Less than a Day", *The Verge*, 24/03/2016. En línea: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Wang, Yiling; Xiong, Meng y Olya, Hamed (2020). "Toward an Understanding of Responsible Artificial Intelligence Practices", *Proceedings of the 53rd Hawaii International Conference on System Sciences* (HICSS), 4962-4971. https://doi.org/10.24251/hicss.2020.610

Entornos tecnológicos y violencias machistas: la IA en la era digital

Cynthia Gálvez López
Mulleres Tech

La transformación digital ha modificado profundamente nuestras interacciones cotidianas, abriendo nuevas oportunidades, pero también generando nuevos riesgos. Entre ellos, las violencias machistas han encontrado en los entornos digitales un nuevo espacio de actuación, donde las redes sociales, los videojuegos y otras plataformas *online* se han convertido en escenarios de acoso y control, especialmente hacia las mujeres jóvenes. A medida que estas tecnologías se han democratizado, llegan a audiencias más jóvenes y vulnerables, blancos fáciles de ataques, como los y las adolescentes, que se enfrentan a estos riesgos sin las herramientas adecuadas para protegerse y que pueden llegar a impactar en su autoestima y seguridad personal. En este contexto, la Inteligencia Artificial (IA), uno de los avances más significativos de nuestro tiempo, puede ser tanto una aliada como una amenaza en la lucha contra la violencia de género.

1. Violencias machistas en los entornos digitales

Ciberacoso, sextorsión, *grooming*, *doxing*, *flaming*, troleo misógino, pornografía de venganza, suplantación de identidad y *stalking* digital. Son solo algunos de los términos que definen formas de violencia digital, muchas de ellas con un enfoque particular en las agresiones de género. El anonimato que facilitan los entornos digitales permite que agresores utilicen Internet para ejercer control, difamar o acosar a sus víctimas, muchas veces niñas y adolescentes.

Hay múltiples ejemplos conocidos, como es el caso de Zoë Quinn en 2014, desarrolladora de videojuegos que sufrió acoso masivo, amenazas de muerte y ataques de *doxing* (revelación de información privada) tras ser criticada públicamente en el incidente conocido como "Gamergate". Otro caso ha sido la crítica de la comunidad *gamer* hacia las jugadoras que usaron la voz femenina del personaje principal tras el lanzamiento de *Assassin's Creed: Valhalla* en 2020. O el primer caso reportado de acoso sexual en el metaverso ocurrido en Horizon Worlds (la plataforma de realidad virtual de Meta) en 2021 donde una usuaria denunció haber sido víctima de una agresión sexual virtual mientras exploraba el entorno inmersivo. Son todo ejemplos que evidencian la hostilidad en la comunidad hacia mujeres subrayando el sesgo y menosprecio machista que persiste en algunos sectores del *gaming*.

2. La Inteligencia Artifical y sus sesgos de género

En la realidad actual, la IA juega un papel clave, no solo como herramienta para combatir estas violencias, sino también como elemento que, si no es bien gestionado, puede amplificarlas. Casos recientes muestran que una IA mal implementada puede reproducir sesgos de género preexistentes y perpetuar estereotipos machistas.

Uno de los primeros casos que demostró los peligros de los sesgos fue en 2016, cuando Microsoft lanzó el *chatbot Tay* en Twitter. Este fue desactivado en menos de 24 horas debido a que, tras ser manipulado por *trolls*, empezó a generar mensajes racistas y sexistas. Un ejemplo similar fue la herramienta de reclutamiento de Amazon en 2018, cuyo algoritmo mostró preferencia por candidatos hombres debido a que se entrenó con datos históricos sesgados.

Los algoritmos de reconocimiento facial, como los de IBM y Amazon en 2019, también revelaron fallas importantes, mostrando mayor error en la identificación de mujeres, especialmente mujeres negras. Los asistentes virtuales como Siri y Alexa han sido criticados por usar voces femeninas sumisas, lo que refuerza estereotipos de género; sin embargo, ambos sistemas fueron modificados para responder de manera más asertiva tras las críticas.

Incluso LinkedIn mostró sesgos en 2021, cuando su función de autocompletado sugería pronombres masculinos para puestos como "ingeniero" o "jefe", y femeninos para roles subordinados como "asistente". GPT-3, el modelo de lenguaje de OpenAI, también ha demostrado sesgos al asociar profesiones y razas de manera inapropiada. Estos casos evidencian cómo la IA, al aprender de datos sesgados, perpetúa desigualdades preexistentes, resaltando la necesidad de un enfoque inclusivo y consciente en su desarrollo.

3. Oportunidades basadas en IA

Las innovaciones tecnológicas no pueden ser ignoradas, detenidas o prohibidas, ya que son una parte esencial del progreso social. En lugar de oponernos a ellas, es fun-

damental comprender su funcionamiento y adaptarlas al contexto actual, garantizando que se alineen con los principios de equidad, seguridad y justicia. Solo de este modo podremos aprovechar su potencial para mejorar la vida de las personas. En la lucha contra la violencia de género, la IA representa tanto un desafío como una oportunidad. Aunque los avances tecnológicos han facilitado nuevas formas de acoso y control, la IA ofrece herramientas innovadoras para prevenir, detectar y combatir estas violencias. Algunos ejemplos incluyen:

1. Prevención y detección de la violencia de género

- Detección temprana de acoso: IA que monitorea plataformas *online* para identificar patrones de abuso o amenazas en tiempo real, permitiendo alertas y respuestas rápidas.
- Análisis predictivo: Identificación de patrones mediante grandes volúmenes de datos que sugieran la probabilidad de que alguien sea víctima de violencia.
- Análisis de sesgos: La IA puede evaluarse para detectar y corregir sesgos de género en sus propios modelos y algoritmos, previniendo discriminaciones.

2. Asistencia a las víctimas

- *Chatbots* de asistencia: Herramientas impulsadas por IA que proporcionan apoyo psicológico, legal o informativo a víctimas de violencia de género en tiempo real.
- Simulación y formación: Sistemas que entrenan a profesionales o a víctimas para la detección y manejo de casos de violencia de género.

3. Protección y seguridad digital

- Sistemas de seguridad digital: La IA puede proteger los datos y cuentas personales de las mujeres, evitando que los agresores accedan o ejerzan control sobre su vida digital.
- Bloqueo automático de acosadores: Identificación y bloqueo inmediato de usuarios involucrados en acoso o *ciberstalking*.

4. Intervención y eliminación de contenido nocivo

- Eliminación de contenido ofensivo: IA que detecta y elimina automáticamente imágenes no consensuadas, como casos de pornografía de venganza, o comentarios abusivos.
- Prevención de la difusión de contenido: IA que impide la difusión de contenido explícito o imágenes compartidas sin consentimiento, protegiendo la privacidad de las víctimas.

5. Investigación y recolección de pruebas

- Análisis forense digital: Recolección, análisis y procesamiento de pruebas digitales en casos de violencia, lo que facilita la creación de perfiles de agresores basados en patrones previos.

Estas soluciones muestran el potencial de la IA para abordar la violencia de género en entornos digitales, siempre que su desarrollo se realice de manera ética e inclusiva, evitando perpetuar los sesgos presentes en la sociedad.

4. Un marco de trabajo para un futuro seguro

Para que las soluciones basadas en IA sean efectivas, deben apoyarse en tres pilares: transparencia, colaboración y privacidad. Es crucial que las tecnologías de IA se diseñen desde una perspectiva inclusiva y feminista, involucrando a expertos en género y utilizando datos diversos que representen adecuadamente a todas las mujeres, especialmente a las comunidades más marginadas.

En cuanto a la privacidad, se debe proteger a las víctimas y asegurar que los datos no sean mal utilizados, estableciendo regulaciones y mecanismos de supervisión para evitar la perpetuación de violencias machistas. Esto requiere la colaboración entre gobiernos, empresas tecnológicas y organizaciones para garantizar que las innovaciones respeten los derechos humanos y promuevan la igualdad de género.

