Los engranajes de la máquina. Poder y desigualdades en la inteligencia artificial

VV.AA.

Documento de trabajo.
Diciembre 2024

Direcció de Justícia Global i Cooperació Internacional de l'Ajuntament de Barcelona

Ajuntament de Barcelona

Comisariado:
Oxfam Intermón

Fecha de publicación:
Diciembre de 2024

Coordinación:
Carlos Bajo Erro

Autoría: Afef Abrougui, Sarah Chander, Ervin Félix, Pablo Jimenez Arandia, Pelonomi Moiloa, No Tech For Apartheid, Salvatore Romano, Sofia Scasserra, Sara Suárez-Gonzalo, Ana Valdivia, Paola Villareal y Thomas Wright.

Revisión: Liliana Arroyo Moliner, Carlos Bajo Erro, Aina Gallego, Thai Jungpanich, Judith Membrives i Llorens, Paz Peña, Ismael Peña-López, Natalia Pereira Martín, Sebastián Ruiz Cabrera, Susana Ruiz Rodríguez, Hernán Saenz Cortés.

Traducción en la versión en castellano: Arantxa Albiol Benito, Belén Carneiro y Camino Villanueva.

Traducción en la versión en inglés: Kim Causier y Teri Jones-Villeneuve.

Maquetación:
Jimena Zuazo.

Publicado por la Direcció de Justícia Global del Ajuntament de Barcelona.

"Esta publicación es una aportación a la discusión sobre un tema controvertido. Ni el abordaje, ni el desarrollo de los estudios, ni las conclusiones extraídas por cada una de las expertas tienen por qué reflejar o ser compartidas por la Direcció de Justícia Global del Ajuntament de Barcelona, ni por Oxfam Intermón, ni por el resto de autoras o las organizaciones en las que participan".

Contenido

Aclaraciones metodológicas previas 1

INTRODUCCIÓN

Capítulo 1. Introducción: inteligencia artificial y desigualdades 3

Sara Suárez-Gonzalo

PARTE I: IMPACTO EN LOS ÁMBITOS ESTRUCTURALES

Capítulo 2. Código colonizado: Desentrañando las desigualdades económicas en la era de la IA 13

Sofia Scasserra

Capítulo 3. El impacto medioambiental de la Inteligencia Artificial: No es una nube, sino una nave industrial 25

Ana Valdivia

PARTE II: IMPACTO EN LOS ÁMBITOS DE LA VIDA

Capítulo 4. Aceleracionismo militar: Inteligencia Artificial, *big tech* y genocidio en Gaza 37

No Tech For Apartheid

Capítulo 5. Raza y resistencia: Explorando las economías políticas de la inteligencia artificial 56

Sarah Chander

Capítulo 6. El impacto de la Inteligencia Artificial Lingüística en el acceso al conocimiento y su producción 70

Pelonomi Moiloa

Capítulo 7. Decisiones algorítmicas y derechos vulnerados en la *gig economy* 85

Paola Villareal / Ervin Félix

Capítulo 8. Algoritmos que señalan y castigan. ¿Qué hay en juego con la automatización de las prestaciones sociales? 101

Pablo Jimenez Arandia

Capítulo 9. IA y elecciones: cómo afectaron los chatbots y las imágenes de IA generativa a la campaña electoral en las elecciones europeas de 2024 112

Salvatore Romano / Thomas Wright

Capítulo 10. Exacerbación de la violencia, la vigilancia y la exclusión económica: El impacto de género de AI en la región MENA 126

Afef Abrougui

Recapitulación: Hilvanar las dimensiones y tejer otro modelo de inteligencia artificial 147

Carlos Bajo Erro

Nota metodológica aclaratoria

Durante los últimos años, las herramientas basadas en inteligencia artificial y los sistemas algorítmicos se han ido haciendo presentes en más y más ámbitos de las vidas cotidianas de una parte creciente de la población del planeta. Al mismo tiempo, el ciclo de vida de estos sistemas, el diseño, el desarrollo y la implantación, básicamente, se ha ido expandiendo y complejizando, dando lugar a una de las cadenas de suministro más globales e interconectadas de la economía global. Esta circunstancia ha hecho que la influencia del fenómeno de expansión de la inteligencia artificial y de los sistemas algorítmicos tenga impactos más allá de las personas usuarias o las destinatarias directas. Aquellos territorios o grupos de población que se encuentran más alejados del proceso global de digitalización, los y las desconectadas, también experimentan las consecuencias de la adopción creciente de esta tecnología.

La Direcció de Justícia Global del Ajuntament de Barcelona se propuso impulsar un volumen en el que se explorasen las diferentes capas de estos impactos en las vidas de las personas, como una aportación a la comprensión social de todas las caras de este fenómeno y desde la promoción de una mirada crítica a las intersecciones de la digitalización, los derechos, la democracia y las desigualdades. La administración municipal encargó el comisariado de este documento a Oxfam Intermón, en el marco de la tercera fase del proyecto de Justicia Digital Global que comparten las dos entidades.

Esa obra se concibió como un espacio de encuentro de diferentes voces, expertas en diferentes ámbitos, cada una de ellas con las particularidades de sus enfoques y de sus propias experiencias, que compartían el resultado de sus investigaciones en sectores concretos. El texto, intencionadamente diverso y heterogéneo, pretendía mostrar la complejidad del fenómeno y sus múltiples ramificaciones, interacciones e, incluso, interpretaciones. Para sistematizar la mirada hacia un fenómeno con tantas aristas como el de los impactos de la inteligencia artificial en el momento actual, era necesario buscar un marco que contemplase no solo la existencia de las numerosas caras de ese prisma, sino también sus vínculos, sus conexiones y sus relaciones. Por ello, Oxfam Intermón recurrió al Marco de las Desigualdades Multidimensionales (MIF, por sus siglas en inglés), un armazón conceptual y metodológico para medir las desigualdades desde un enfoque multidimensional que fue concebido y elaborado por el Centre for Analysis of Social Exclusion (CASE) de la London School of Economics and Political Science (LSE), la School of Oriental and African Studies (SOAS) y Oxfam.

El MIF se basa en el Enfoque de Capacidades formulado por Amartya Sen, que establece las dimensiones vitales que influyen en la consecución del bienestar por parte de las personas. El trabajo que el MIF hace sobre esa base teórica y filosófica construye un marco para medir de una manera sistemática, rigurosa y completa las desigualdades y, por ello, identifica una serie de indicadores objetivables y mesurables, para evaluar la desigualdad en la calidad de vida de las personas. Esta pretensión de medición excede el interés del volumen sobre el impacto de la inteligencia artificial en las desigualdades, sin embargo, la identificación clara de esas dimensiones que condicionan el acceso al bienestar, sí que sirve como un sustrato útil para establecer los ámbitos de la vida a los que es necesario prestar atención. De esta manera, el MIF sirve como cimiento y como marco conceptual, para una aproximación a los impactos diversos de la inteligencia artificial.

Siguiendo ese marco que ha sido ampliamente aceptado, contrastado y utilizado por diferentes instituciones internacionales y centros de estudio de alcance global, el volumen impulsado por la Direcció de Justícia Global del Ajuntament de Barcelona recoge los siete "dominios" que identifica el MIF como ámbitos principales de la vida para determinar el impacto de las desigualdades: vida y salud; seguridad física y legal; educación y aprendizaje; seguridad financiera y trabajo digno; condiciones de vida adecuadas; participación, influencia y voz; vida individual, familiar y social. Para terminar de adaptar el marco a las necesidades concretas de este volumen, se han añadido dos dimensiones con una mirada global, que no estaban contemplados en el MIF, debido a su enfoque en el bienestar individual: la influencia geoestratégica con una mirada económica y el impacto ecosocial.
Como se ha visto, el marco general que ofrece el MIF tiene que ser adecuadamente adaptado, para responder a algunas necesidades concretas. En este caso, por ejemplo, a pesar de que la clasificación establece dimensiones independientes, se realiza un esfuerzo para aproximarse a las interseccionalidades, la interrelación entre algunas de esas capas. Del mismo modo, bajo el sustrato del MIF existe una aproximación concreta a la idea de las desigualdades, desde el tratamiento en plural del fenómeno, como una expresión de múltiples asimetrías o una huida de una mirada exclusivamente socioeconómica y el abordaje de las desigualdades desde otras miradas; o la atención a otras cuestiones simbólicas y subjetivas que tiene que ver con la construcción de significado, además de las dimensiones concretas.