Estamos en un momento crucial de implementación de la IA, donde es indispensable un proceso de co-creación inclusivo. Solo así se podrá garantizar que estas tecnologías reflejen las necesidades de toda la sociedad y no perpetúen desigualdades. La IA tiene el potencial de transformar los entornos digitales en espacios más seguros para las mujeres. Pero para lograrlo, es crucial poner la igualdad de género en el centro del desarrollo tecnológico, creando entornos digitales más justos y libres de discriminación.

5. Referencias bibliográficas

Basu, Tanya (2021). "The metaverse has a groping problem already", *MIT Technology Review*. En línea: https://www.technologyreview.com/2021/12/16/1042516/the-metaverse-has-a-groping-problem/

Dastin, Jeffrey (2018). "Amazon abandona un proyecto de IA para la contratación por su sesgo sexista", *Reuters*, 14/10/2018. Disponible en: https://www.reuters.com/article/world/amazon-abandona-un-proyecto-de-ia-para-la-contratacin-por-su-sesgo-sexista-idUSKCN1MO0M4/

Li, Lucy y Bamman, David (2021). "Gender and Representation Bias in GPT-3 Generated Stories". En Nader Akoury et al. (eds.), *Proceedings of the Third Workshop on Narrative Understanding*, Association for Computational Linguistics, 48-55. Disponible en: https://aclanthology.org/2021.nuse-1.5

Navarro, Beatriz (2020). "IBM y Amazon abjuran de la tecnología de reconocimiento facial por su sesgo racista", *La Vanguardia*, 11/06/2020. Disponible en: https://www.lavanguardia.com/internacional/20200611/481710398480/ibm-reconocimiento-facial-racismo-tecnologia-negros.html

Navarro, Víctor (2014). "Feminismo, medios y #GamerGate: por qué está en guerra el mundo de los videojuegos", *El País*, 2/11/2014. Disponible en: https://verne.elpais.com/verne/2014/11/02/articulo/1414911892_000081.html

Reese, Hope (2016). "Why Microsoft's 'Tay' AI bot went wrong", *TechRepublic*, 24/03/2016. En línea: https://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/

Impacte de la intel·ligència artificial en la violència masclista i l'agenda del feminista

Lourdes Muñoz Santamaría
Presidenta de Dones en Xarxa

L'avenç de les tecnologies digitals està provocant profundes transformacions en la societat en les últimes dècades. En una primera etapa, la introducció de la computació i de connexions a través de les xarxes a nivell mundial va produir una evolució cap a l'anomenada "societat digital" que ha implicat grans canvis socials; actualment, la irrupció de les tecnologies de dades, i en concret, de la intel·ligència artificial (IA), implicarà una nova revolució per l'automatització, que impactarà de manera rellevant en persones, organitzacions i espai públic.

En aquest context, el feminisme de dades proposa establir una agenda per evitar els riscos —com els biaixos que ja estan afectant les dones— així com aprofitar les oportunitats que ofereixen aquestes tecnologies per a abordar els reptes de l'igualtat de les dones. Aquestes tecnologies d'IA estan impactant en les oportunitats de les dones i, en concret, en les violències masclistes. Per tant, aquesta mirada s'ha de contemplar en les agendes públiques de digitalització i regulació dels drets digitals.

1. Reptes de la intel·ligència artificial respecte a la violència masclista

Gran part del processos de digitalització, consisteixen en automatitzar els procediments de presa de decisions de les organitzacions. Aquesta digitalització esta basada en processaments de dades i algorismes predictius d'IA; per aquesta raó, disposar de dades que reflecteixen la realitat i les necessitats de les dones és un requisit imprescindible per

avançar en la igualtat i en els drets de les dones. L'alternativa és un retrocés, que la pròpia intel·ligència artificial és capaç de prevenir.

Si fem l'exercici de preguntar a *chatgpt* sobre les conseqüències de bases de dades sense aquest perspectiva, n'indica 5 per a les dones:

1. Biaixos: els algorismes perpetuen i amplifiquin els biaixos existents.
2. Desigualtat laboral: es limita l'accés de les dones a llocs de treball d'alt valor o economia digital.
3. Impacte en la salut: s'obtenen diagnòstics erronis o tractaments menys efectius.
4. Seguretat en xarxes: podria no identificar o respondre de manera efectiva si no preveuen que les dones pateixen més assetjament i violència digital.
5. Representació en el disseny de l'IA: no s'ofereixen solucions que beneficiïn les dones

El principal repte doncs, consisteix en disposar de bases de dades públiques sobre les situacions específiques de les dones i de les qüestions que conformen l'agenda feminista, ja que, si aquestes dades no estan incorporades en els procediments automatitzats, no és només que no puguem recórrer a les tecnologies data per a analitzar i abordar reptes d'igualtat, és que s'estan generant noves bretxes i discriminacions a conseqüència del desenvolupament d'algorismes esbiaixats.

En aquest sentit, una de les mancances i dels reptes de la recopilació de dades obertes públiques és que les Administracions públiques han de publicar en formats oberts o *open data* amb criteris de drets de les dones; però ens trobem que publiquen menys dades sobre aquesta matèria que altres àmbits com mobilitat o urbanisme. Cal abordar la desagregació de dades per sexe en gran part de les dades obertes publicades per a les administracions públiques, i la publicació de bases de dades relacionades amb la igualtat efectiva de les dones i la violència masclista.

Si les administracions públiques no recopilen dades relatives a les dones que es publiquen en formats oberts, és impossible que aquests siguin utilitzats en les eines d'IA, i queden excloses dels resultats que generen aquests aplicatius. La publicació de dades obertes de les administracions ha d'aportar fiabilitat i enfocament social a l'ecosistema de dades.

Més enllà dels riscos que implica l'omissió de dades relatives a les dones en les bases de dades que alimenten les IA, cal aplicar la perspectiva de gènere en tot el procés de la creació de projectes d'IA. En la pròpia construcció de l'algoritme, l'entrenament i l'enfoc, cal que la presa de decisions no reprodueixi biaixos sexistes i tingui present, en el punt de partida, les diferències entre homes i dones, i les discriminacions que les dones pateixen a la societat.

Un exemple il·lustratiu de mala praxi és el d'un programa de reclutament basat en tecnologies d'IA que va desenvolupar la companyia Amazon, i que va acabar contractant únicament homes per a càrrecs intermedis i de direcció. Això es va deure al fet que l'algoritme s'havia entrenat amb dades de les seleccions de personal que s'havien realitzat

anteriorment, i, com la majoria de vegades les persones triades havien estat homes, va interpretar que ser dona era una característica no adequada per al lloc de treball a cobrir. Si fem servir bases de dades que reflecteixen la discriminació social sense aplicar correccions, els algoritmes repetiran aquesta discriminació com a criteri.

Per a entendre els efectes que pot arribar a tenir la utilització de bases de dades esbiaixades, tenim el cas d'alguns sistemes de reconeixement facial, els quals tenen més dificultats per a reconèixer rostres de dones, i especialment, rostres de dones negres (Ortiz, 2023). És a dir, aquestes IA van ser creades a partir de bases de dades en les quals hi havia moltes més dades relatives a homes que a dones, de manera que van desenvolupar una capacitat major per a reconèixer rostres d'homes.

Tenint en compte que aquesta diferència en la qualitat dels resultats entre homes i dones pot anar des d'aplicacions de reconeixement facial fins a sistemes de reclutament laboral o diagnòstic mèdic, és evident la urgent necessitat d'introduir la perspectiva feminista en el desenvolupament i funcionament de les tecnologies basades en dades, com és el cas de les IA.