Los nueve ámbitos identificados a partir de la adaptación del MIF conforman la estructura del documento. Para cada uno de esos ámbitos hay múltiples abordajes posibles y en este volumen exploratorio se ha tenido que elegir uno concreto que pudiese ser representativo e interesante para el público y para la construcción de la mirada poliédrica que se busca. Para ello, se ha realizado un proceso de consulta colectivo, comisariado por Oxfam Intermón. Como obra de contexto, la reflexión y el análisis sobre cada uno de esos ámbitos ha sido encargado a una organización o una persona experta especializada en esa dimensión y que ya hubiese hecho investigaciones en ese sentido. Ni el abordaje, ni el desarrollo de esos estudios, ni las conclusiones extraídas por cada una de las expertas tienen por qué ser compartidas por la Direcció de Justícia Global del Ajuntament de Barcelona, ni por Oxfam Intermón, pero sus aportaciones se consideraban valiosas para promover y articular el debate.

Precisamente, por el objetivo y el sentido de la publicación, que pretende ser una aportación a la discusión sobre un tema controvertido, las y los autores de las diferentes piezas han transmitido el resultado de sus investigaciones sin tener que cumplir directrices específicas de contenido, más allá del rigor de sus análisis. Esta circunstancia hace que cada una de las piezas de este puzle no refleje las opiniones del resto de autoras o de las organizaciones implicadas en su publicación y que tampoco refleje las posiciones individuales de las otras expertas participantes, ni de las entidades impulsoras del proyecto.

1. Introducción: inteligencia artificial y desigualdades

Sara Suárez-Gonzalo
Universitat Oberta de Catalunya - Communication Networks and Social Change / Internet Interdisciplinary Institute

Los datos más recientes muestran que las desigualdades, lejos de remitir, son un problema creciente. Los desequilibrios son profundos y las barreras son cada vez mayores para vivir una vida libre y en igualdad de oportunidades. Aspectos como la identidad de género, el lugar de origen o residencia, los rasgos físicos, la edad, el nivel de estudios o de ingresos de las personas y, especialmente, las combinaciones de varios de estos factores, determinan, todavía de forma clara, su posición social, sus capacidades y sus oportunidades. Como causa y, al mismo tiempo, consecuencia de ello, la riqueza se concentra cada vez más en las manos de los propietarios de las grandes empresas y, más concretamente, de las corporaciones tecnológicas. En un contexto como este, no es de extrañar que el desarrollo tecnológico y, en particular, de una tecnología con cada vez más efectos sobre la vida cotidiana como la inteligencia artificial, contribuya a profundizar los desequilibrios, en lugar de favorecer la mejora de las condiciones de vida del conjunto de la población.

Se suele decir que la inteligencia artificial (IA) “ha venido para quedarse”. Pero no ha venido, alguien la ha traído. Entonces quién, cómo, a dónde, por qué o para qué (la ha traído) devienen preguntas fundamentales. Y también cómo y a quién afecta todo ello.

Este volumen, impulsado por la Direcció de Justícia Global del Ajuntament de Barcelona y comisionado por Oxfam Intermón, se plantea, precisamente, con el objetivo de arrojar luz sobre estas y otras preguntas que ofrecen un marco para entender un poco mejor cómo se relaciona la inteligencia artificial con las desigualdades. Además, da algunas pistas sobre cómo avanzar hacia un futuro más libre, justo y equitativo.

Las desigualdades, en auge

Para entender las implicaciones presentes y futuras de la inteligencia artificial es imprescindible entender el contexto en el que ésta se desarrolla y en el cual se integra y que se trata de uno marcado por profundas y crecientes desigualdades socio-económicas. Empecemos por tanto por la que quizás es la pregunta menos habitual: *¿a dónde?* O lo que es lo mismo: ¿cuál es el contexto en el que se despliega la inteligencia artificial?

Los datos demuestran que **las desigualdades son un problema acuciante a escala global.** El caso español no es una excepción. El reciente informe *Desigualdad S.A. El poder empresarial y la fractura global: la urgencia de una acción pública transformador*, publicado por Oxfam Intermón (Riddell, Ahmed, Maitland, Lawson y Taneja, 2024) pone de manifiesto que, 4.800 millones de personas son más pobres hoy que hace un lustro. Las mujeres, las personas racializadas y otros grupos sociales en situación de exclusión social son las más afectadas por esta situación que, además, se traduce en una brecha cada vez mayor entre el Norte y el Sur globales, de la que salen muy beneficiados los países del Norte, pese al bajo porcentaje de la población mundial que vive en ellos (según datos del citado informe, solo el 21%). Al mismo tiempo, señala una concentración creciente de la riqueza, a nivel global, en pocas manos. Generalmente, las de los propietarios de las grandes empresas que dominan los mercados y, en particular, el tecnológico. Precisamente, en la carrera por el monopolio del sector tecnológico, los propietarios de gigantes digitales (también conocidos como las *big tech*) como Meta, Alphabet o Amazon se encuentran ya entre los más ricos del mundo (Fortune, 2023; Murphy and Schifrin, 2024; Statista, 2024).

Las causas y las dimensiones en las que se reflejan las desigualdades son múltiples y complejas, y están profundamente interrelacionadas. Su estudio y su denuncia forman parte de la extensa trayectoria de Oxfam Intermón. ***El Multidimensional Inequality Framework*** (Marco de Desigualdad Multidimensional) es una contribución basada en el trabajo de Amartya Sen, elaborado como fruto de una colaboración entre académicos del Centro de Análisis de la Exclusión Social (CASE) de la London School of Economics (LSE) y la Escuela de Estudios Orientales y Africanos (SOAS), dirigidos por Abigail McKnight, y profesionales de Oxfam. El marco se centra en la medición de determinados indicadores, en base a los cuales se identifican una serie de dominios de la vida de las personas en las que se manifiestan las desigualdades: la vida individual, familiar y social; la vida y la salud; la participación, influencia y la voz; las condiciones de vida adecuadas; la seguridad financiera y el trabajo digno; la educación y el aprendizaje; y la seguridad física y legal. Por otra parte, el *VI Informe sobre la desigualdad en España 2024*, publicado por la Fundación Alternativas, recoge datos actuales y reflexiones pertinentes a este respecto. El capítulo a cargo de García López (2024) interpreta datos relevantes de la encuesta de Oxfam Intermón y 40dB sobre percepciones sociales de las desigualdades en España realizada durante el segundo semestre del 2023. Según este estudio, **ocho de cada diez personas en España creen que existen desigualdades**. La percepción es más intensa entre las mujeres, las personas de las cohortes de edad más avanzada (más de 65 años) o con niveles adquisitivos medios o medios-altos y aquellas que se auto-identifican como blancas o caucásicas. Si

nos fijamos en algunas de las distintas dimensiones (aunque estén interconectadas) de las desigualdades sociales, la económica (diferencias entre ricos y pobres) es la que más percibe la población española, seguida de la migratoria (especialmente entre las personas en situaciones administrativas irregulares) y las territoriales, entre los distintos barrios que componen los espacios urbanos. No obstante, la mayor parte de la población considera que estas desigualdades son erradicables, principalmente, si el gobierno central, la Unión Europea, las comunidades autónomas, los ayuntamientos, los medios de comunicación, los movimientos sociales y las empresas (por este orden de importancia percibida) les prestan la atención y proveen los recursos necesarios para ello.

Los procesos de digitalización de la sociedad, acelerados por las medidas de gestión de la pandemia por Covid-19 (Ayala Cañón et al., 2022) también contribuyen, de distintas formas, a la reproducción y la profundización de estas y otras desigualdades. Por una parte, las inversiones en materia de digitalización son profundamente dispares. En el territorio español son una muestra de ello las diferencias en la captación por parte de las comunidades autónomas de fondos como los europeos Next Generation en materia de investigación, desarrollo e innovación (I+D+i) y en materia de digitalización. En ambos casos, la Comunidad de Madrid es la más beneficiada. Por otra parte, existe una profunda brecha socio-digital, que provoca que la digitalización, especialmente, pero no solo, de aquellos servicios y productos básicos para la vida cotidiana (como aquellos que median en la relación con las administraciones públicas, bancos, servicios médicos, pero también los que permiten la comunicación y la relación con otras personas y agentes públicos y privados) afecte de maneras muy desiguales a la población (Pons y Gordo, 2024). Informes recientes, como los publicados por la Fundació Ferrer i Guàrdia (*La brecha digital en España. Conocimiento clave para la promoción de la inclusión digital, 2023*) o, a nivel autonómico, por la consultora KPMG y la Generalitat de Catalunya (Acebo Pérez, 2022) o por el Consell Assessor del Parlament sobre Ciència i Tecnologia (Fernández-Ardèvol, Suárez-Gonzalo y Sáenz-Hernández, 2024) lo demuestran. En líneas generales, estos estudios muestran que **la desigualdad digital es, en definitiva, una forma más de desigualdad socioeconómica**, de carácter interseccional, que afecta especialmente a las personas mayores de 65 años (y de forma más intensa a las de 75 años o más), con bajos niveles educativos y de renta, con alguna discapacidad, que viven en zonas rurales y las mujeres.