D'altra banda, és important augmentar la presència de dones en els equips desenvolupadors de projectes d'IA per a disminuir la segregació horitzontal; en qualsevol cas, cal garantir que els algorismes no discriminin les dones encara quan aquestes siguin una minoria en els equips de programació; d'una altra manera, acceptaríem que el conjunt de les dones siguin discriminades pels projectes digitals pel fet que abans han estat discriminades en l'accés, i obviaríem la responsabilitat ètica dels equips de desenvolupament amb el conjunt de la societat.

1.1. Mals usos de les tecnologies data

A banda de la pròpia construcció de la tecnologia, existeixen els mals usos que se'n poden fer com, per exemple, utilitzar-les per a assetjar i violentar les dones. En aquest cas es converteixen en nous vehicles o formes de manifestació de la violència masclista.

1.2. Violència digital: violència en xarxes socials

Les dones pateixen una sèrie de formes de violència masclista específiques de l'àmbit digital, i en concret de les xarxes socials, com és l'assetjament en línia i la difusió de continguts íntims sense consentiment. Cal sempre subratllar que la violència masclista contra les dones és prèvia, ara el masclisme troba noves eines i estratègies per a exercir violència contra les dones.

Existeixen diverses formes de violència masclista en l'àmbit digital: el control i la vigilància sobre les activitats en línia de les dones, la suplantació de la seva identitat amb

la finalitat de desprestigiar-les, el ciberassetjament sexual i la violència simbòlica que implica la difusió d'estereotips de gènere o el llenguatge sexista.

A més, existeixen diverses formes de manifestació d'aquesta violència masclista en l'àmbit digital. En primer lloc, les possibilitats de difusió i coordinació que ofereixen les eines digitals, són utilitzades per organitzar atacs col·lectius en xarxes amb un gran impacte en les dones a les quals van dirigits. Això és especialment rellevant tenint en compte que el 58% de les dones líders sofreixen aquest tipus d'atacs en les seves xarxes socials (Morena-Balaguer *et al.*, 2022). Aquests fenòmens volen silenciar les veus de les dones a les xarxes, i les dones estan patint les conseqüències, com són l'eliminació dels seus perfils, autocensura o baix perfil en xarxes; a més d'una barrera que frena les dones que vulguin tenir veu en l'esfera pública en xarxes.

Les xarxes socials tenen una relació directa amb la IA, les xarxes funcionen amb algorismes d'IA, aquestes potencien la confrontació i això afecta de forma més negativa les dones, donat que les dones no parteixen del mateix punt en l'accés a eines o en rebre assetjament o consideració social. Aquestes xarxes representen gran part de l'espai de relació social i ciutadana.

Les possibilitats que ofereixen aquestes eines i la seva capacitat de difusió, multipliquen l'impacte negatiu que té en les dones els casos de violència masclista. Un bon exemple és el cas de l'empresa Iveco (Durán *et al.*, 2019), en el qual un treballador de l'empresa va gravar un vídeo sense consentiment mentre va mantenir relacions sexuals amb una companya de feina i ho va difondre entre la resta de companys de l'empresa. La dona va arribar a sofrir com a conseqüència tant d'estrès i ansietat, que va acabar suïcidant-se.

Les adolescents pateixen un important impacte pels continguts, pressions sexistes i interaccions degradants a les xarxes. De fet, existeix una correlació entre la salut mental i la interacció amb la tecnologia de les persones joves (Asociación Internacional Teléfono de la Esperanza, 2022), afectant més les dones, entre les quals hi ha un 30% de prevalença de pràctiques d'autolesions i l'ús compulsiu d'Internet entre els 12 i els 18 anys (Fundación Save the Children, 2022).

En aquest sentit, és prioritari analitzar l'impacte multiplicador que poden tenir les eines d'IA dissenyades sense perspectiva de gènere en aquesta difusió d'estereotips de gènere i llenguatge sexista.

Les tecnologies de dades, per ser ètiques, cal que tinguin present el context social masclista i el potencial d'un ús violent i discriminatori contra les dones d'aquestes eines. Donat el seu paper social com espai de socialització i espai de debat públic, cal que estableixin mecanismes per garantir la privadesa i seguretat específics per a les dones, que han de tenir present que les dones estan exposades a certes formes d'assetjament i violència digital, i cal que siguin capaces d'identificar, prevenir o respondre de manera efectiva aquestes situacions.

Aquest anàlisi ha d'incorporar l'impacte de la bretxa digital de gènere, que està frenant l'avenç de les dones en qualsevol àmbit. En el cas de la violència masclista, el desequilibri en l'ús d'eines digitals avançades entre dones i homes —la tercera bretxa digital— situa les dones en desavantatge i en molts casos en vulnerabilitat en cas de violència masclista utilitzant eines digitals o a l'espai digital.

Tot aquest nou context fa necessari construir noves estratègies per l'abordatge de la violència masclista, ja que la violència digital contra les dones és una manifestació més d'altres formes de violència masclista que es donen de forma simultània.

2. Intel·ligència artificial amb perspectiva de drets de les dones

Cal desenvolupar tecnologies digitals de dades que incorporin la perspectiva de gènere per evitar que aquestes contribueixin a perpetuar biaixos per raó de sexe i siguin beneficioses per a les dones. Per a això cal aplicar les "mateixes" receptes que les agents d'igualtat han utilitat en d'altres àmbits, com les empreses, i no partir del supòsit de neutralitat, sinó que sempre s'ha d'estudiar el punt de partida d'homes i dones. En aquest sentit, cal aplicar aquests criteris als projectes d'IA en tres àmbits:

- Bases de dades
- Entrenament i ajust de l'algorisme (corregir biaixos entre homes i dones)
- Formulació de l'algoritme (incorporar la perspectiva de gènere, el punt de partida del 51% de la població, les dones)

D'altra banda les tecnologies d'IA tenen un gran potencial com a eines que ajudin a abordar els reptes dels drets de les dones, com és el cas de la violència masclista. De la mateixa manera que s'estan desenvolupant i automatitzant molts sectors (aigua, clima, mobilitat, turisme…) amb finançament de la UE per a la digitalització, cal incorporar a aquesta Agenda de Digitalització l'àmbit d'atenció i prevenció de les violències masclistes i drets de les dones. Les iniciatives en aquest àmbit són anecdòtiques encara.

D'aquesta forma, la agenda feminista ha de prioritzar el feminisme de dades per a garantir que les eines digitals que es desenvolupen per a optimitzar i facilitar processos de diversos àmbits, no generin ni perpetuïn biaixos per raó de sexe, així com reclamar i liderar processos de digitalització de les administracions que també inclouen els serveis i reptes de polítiques de drets de les dones.

L'apoderament de l'àmbit feminista respecte a les eines IA passa per utilitzar-les i participar en el seu entrenament, per assegurar que s'introdueixi la perspectiva de gènere; però també liderar la creació d'eines avançades per abordar els reptes pels dels drets de les dones, com la violència masclista. Cal que aquests projectes estiguin liderats i pilotats per dones i equips feministes, per crear models de construcció de projectes digitals i generació de pràctiques en l'àmbit de les tecnologies d'IA.