La inteligencia artificial en un contexto de desigualdades

Antes de continuar con el resto de preguntas planteadas al inicio de este capítulo (quién, cómo, para qué -ha traído la inteligencia artificial- y a quién y cómo afecta), es imprescindible partir de la definición comprensible de un concepto central en este volumen que, no obstante, es especialmente complejo: *inteligencia artificial.*

La inteligencia artificial es **una disciplina de conocimiento** dedicada a generar agentes que, partiendo del procesamiento de conjuntos masivos de datos, producen resultados indistinguibles de los creados por las personas humanas, con un cierto nivel de autonomía y capacidad de adaptación a contextos o casos nuevos. Estas son las características de un sistema "artificialmente inteligente". ¿Pero cómo "aprenden" los sistemas de inteligencia artificial? El aprendizaje de máquinas (*machine learning*) es el paradigma actual de la inteligencia artificial: un conjunto de técnicas dedicadas a crear y entrenar sistemas basados en el procesamiento de enormes cantidades de datos sobre los "problemas" a los que se aplicarán y las "soluciones" a las que se pretende llegar. Hoy en día, la inteligencia artificial está más presente que nunca antes en el imaginario colectivo y, también, en las vidas cotidianas de las personas.

Pese a que, como tal, el término "inteligencia artificial" fue acuñado el año 1956 (ya con cierta polémica) los estudios en el campo ya llevaban años de desarrollo. Desde entonces y especialmente en los últimos años **los avances en el campo de la inteligencia artificial han sido muchos y notorios y sus usos se han multiplicado en una infinidad de sectores diversos**. Entre otros ejemplos, se han desarrollado sistemas dedicados a personalizar tratamientos médicos frente a enfermedades complejas, a tomar decisiones sobre quién debe o no recibir prestaciones sociales, a mejorar la calidad del transporte urbano, a prever el riesgo de reincidencia de personas con antecedentes criminales o a acelerar los procesos de contratación de personal. No obstante, no ha sido hasta la irrupción de los sistemas de inteligencia artificial "generativa" (una sub-área de la disciplina dedicada a la producción autónoma de textos, imágenes o contenidos en otros formatos como el audiovisual) como ChatGPT, a finales de 2022, cuando la mayoría de la población ha percibido que la inteligencia artificial ha entrado a formar parte de sus vidas cotidianas.

Estos avances, no obstante, no han afectado de la misma manera a las distintas capas de la población. En los últimos años se han multiplicado los estudios, las acciones y las denuncias sobre las consecuencias negativas y, especialmente, los efectos desiguales, en muchos casos discriminatorios, del desarrollo y el uso de los sistemas y aplicaciones de inteligencia artificial. Nada extraño en un contexto de fuertes desigualdades como el actual, que contribuye a que el desarrollo tecnológico, en este caso de la inteligencia artificial, en lugar de favorecer la mejora de las condiciones de vida del conjunto de la población, esté contribuyendo a profundizar los desequilibrios de poder, riqueza y bienestar (Eubanks, 2018).

Pero, más allá de la importancia del contexto en el que se produce, es imprescindible comprender **quién tiene la capacidad de influir en cómo y para qué se desarrolla y se usa la inteligencia artificial** que, fundamentalmente, son las grandes corporaciones tecnológicas. Actualmente, las grandes empresas del sector son las únicas capaces de desarrollar soluciones de inteligencia artificial y de introducirlas en un mercado que ya dominan, de una manera casi monopolística. Lo hacen gracias al control privativo que, desde hace años, ostentan sobre: por una parte, los datos necesarios para crear y entrenar los sistemas; y, por otra, los estándares técnicos que permiten la interoperabilidad entre sistemas y que fomentan el uso de determinados productos y servicios frente a otros; y, además, las infraestructuras físicas que soportan el funcionamiento de todo ello. Y su preeminencia se sostiene, además, gracias a un desequilibrio geopolítico general del poder y la riqueza (McChesney, 2013; Tufekci, 2017; Zuboff, 2019).

Consecuentemente, el desarrollo de la inteligencia artificial ha estado (y está) intensamente marcada por los intereses económicos de estas grandes empresas que marcan cómo son estas tecnologías, para qué se utilizan o quién tiene acceso a ellas, y determinan que todo ello se haga en unas condiciones de opacidad y secretismo que contribuyen al éxito de su negocio. Frente a ello, las instituciones públicas, las organizaciones de la sociedad civil y el conjunto de la ciudadanía se encuentran en una posición de debilidad, con importantes dificultades para influir en todo ello. Así, los principios democráticos que deberían guiar la transformación digital quedan relegados, las desigualdades se consolidan cada vez más, y **las personas y los grupos en situaciones más desfavorables son quienes, una vez más, se ven afectadas de forma más negativa** por una nueva forma de reproducción de los factores que provocan las brechas sociales.

Fruto de un entramado como este, en los últimos años, **han salido a la luz distintos sistemas de inteligencia artificial que han demostrado ser imprecisos, inútiles, indeseables o injustos**, por diversos motivos. Gran parte de la atención se ha puesto en el hecho grave de que, frecuentemente, se están utilizando para objetivos que perjudican al conjunto de la población o, de manera más intensa, a movimientos socio-políticos y a grupos o personas en situaciones desfavorables. La toma de decisiones basada en criterios arbitrarios e injustos, la limitación de oportunidades o la discriminación de, por ejemplo, las personas

migrantes o en situaciones laborales precarias, de pobreza o de privación de libertad, han estado en el punto de mira. Además, de manera más reciente, ha ganado terreno también la proyección pública de otros aspectos negativos, ligados al despliegue de la inteligencia artificial, como su impacto medioambiental, o los efectos en la percepción pública de los discursos interesados que lo acompañan. Las organizaciones de la sociedad civil, los movimientos en defensa de los derechos fundamentales, e investigaciones académicas, periodísticas e independientes, desempeñan un papel central en la denuncia y la toma de conciencia colectiva sobre todo ello. **La presente publicación representa una aportación ineludible a este respecto.**

Las raíces y los objetivos de esta publicación

El presente volumen colectivo se enmarca en la **tercera fase del proyecto Justicia Digital Global impulsado por Oxfam Intermón y la Dirección de Justicia Global del Ayuntamiento de Barcelona**[1] que se centra, de manera más específica que las fases anteriores, en analizar si el desarrollo de los sistemas de inteligencia artificial está o no agravando las desigualdades existentes o creando otras nuevas.

Publicado en el marco de la fase anterior, el informe *Desplazar los ejes: alternativas tecnológicas, derechos humanos y sociedad civil a principios del siglo XXI* (Calleja-López, Cancela y Cambronero, 2022), elaborado desde la unidad de Tecnopolítica del grupo de investigación Communication Networks and Social Change (Internet Interdisciplinary Institute, Universitat Oberta de Catalunya) apuntaba ya algunos de los principales problemas relacionados con el desarrollo y el uso de la inteligencia artificial. Entre ellos, las limitaciones derivadas del entrenamiento de estos sistemas; el hecho de que sea una tecnología propietaria, diseñada y controlada por grandes empresas tecnológicas; las promesas de dudosa fiabilidad que lanza de forma constante la industria; la propiedad de las infraestructuras físicas detrás de estos sistemas; las relaciones económicas y geopolíticas de poder de las que todo ello depende; las lógicas de opresión que producen y reproducen estas tecnologías; o la falta de transparencia.

Partiendo de este punto, **el objetivo del presente volumen colectivo es profundizar en cómo el desarrollo y los usos actuales de la inteligencia artificial se relacionan con las desigualdades**. La propuesta a las 13 personas que firman los distintos capítulos es tratar esta cuestión desde una perspectiva situada, decolonial y feminista, partiendo de preguntas como las siguientes: ¿cuáles son las implicaciones del avance de la inteligencia artificial en materia de (des)igualdades?, ¿contribuyen estas tecnologías al bienestar general y favorecen los intereses del conjunto de la población?, ¿privilegian intereses particulares?, ¿se ve comprometido el cumplimiento de algún derecho fundamental o el ejercicio de las libertades básicas de la ciudadanía?, ¿cuáles son las características de los contextos en los que podemos dar respuesta a las preguntas anteriores?