3. Referències bibliogràfiques

Asociación Internacional Teléfono de la Esperanza (2021). "El Teléfono de la Esperanza sonó más de 183.000 veces en 2021". En línia: https://telefonodelaesperanza.org/noticias/el-telefono-de-la-espera nza-sono-mas-de-183000-veces-en-2021

Durán, Luis F.; Fernández García, Sara i Núñez Villaveirán, Luis (2019). "Una empleada de Iveco se suicida tras viralizarse en la empresa un vídeo sexual", *El Mundo*, 29 de maig. Disponible en: https://www.elmundo.es/madrid/2019/05/28/5ced493efdddffb075 8b48fb.html

Fundación Save the Children (2022). "Suicidios adolescentes en España: Factores de riesgo y datos", 8 de febrer. En línia: https://www.savethechildren.es/actualidad/suicidios-adolescentes-espana-factores-riesgo-datos

Morena-Balaguer, Diana; García-Romeral, Gloria i Binimelis-Adell, Mar (coords.) (2022). *Diagnòstic sobre les violències de gènere contra activistes feministes en l'àmbit digital*. Servei de Publicacions de la Universitat de Vic. Disponible en: https://mon.uvic.cat/ceig/files/2022/01/Calala-CAT_DEF.pdf

Ortiz de Zárate Alcarazo, Lucía (2023). "Sesgos de género en la inteligencia artificial", *Revista de Occidente*, 502(1), 5-20. Disponible en: https://dialnet.unirioja.es/servlet/articulo?codigo=8853265

Indicadores de estereotipos en inteligencia artificial. Test para su detección

María Marta Elustondo
Miguel Ángel Blanco

1. La otra cara de la IA: sesgos de género y violencia machista

La inteligencia artificial (IA) se ha erigido como una fuerza transformadora en diversos ámbitos, prometiendo una nueva era de eficiencia y precisión sin precedentes. Desde los avances en medicina hasta las sofisticadas herramientas financieras, la IA parece estar destinada a revolucionar nuestras vidas (Dorado *et al.*, 2019; Rodríguez Rodríguez *et al.*, 2021; Berrones Salgado, 2023; García y Sánchez, 2023). Sin embargo, detrás de esta fachada de progreso, se esconden problemáticas profundamente arraigadas que amenazan con socavar los beneficios de esta tecnología.

Uno de los desafíos más preocupantes que enfrenta la IA es la perpetuación de los sesgos de género y la violencia machista (Sáinz *et al.*, 2020; Betanzos, 2023; de Zárate, 2023; Ho *et al.* 2025) Estos sesgos no son simples fallas técnicas, sino reflejos de las desigualdades y prejuicios presentes en nuestra sociedad, que se filtran en los sistemas de IA a través de los datos y algoritmos utilizados para crearlos (O'Neil, 2016; Caliskan, 2017; Barocas, 2019; Mehrabi *et al.*, 2021).

Los ejemplos de estos sesgos son numerosos y alarmantes. Desde sistemas de reconocimiento facial que tienen dificultades para identificar rostros de mujeres con tez oscura, hasta herramientas de evaluación de riesgos que califican injustamente a los hombres negros como "de alto riesgo", la IA ha demostrado una tendencia preocupante a reflejar y amplificar los estereotipos y las desigualdades existentes.

La experiencia docente que hemos adquirido en el dictado del Módulo de Violencia de Género, nos demuestra que es difícil deconstruir los estereotipos y prejuicios que están arraigados fuertemente en la comunidad y que se perpetúan sin revisión y reflexión en este sentido (Vázquez, 2015). Es sencillo declamar derechos y equidad, pero, a la hora de deponer algunos privilegios que se asientan en esos desequilibrios, los discursos se acaban y vuelven a aparecer las desigualdades.

Respecto al tema que nos ocupa, los sesgos de género no se limitan a los algoritmos. Incluso los asistentes virtuales, diseñados para interactuar con los usuarios y usuarias, han sido objeto de críticas por su lenguaje y comportamiento denigrante hacia las mujeres. En 2016, el *chatbot* Tay, creado por Microsoft para la plataforma de Twitter, tuvo que ser retirado después de solo 16 horas en línea debido a que comenzó a generar comentarios misóginos, racistas y ofensivos después de interactuar con usuarios de Internet. Este *chatbot* había sido diseñado para reproducir los patrones lingüísticos de una joven estadounidense aprendiendo de las interacciones reales mantenidas en dicha red social. El comportamiento inapropiado de Tay mostró la importancia de utilizar filtros y reglas robustas preventivas en el diseño de la aplicación para reducir la exposición a entornos tóxicos, la necesidad de monitorear de manera continua el sistema para detectar y corregir inmediatamente cualquier desviación, así como la exigencia de que el desarrollo de las aplicaciones incluya mecanismos de seguridad ante ataques y manipulaciones, como los que se produjeron en aquella ocasión.

Estos incidentes no son meros contratiempos técnicos; son síntomas de un problema más profundo enraizado en la forma en que se diseñan, desarrollan e implementan los sistemas de IA. Si no se abordan adecuadamente, estos sesgos pueden conducir a decisiones discriminatorias y perpetuar la violencia simbólica y estructural contra las mujeres y otros grupos históricamente marginados.

Es crucial que los desarrolladores y desarrolladoras de IA sean conscientes de estos sesgos y tomen medidas proactivas para mitigarlos. Esto implica diversificar los conjuntos de datos utilizados para entrenar los algoritmos, implementar auditorías de ética y equidad, y fomentar una mayor participación de mujeres y otros grupos subrepresentados en el campo de la IA.

Además, es esencial que el público usuario final sea crítico y cuestione los resultados y recomendaciones generados por los sistemas de IA, en lugar de aceptarlos ciegamente como verdades objetivas. La IA es una herramienta poderosa, pero no es infalible ni está libre de sesgos.

Abordar estos problemas de manera proactiva y colaborativa es responsabilidad de todos nosotros, desde los desarrolladores hasta los usuarios finales. Solo así podremos garantizar que la IA sea una herramienta verdaderamente inclusiva y equitativa, en lugar de perpetuar las desigualdades existentes. Tenemos que exigir sistemas de IA éticos y libres de prejuicios, y los poderes públicos deben implicarse, como lo ha hecho el Ayuntamiento de Lleida, con la iniciativa que ha dado, entre otros frutos, este libro.

2. La medición de sesgos de género y violencia machista. Una necesidad para evaluar

En nuestra experiencia en investigaciones sobre sesgos de violencia machista debimos construir una metodología de nuestra autoría, ante la ausencia de herramientas de este tipo, que nos permitiera determinar indicadores para reconocer los estereotipos de género.

La investigación se realizó en un espacio de educación superior que arrojó información valiosa, pero cuya difusión, por las características de la organización, no está permitida. Esta situación profundizó nuestra vocación de construir una metodología abierta aplicable a otras organizaciones.

Cabe mencionar que esta propuesta la compartimos en este espacio, convencidos de que es una manera eficaz para intentar desterrar las violencias machistas, reconocer su existencia y medir su impacto en la vida de las mujeres.

La metodología que presentamos en este trabajo es una propuesta abierta, con la finalidad de compartir, enriquecer, generar debate e incluir aportes, ya que, como toda herramienta de acercamiento a las realidades humanas, es factible de mejorar.

Transpolar la metodología a la IA es un desafío que requiere la suma de saberes técnicos y humanísticos. El objetivo de esta propuesta es evaluar la equidad y la existencia de estereotipos en diferentes aspectos de los sistemas de IA y analizar cómo estos sistemas pueden perpetuar o amplificar prejuicios y generalizaciones injustas sobre ciertos grupos.

3. Determinación de indicadores de estereotipos

La determinación de Indicadores de estereotipos en el uso de la Inteligencia Artificial (IA) implica analizar cómo los sistemas de IA pueden perpetuar o amplificar prejuicios y generalizaciones injustas sobre ciertos grupos.

Aquí presentamos algunos pasos que consideramos imprescindibles para identificar estos estereotipos:

- Análisis de datos de entrenamiento:
  Revisar los datos con los que se entrenó el modelo de IA. Si los datos contienen sesgos o representaciones desproporcionadas de ciertos grupos (por ejemplo, un sesgo hacia cierto género o raza), es probable que el modelo reproduzca esos estereotipos.
- Evaluación de resultados:
  Examinar los resultados o predicciones del sistema de IA. ¿Las decisiones o recomendaciones varían significativamente entre diferentes grupos demográficos (como género, raza, edad, etc.)? Si es así, puede ser un indicio de la presencia de estereotipos.

- Pruebas con escenarios diversos:
  Someter el sistema de IA a una variedad de escenarios que incluyan diferentes perfiles de usuarios y usuarias. Observar si las respuestas del sistema son equitativas y justas para todos los grupos o si favorece a un grupo sobre otros.
- Análisis de lenguaje y comunicación:
  Si la IA utiliza lenguaje natural, revisar cómo se comunica. ¿Usa términos, frases o asociaciones que pueden ser considerados ofensivos o estereotipados? El análisis del lenguaje utilizado por la IA puede revelar sesgos subyacentes.
- Revisión de casos reales:
  Estudiar ejemplos de uso de IA en la práctica y analizar incidentes donde la IA haya sido criticada por perpetuar estereotipos. Esto puede dar una idea de dónde suelen aparecer estos problemas y cómo podrían estar presentes en otros sistemas.
- Consulta con personas expertas en Diversidad y Ética:
  Trabajar con expertos y expertas en ética, derechos humanos y diversidad para identificar posibles estereotipos en los sistemas de IA. Su perspectiva puede ayudar a detectar sesgos que no son inmediatamente evidentes.
- Implementación de métricas de equidad:
  Utilizar métricas específicas para medir la equidad y la representación justa de diferentes grupos en los resultados del sistema de IA. Esto puede incluir la evaluación de paridad de resultados, tasas de error diferenciales, etc.
- Revisión de sesgos implícitos:
  Considerar los sesgos implícitos que pueden estar presentes en el desarrollo del sistema. Estos son sesgos que no son intencionales, pero que pueden influir en el comportamiento del modelo de IA debido a decisiones inconscientes durante el proceso de diseño y desarrollo.

Estos indicadores permiten detectar y mitigar estereotipos en la IA, lo que resulta esencial para desarrollar sistemas justos y equitativos, especialmente en contextos sensibles como la contratación, la educación, la salud, y la justicia**.**

4. Test para identificar los estereotipos

Luego de analizar los indicadores, proponemos el siguiente test para identificar estereotipos en el uso de la IA. Se trata de una tabla organizada en categorías clave y enfoque binario, con preguntas específicas para evaluar la equidad y la ausencia de estereotipos en diferentes aspectos de los sistemas de IA. Su función es únicamente diagnóstica.

Tabla 1. Test para Identificar Estereotipos en la IA

Categoría	Pregunta	Respuesta (Sí/No)	Comentarios/ Acciones
1. Representación demográfica	1. ¿El sistema de IA toma decisiones o hace recomendaciones sin depender exclusivamente de características demográficas (género, raza, edad, etc.)?		
	2. ¿El sistema de IA funciona con la misma eficacia para personas de diferentes orígenes culturales o socioeconómicos?		
	3. ¿El conjunto de datos utilizado para entrenar la IA incluye una representación equilibrada de diferentes géneros, razas, y grupos sociales?		
2. Igualdad en resultados	4. ¿Las decisiones del sistema de IA no muestran una tendencia a favorecer a un grupo específico sobre otros?		
	5. ¿Los errores o fallos del sistema de IA afectan por igual a todos los grupos demográficos?		
	6. ¿El sistema de IA ofrece las mismas oportunidades y resultados a personas con diferentes características demográficas?		
3. No perpetuación de estereotipos	7. ¿El lenguaje o las imágenes generadas por la IA evitan reforzar estereotipos de género, raza o clase social?		
	8. ¿El sistema de IA no asocia roles o profesiones específicas con un género o grupo racial particular?		
	9. ¿El sistema de IA presenta a las personas de manera justa, sin utilizar sesgos visuales o de lenguaje que perpetúen estereotipos?		

Categoría	Pregunta	Respuesta (Sí/No)	Comentarios/ Acciones
4. Inclusión y diversidad	10. ¿El diseño y desarrollo de la IA incluyó la participación de personas de diversos orígenes y experiencias?		
	11. ¿El sistema de IA considera la diversidad cultural y no impone un estándar único de comportamiento o apariencia?		
	12. ¿El sistema permite la personalización para adaptarse a las necesidades específicas de diferentes usuarios y usuarias?		
5. Retroalimentación y mejora continua	13. ¿El sistema de IA incorpora mecanismos para recibir retroalimentación sobre posibles sesgos o estereotipos?		
	14. ¿Existe un proceso establecido para revisar y corregir el funcionamiento de la IA cuando se detectan sesgos o estereotipos?		
	15. ¿El equipo de desarrollo realiza auditorías periódicas para asegurar que la IA no perpetúe estereotipos?		

4.1. Interpretación de resultados

- **Cumple con todos los criterios (15/15)**: La IA muestra un excelente nivel de equidad e inclusión, con un bajo riesgo de perpetuar estereotipos. Se recomienda mantener las prácticas actuales y continuar monitoreando el desempeño.
- **Cumple con la mayoría de los criterios (11-14/15)**: La IA es generalmente equitativa, pero se debe realizar una revisión en las áreas donde no cumple para mejorar la equidad y minimizar estereotipos.
- **Cumple con algunos criterios (6-10/15)**: La IA tiene áreas significativas que necesitan mejoras para asegurar que no perpetúe estereotipos. Se recomienda realizar un análisis más profundo y ajustar el diseño y los datos de entrenamiento.
- **Cumple con pocos criterios (0-5/15)**: La IA tiene un alto riesgo de perpetuar estereotipos y requiere revisiones significativas en múltiples áreas. Es necesario un rediseño o una auditoría exhaustiva para corregir estos problemas.

4.2. Cómo usar la tabla

- **Responder a cada pregunta**: Evalúa cada aspecto de tu sistema de IA respondiendo "Sí" o "No" en la columna de "Respuesta".
- **Agregar comentarios o acciones**: En la columna "Comentarios/Acciones", anota observaciones, acciones correctivas o áreas que necesiten una revisión más profunda.
- **Revisión periódica**: Repite este test regularmente para asegurarte de que la IA sigue cumpliendo con los criterios de equidad y no perpetuación de estereotipos.

5. El desafío futuro: Método de Validación de la Herramienta Propuesta

5.1. Objetivo

El objetivo de este método es validar la herramienta propuesta para identificar estereotipos en la IA mediante la evaluación de su efectividad, confiabilidad y aplicabilidad en diferentes contextos.

5.2. Metodología

5.2.1. Selección de casos de prueba

Selección de sistemas de IA: Identificar sistemas de IA que han sido previamente criticados por sesgos o estereotipos conocidos. Estos sistemas servirán como casos de prueba para la herramienta.

Diversidad de contextos: Incluir sistemas de IA en diversos campos como contratación, educación, salud y justicia para asegurar que la herramienta sea aplicable en diferentes contextos.

5.2.2. Evaluación de la herramienta

Aplicación del test: Aplicar el test propuesto en la tabla de identificación de estereotipos a cada sistema de IA seleccionado.

Recopilación de datos: Registrar las respuestas y comentarios para cada pregunta en la tabla.

5.2.3. Análisis de resultados

Interpretación de resultados: Seguir las directrices de interpretación de resultados (4.1) para evaluar el cumplimiento de los criterios de equidad y ausencia de estereotipos.

Comparación con Casos Reales: Comparar los resultados del test con los incidentes conocidos de sesgo en los sistemas de IA seleccionados.

5.2.4. Validación por personas expertas

Panel de personas expertas: Involucrar a un panel de especialistas en IA, ética y derechos humanos para revisar los resultados y proporcionar retroalimentación sobre la efectividad de la herramienta desde el punto de vista de la especialidad de cada uno/a.

Evaluación de personas expertas: Solicitar a especialistas que evalúen la precisión y utilidad de la herramienta en la identificación de estereotipos.

5.2.5. Iteración y mejora

Feedback de personas expertas: Incorporar el *feedback* de los expertos y expertas para mejorar la herramienta.

Revisión de la herramienta: Realizar ajustes en la herramienta basados en las recomendaciones de las personas expertas y en los resultados de la validación.

5.3. Criterios de validación

5.3.1. Efectividad

Detección de sesgos: La herramienta debe identificar correctamente los sesgos y estereotipos conocidos en los sistemas de IA seleccionados.

Precisión: La herramienta debe tener una alta precisión en la identificación de sesgos, minimizando los falsos positivos y negativos.

5.3.2. Confiabilidad

Consistencia: La herramienta debe proporcionar resultados consistentes cuando se aplica a diferentes sistemas de IA similares.

Reproducibilidad: Los resultados deben ser reproducibles por diferentes evaluadores y evaluadoras utilizando la misma herramienta.

5.3.3. Aplicabilidad

Versatilidad: La herramienta debe ser aplicable en diferentes contextos y campos de aplicación de la IA.

Facilidad de uso: La herramienta debe ser fácil de usar y comprender para diferentes usuarios y usuarias, incluyendo desarrolladores y desarrolladoras de IA y especialistas en ética.

5.4. Resultados esperados

Herramienta validada: La herramienta será validada como efectiva y confiable para identificar estereotipos en sistemas de IA.

Mejoras implementadas: Se implementarán mejoras en la herramienta basadas en el *feedback* de personas expertas y en los resultados de la validación.

Documentación: Se documentará el proceso de validación y los resultados para futuras referencias y mejoras continuas.

6. Reflexiones finales

Esta estrategia propuesta es parte de un enfoque integral para abordar el sesgo en la aplicación de la IA, que busca mejorar la equidad y la justicia en el uso de sus modelos. Su enfoque diagnóstico sirve, por un lado, para que la gente común pueda percatarse, de manera sencilla, del trato desigual que algunos sistemas de IA dan a ciertos colectivos y, por otro, como primer paso para un estudio más profundo que quieran realizar especialistas en derechos humanos, en violencia digital, en estudios de género, o los propios poderes públicos.

El reconocimiento y la mitigación de los sesgos son cruciales para desarrollar sistemas de IA más justos y equitativos. El sesgo en la IA puede perpetuar desigualdades, afectando justicia y calidad de vida. Es vital desarrollar tecnologías de IA con enfoque en equidad e inclusión para mitigar estos efectos.

El método de validación intenta asegurar que la herramienta propuesta sea efectiva, confiable y aplicable en diversos contextos. La validación por personas expertas y la iteración basada en el *feedback* garantizarán que pueda cumplir con los estándares necesarios para promover la equidad y la justicia en el uso de la IA.

7. Referencias bibliográficas

Barocas, Solon; Hardt, Moritz y Narayanan, Arvind (2019). *Fairness and Machine Learning*. Cambridge (MA): The MIT Press. Disponible en: fairmlbook.org.

Berrones Yaulema, Laura Pilar y Salgado Oviedo, Sebastián Alejandro (2023). "La aplicación de la inteligencia artificial para mejorar la enseñanza y el aprendizaje en el ámbito educativo". *Esprint Investigación*, *2*(1), 52–60. https://doi.org/10.61347/ei.v2i1.52

Betanzos, Amparo Alonso (2023). "Inteligencia Artificial y sesgos de género". *Gender on Digital. Journal of Digital Feminism*, 1, 11-32. https://doi.org/10.35869/god.vii.5060

Caliskan, Aylin; Bryson, Joanna J. y Narayanan, Arvind (2017). "Semantics derived automatically from language corpora contain human-like biases". *Science*, 356(6334), 183-186.

de Zárate Alcarazo, Lucía Ortiz (2023). "Sesgos de género en la inteligencia artificial." *Revista de Occidente*, 502(1), 5-20. Disponible en: https://ortegaygasset.edu/wp-content/uploads/2023/03/RevistadeOccidente_Marzo2023_L.Ortiz_de_Zarate.pdf

Dorado-Díaz, P. Ignacio; Sampedro-Gómez, Jesús; Vicente-Palacios, Víctor y Sánchez, Pedro L. (2019). "Aplicaciones de la inteligencia artificial en cardiología: el futuro ya está aquí". *Revista Española de Cardiología*, 72(12), 1065-1075. Disponible en: https://www.sciencedirect.com/science/article/abs/pii/S0300893219302507

García Moreno, Elizabet y Sánchez Balcázar, Marcela del Carmen (2023). "Efectos de la aplicación de la Inteligencia Artificial en la contabilidad y la toma de decisiones". *Gestión*, 1(1), 37-43. Recuperado de: https://revistap.ejeutap.edu.co/index.php/Gestion/article/view/71

Ho, Jerlyn Q.H.; Hartanto, Andrée; Koh, Andrew y Majeed, Nadyanna M. (2025). "Gender biases within Artificial Intelligence and ChatGPT: Evidence, Sources of Biases and Solutions". *Computers in Human Behavior: Artificial Humans*, 4, 100145. https://doi.org/10.1016/j.chbah.2025.100145

Mehrabi, Ninareh, Morstatter, Fred, Saxena, Nripsuta, Lerman, Kristina y Galstyan, Aram (2021). "A Survey on Bias and Fairness in Machine Learning". *ACM Computing Surveys* (CSUR), 54(6), 1-35. https://doi.org/10.1145/3457607

O'Neil, Cathy (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.

Rodríguez Rodríguez, Alberto; Romero Castro, Vicente Fray; Rodríguez González, Antonieta del Carmen; Cabezas Baque, Nicolás Alfonso y Pino Tarragó, Julio César (2021). "Aplicaciones de la Inteligencia Artificial en técnicas de minería de procesos". *Serie Científica de la Universidad de las Ciencias Informáticas,* 14(7), 136-155. Disponible en: https://dialnet.unirioja.es/servlet/articulo?codigo=8590663

Sáinz, Milagros; Arroyo, Lidia y Castaño, Cecilia (2020). "Participación de las mujeres en el diseño, la producción y la aplicación de las RIC para la vida, así como la toma de decisiones vinculadas a las TIC". En *Mujeres y digitalización. De las brechas a los algoritmos*. Madrid: Instituto de la Mujer y para la Igualdad de Oportunidades. Disponible en: https://www.inmujeres.gob.es/diseno/novedades/M_MUJERES_Y_DIGITALIZACION_DE_LAS_BRECHAS_A_LOS_ALGORITMOS_04.pdf

Vázquez Cupeiro, Susana (2015). "Ciencia, estereotipos y género: una revisión de los marcos explicativos". *Revista de Ciencias Sociales*, 68, 177-202. Disponible en: https://dialnet.unirioja.es/servlet/articulo?codigo=5209884

Las mujeres en la era de la ingeniería y la tecnología: un camino hacia la igualdad

Marta Rosa Poiasina García

La figura de la mujer en el ámbito de la ingeniería y la tecnología ha tomado un protagonismo creciente en los últimos años. Voces influyentes como la de María Laura Orfanó, ingeniera de sistemas por la Universidad Tecnológica Nacional de Buenos Aires y fundadora y CEO de la empresa tecnológica *Simbel Ecommerce*, resaltan los logros alcanzados por las mujeres y los desafíos que deparan el presente y el futuro inmediato. En su emotivo discurso pronunciado en el acto de entrega de títulos en la Universidad Tecnológica Nacional de Buenos Aires, puso en valor el talento de las mujeres en el mundo de las ingenierías, al mismo tiempo que destacó la necesidad de incrementar su presencia en las universidades, como estudiantes, profesoras e investigadoras, así como en de las empresas tecnológicas.

A medida que nos adentramos en la era de la Inteligencia Artificial (IA), la necesidad de mujeres en todos estos campos es más urgente que nunca. Este artículo es una breve reflexión sobre la relación de la mujer con la IA, su impacto en el ámbito empresarial y académico, y su creciente papel en la evolución del desarrollo tecnológico.

1. Las mujeres y su relación con la IA

La inteligencia artificial es, sin duda, uno de los campos más revolucionarios de nuestra época y, sin embargo, sigue siendo un terreno desproporcionadamente dominado por hombres. En consecuencia, las ingenieras pueden y merecen tener un papel protagonista en esta revolución tecnológica.

Los datos del Instituto de Estadística de la Unesco, recogidos en el estudio de Naciones Unidas sobre la brecha de género en las carreras STEM señalan que, a nivel mundial, la tasa promedio de mujeres STEM en el ámbito de la investigación apenas alcanzaba el 30% en el año 2019 (Bello, 2020). A pesar del creciente aumento de mujeres en este tipo de estudios, su presencia continúa siendo inferior en los niveles de especialización y de doctorado, en aquellos precisamente que capacitan para la investigación, resultando una pérdida de talento que debemos revertir. Este mismo estudio pone de relieve que en América Latina y El Caribe, se ha logrado alcanzar la paridad en las graduaciones universitarias, una paridad que se va perdiendo en los estudios postuniversitarios y, con ella, las mujeres pierden oportunidades para acceder a los empleos bien pagados que se están generando en este sector. El mismo estudio revela que los empleos mejor pagados del ámbito STEM son los que están vinculados a la IA (Bello, 2020: 31).

El beneficio de la presencia de mujeres en el desarrollo de la IA es crucial, no solo en el nivel individual si no también y sobre todo en el desarrollo social. La diversidad de pensamiento que aportan es necesaria para el desarrollo de tecnologías inclusivas, de herramientas y soluciones que respondan a las necesidades de toda la sociedad, teniendo en cuenta la diversidad de situaciones y colectivos.

Fomentar el interés de las mujeres en la IA desde una edad temprana es vital; las iniciativas que estimulan la curiosidad en la ciencia y la tecnología son elementos fundamentales para cultivar una nueva generación de innovadoras. La educación, por tanto, se convierte en el primer paso hacia un futuro en el que la IA refleje la diversidad de la humanidad.

2. La importancia de las mujeres en las empresas tecnológicas

La trascendencia de la participación femenina en las empresas tecnológicas, es un mensaje crucial que debemos transmitir. Nuevamente las estadísticas dan argumentos a favor de la inclusión de las mujeres al mostrarnos que las empresas con mayor diversidad de género tienden a tener mejores resultados financieros y una cultura de trabajo más dinámica.

La inclusión de mujeres en cargos de liderazgo no solo es una cuestión de justicia social, sino también una estrategia empresarial inteligente. Cada vez más, los informes evidencian que las empresas que promueven la diversidad y la inclusión son más innovadoras y están mejor posicionadas para resolver problemas complejos debido a la variedad de perspectivas que se reúnen.

Además, la experiencia sirve como un poderoso ejemplo de cómo las mujeres pueden triunfar y liderar en un ámbito previamente considerado dominado por hombres. Visualizar esos liderazgos y sus éxitos, sienta un precedente inspirador que puede ayudar a

atraer y retener a más mujeres en este sector y servir de referente a las niñas y jóvenes en la elección de sus estudios y en cómo se proyectan de mayores como profesionales

3. Las mujeres en la universidad

La educación es la piedra angular para facultar a las mujeres en el campo de la ingeniería. Hay que considerar el incremento del 38% en el número de mujeres que eligen carreras de ingeniería en la última década a escala mundial.

Sin embargo, la vida universitaria también presenta desafíos únicos; muchas mujeres se enfrentan a la presión de equilibrar sus estudios con expectativas culturales y familiares. Por lo tanto, es vital que las universidades no solo ofrezcan apoyo académico, sino que también promuevan un entorno inclusivo. Implementar programas de mentorización, donde ingenieras experimentadas guíen a las nuevas generaciones, puede ser un paso efectivo para facilitar esta transición. También el desarrollo de programas de corresponsabilidad de los hombres en las tareas de cuidados y de conciliación de la vida personal, familiar y laboral para facilitar el desarrollo académico y profesional de las mujeres en la universidad.

Además, la promoción de actividades que permitan la colaboración de mujeres y hombres en la formación e investigación ayudará a romper los estereotipos y a construir una comunidad universitaria más cohesionada en la que todos los talentos se pongan en valor, sin exclusiones de género.

4. Las mujeres y la tecnología

La relación de las mujeres con la tecnología va mucho más allá de su papel como usuarias de herramientas: son creadoras, innovadoras y líderes de proyectos. Por lo tanto, quiero subrayar que no solo es necesaria la habilidad de las ingenieras en su capacidad técnica, sino también en su enfoque humanista y su capacidad para captar los problemas sociales y buscar soluciones tecnológicas. En un mundo donde la tecnología avanza a pasos agigantados, es crucial que las mujeres ofrezcan buenas prácticas y valores éticos en el desarrollo tecnológico, más allá de los objetivos de mercado.

Fomentar una cultura de colaboración y respeto en el ámbito tecnológico no solo mejora el ambiente laboral, sino que también garantiza que las innovaciones sean accesibles y útiles para todos.

Además, el liderazgo femenino puede ser el catalizador que impulse el cambio social; las ingenieras pueden utilizar sus plataformas para abogar por políticas que promuevan la equidad y la inclusión dentro de sus empresas y en la sociedad.

5. Conclusiones

El discurso de Maria Laura Orfanó a las graduadas de la Universidad Tecnológica Nacional de Buenos Aires con el que empecé este artículo me permite cerrar con la reflexión de que las mujeres son una fuerza vital para el avance de la ingeniería y la tecnología en el actual panorama global.

Después de mis más de sesenta años de docencia en la escuela primaria y secundaria y en la universidad, enseñando matemáticas a distintas generaciones de ingenieras e ingenieros, me permito pasar el testigo a las ingenieras de hoy para que tomen el relevo en la responsabilidad de modelar un futuro en el que la diversidad y la igualdad de género formen parte de desarrollo tecnológico.

Al empoderar a la próxima generación con educación, apoyo y un entorno laboral equitativo, estamos construyendo los cimientos para un mundo mejor, donde tanto hombres como mujeres trabajen juntos en la creación de soluciones innovadoras y sostenibles.

Hoy en día, el desafío para cada ingeniera no es solo el desarrollo profesional en el ámbito del conocimiento adquirido, sino ser también agente de cambio. La celebración de cada logro es esencial en este camino.

6. Referencia bibliográfica

Bello, Alessandro (2020). *Las mujeres en ciencias, tecnología, ingeniería y matemáticas en América Latina y el Caribe*. Montevideo: ONU Mujeres. Disponible en: https://lac.unwomen.org/es/digiteca/publicaciones/2020/09/mujeres-en-ciencia-tecnologia-ingenieria-y-matematicas-en-america-latina-y-el-caribe

Breve glosario de términos sobre IA, redes y violencia de género

Mª Ángeles Calero Fernández
Universidad de Lleida

algoritmo: En el lenguaje informático, procedimiento computacional consistente en definir una serie de valores de entrada y aplicar una secuencia de operaciones preestablecidas con el objeto de llegar a la solución de un problema o resolver una incógnita.

app(s): Aplicación informática para dispositivos móviles y tabletas. También se utiliza para referirse a una aplicación web, es decir, un programa de software que no se descarga, sino que se ejecuta en un navegador o en un servidor. Es una abreviatura de la palabra inglesa *application*, usada en estos sentidos.

big tech: Nombre que se da a las grandes empresas tecnológicas que desarrollan su actividad a nivel mundial y que, dado su gran alcance operacional y su potencial económico, pueden llegar a ofrecer incluso servicios financieros. Es un anglicismo cuya traducción es 'gigante tecnológico'.

chatbot: Programa informático basado en Inteligencia Artificial que simula mantener una conversación con el usuario o usuaria bien oralmente, bien por escrito, ofreciendo respuestas automáticas. Este programa constituye un modelo de lenguaje construido mediante técnicas de aprendizaje supervisadas y de refuerzo. Ejemplos de chatbot son los asistentes Siri o Alexa. Es un anglicismo a partir de *chat (rob)bot*, cuya traducción es 'robot de charla' o 'robot conversacional'.

chatgpt: Es una aplicación de *chatbot* que desarrolló en 2022 OpenAI, empresa norteamericana especializada en Inteligencia Artificial. Es el acrónimo de *Chat Generative Pre-trained Transformer*, cuya traducción es 'transformador preentrenado que genera conversación'.

ciberacoso: Acoso, espionaje o intimidación que se lleva a cabo a través de herramientas de internet y de redes sociales con el objeto de molestar, humillar o atemorizar a personas o colectivos concretos. Puede manifestarse como difusión de mentiras, publicación de fotografías o vídeos vergonzosos, íntimos o amenazantes, envío de mensajes con similares intenciones, suplantación de una persona para realizar actos inadecuados o delictivos, etc.

ciberstalking: Anglicismo equivalente a *ciberacoso* o *acoso cibernético*.

deepfake: Archivo de vídeo, imagen o voz de contenido falso pero con apariencia de ser auténtico que ha sido construido manipulando otros archivos existentes mediante programas de Inteligencia Artificial. Aunque pueda tener utilidades prácticas (han sido usados, por ejemplo, para generar escenas que no podían ser filmadas por la ausencia del actor o de la actriz), en general se emplean para inducir a error al público destinatario o para perjudicar a ciertas personas a las que se suplanta. Es un anglicismo que se traduce por 'ultrafalso'.

deepnude: Aplicación informática basada en la Inteligencia Artificial que modifica fotos, habitualmente de mujeres, eliminando la ropa que llevan y creando falsos desnudos. Es un anglicismo construido a partir de *deep* ('profundo') y *nude* ('desnudo').

doxing: Revelación de información privada o confidencial de una persona o de un colectivo sin su consentimiento para dañar su imagen pública o profesional. Es un anglicismo construido a partir de *docs*, que es una abreviatura de *documents* ('documentos').

filtro burbuja: Estado de aislamiento de contenidos que sufre una usuaria o usuario resultante de la personalización de páginas web, vídeos, audios, etc. que los algoritmos realizan a partir del historial de búsquedas y de la ubicación geográfica de la usuaria o usuario. Esto genera un sesgo informativo que se retroalimenta. Es una expresión acuñada por Eli Pariser en 2011 en su libro *The Filter Bubble: What the Internet is hiding from you*.

flaming: Publicación de mensajes deliberadamente hostiles, ofensivos o injuriosos en una red social, foro o lista de correo electrónico atacando a una persona o a un colectivo, en el contexto de una discusión, con la voluntad de provocar, intimidar o generar reacciones violentas. Es un anglicismo, del verbo *flame* ('incendiar').

gamer: Persona que juega a videojuegos intensamente. También se designa así a quienes se dedican profesionalmente a esta actividad. Es un anglicismo cuya traducción es 'videojugador/a'.

grooming: Engaño de una persona adulta, a través de redes sociales, de chats de videojuegos o de otros canales de comunicación digital, que contacta con menores haciéndose pasar por alguien de su edad e interactuando hasta ganarse su confianza, con el objetivo final de abusar sexualmente de ellos o ellas. Es un anglicismo que puede traducirse por 'engaño pederasta'.

hater: Persona que insulta, critica o calumnia de manera sistemática y malitencionada a otra(s) persona(s) o colectivo(s) en las redes sociales o en cualquier otra plataforma digital, con el objetivo de humillarla(s) y/o dañar su imagen. Es un anglicismo cuya traducción literal es 'odiador/a', 'el/la que odia'.

inteligencia artificial: Disciplina de las ciencias de la computación que se dedica a diseñar programas informáticos que realizan tareas comparables a las que lleva a cabo la mente humana, como aprender o razonar, a partir de la recopilación de cantidades ingentes de información y de la ejecución de una serie de operaciones previamente establecidas. También es la serie de capacidades cognoscitivas que expresan estos programas informáticos.

inteligencia artificial generativa: Tipo de inteligencia artificial que va más allá de procesar, clasificar y analizar datos, siendo capaz de generar contenidos nuevos y originales de todo tipo: textuales, audiovisuales, de programación...

metadato: Dato que ofrece información de otro dato, no de la realidad. Los metadatos describen características de los datos, lo que permite, por ejemplo, organizar, identificar, ubicar o acceder a recursos, encontrar información relevante, etc.

métrica: En la terminología de las ingenierías, estadística descriptiva para medir una actividad, proceso o realidad cuantitativamente y compararla con otras.

offline: Aquello que puede realizarse en un dispositivo sin estar conectado a la red o todo aquello que es accesible sin usar internet. Es un anglicismo cuya traducción es 'sin conexión', 'desconectado/a' o 'fuera de internet'.

online: Conectado/a a una red de datos o de comunicación, o disponible a través de internet. Es un anglicismo que se traduce, en el primer sentido, por 'en línea', y, en el segundo, por 'digital' o 'electrónico'.

open data: Información de acceso libre que puede ser utilizada, reutilizada o distribuida sin restricciones o sujeta, como mucho, al reconocimiento de la autoría y/o al respeto a la forma en que ha sido originalmente presentada. Es un anglicismo cuya traducción es 'datos abiertos'.

pornografía de venganza: Publicación de fotografías o vídeos eróticos o sexuales privados sin el consentimiento de la persona que aparece en ellos, con el objeto de humillarla y/o dañar su imagen, especialmente en el contexto de una ruptura amorosa.

prompt: Conjunto de instrucciones verbales que un usuario o usuaria proporciona a un sistema de Inteligencia Artificial para conseguir el producto concreto que espera: un diálogo, un texto, una imagen, un vídeo... Se formula en lenguaje natural en forma de pregunta o de un enunciado y debe contener información suficiente para que la herramienta entienda lo que el usuario o usuaria busca obtener. Es un anglicismo que puede traducirse por 'apunte', 'indicación'.

sextorsión: Chantaje consistente en amenazar a una persona con publicar fotografías o vídeos de contenido sexual a través de medios digitales para obtener un beneficio económico, controlarla o victimizarla.

software: En el lenguaje informático, conjunto de programas y de instrucciones que permiten al ordenador realizar tareas concretas.

stalking: Seguimiento obsesivo o vigilancia incesante que alguien realiza sobre otra persona a través de las redes sociales, enviándole mensajes no deseados o creando perfiles falsos que permiten espiarla. Es un anglicismo cuya traducción es 'acecho' o 'acoso'.

STEM: Siglas de la secuencia formada por las palabras inglesas *Science*, *Technology, Engineering and Mathematics* (en español, 'Ciencia, Tecnología, Ingeniería y Matemáticas'), que se refiere a un nuevo modelo educativo que integra todas estas disciplinas y que se empieza a plantear en los años 90 del siglo XX.

streaming: Tecnología que permite transmitir contenidos audiovisuales de manera continua a través de internet, desde un servidor remoto hasta un dispositivo (ordenador, móvil o televisor). Es también la trasmisión concreta de dichos contenidos hasta el usuario o usuaria o el consumo de los mismos en tiempo real. Es un anglicismo que proviene del término *stream* ('arroyo', 'corriente', 'chorro') y que se traduce por 'emisión en continuo' o 'transmisión en directo'.

troleo: Acción de trolear o difundir, en chats, foros digitales y/o redes sociales, mensajes inoportunos, sarcásticos, irreverentes, ofensivos o provocativos para hostigar a alguien, alterar u obstaculizar la conversación o boicotear algo.

troll: Persona que dice cosas negativas o difunde mensajes inoportunos, sarcásticos, irreverentes, ofensivos o provocativos en chats, foros digitales y/o redes sociales para hostigar a alguien, alterar u obstaculizar la conversación provocando enfrentamientos o boicotear algo. Es un anglicismo que puede traducirse por 'provocador/a'.

Universitat de Lleida

Ajuntament de Lleida