Perspectivas y aproximaciones diversas: una obra coral

La publicación es extensa. En ella se abordan en profundidad distintos aspectos fundamentales para entender la relación entre inteligencia artificial y desigualdades. El conjunto de todos ellos ofrece **una aproximación rigurosa que permite entender esta problemática desde un amplio abanico de ámbitos y casos de especial interés.** No obstante, no

1 Accesible en https://www.oxfamintermon.org/es/derechos-digitales-justos-igualitarios

es su objetivo recoger, de forma exhaustiva, todos los posibles temas relacionados, sino una selección oportuna y nutrida, de algunos de los más relevantes, de acuerdo con su relevancia social, su repercusión actual en la esfera pública, y también con la disponibilidad de las personas especializadas en el ámbito de estudio y/o involucradas en actividades relacionadas, que se encargan de la redacción de los diferentes capítulos.

El índice de contenidos es el resultado de un proceso riguroso, dividido en cuatro fases. En primer lugar, Oxfam Intermón elaboró una propuesta de aquellos temas considerados de interés de acuerdo con los trabajos realizados en las fases anteriores del proyecto Justicia Digital Global, y, más en general, con la experiencia de la organización en el estudio y el trabajo de campo relacionado con el tema de este volumen. Esta propuesta sirvió de base para dos sesiones de debate. La primera, celebrada el 4 de julio y conducida por Carlos Bajo en el marco de los encuentros del consejo asesor del proyecto de Justicia Digital Global, con la participación de Renata Ávila, Paz Peña, Cristina Colom, Paola Ricaurte, Eliana Quiroz, Andrea Costafreda, Laura Nathalie Hernández y Marta Peirano y apuntes posteriores de Liliana Arroyo. La segunda, el 12 de julio de 2024, guiada y moderada por Carlos Bajo en representación de Oxfam Intermón y Sara Suárez-Gonzalo (investigadora a cargo de este capítulo de introducción), con la participación de otras cinco personas especializadas en el tema: Andrea Rosales Climent (Universitat Oberta de Catalunya), Marta Galcerán Vercher (Barcelona Centre for International Affairs, CIDOB), Manuel Portela Charnejovsky (Universitat Pompeu Fabra), Núria Vallès Peris (Consejo Superior de Investigaciones Científicas) y Judith Membrives i Llorens (LaFede y AlgoRights). Tras integrar las aportaciones de esta sesión, Sara Suárez-Gonzalo entregó al equipo de Oxfam Intermón una nueva propuesta de temas y contenidos de interés. Más tarde, la organización adaptó el contenido de esas propuestas y de las sesiones al marco del *Multidimensional Inequality Framework*, hasta llegar a la versión definitiva que estructura el texto actual. Finalmente, el contenido de cada uno de los apartados se terminó de perfilar en conversaciones con las potenciales autoras identificadas.

Cabe destacar que, en tanto que obra coral, **cada capítulo se elabora a partir de la visión, los conocimientos y las experiencias particulares de cada una de las personas que los redactan y las organizaciones a las que pertenecen**. Esto quiere decir que la persona lectora no debe buscar una harmonía o una coherencia total a lo largo del volumen, especialmente por lo que respecta al uso de los términos, a la definición o la interpretación de los contextos en los que se enmarca el objeto de estudio, o a las implicaciones particulares de los elementos sobre los que en él se discuten. Precisamente, la polifonía que singulariza esta obra colectiva ha sido intencionadamente buscada por Oxfam Intermón con el objetivo de mostrar una pluralidad de voces suficiente que permita entender la complejidad de un fenómeno sobre el que todavía no hay consensos claros, dado que es altamente contextual y sus implicaciones todavía están emergiendo. Por otra parte, la cuidadosa y exhaustiva labor de coordinación y revisión por parte del equipo de la organización, garantiza la cohesión del texto y la complementariedad de sus distintas partes.

La estructura y los contenidos del volumen colectivo

Tras este primer **capítulo introductorio**, el volumen se organiza, fundamentalmente, en dos partes diferenciadas, seguidas por un apartado de recapitulación y conclusiones.

La **primera parte está dedicada a las cuestiones estructurales** y centra el contexto en el que se realiza esta publicación. Ofrece contribuciones imprescindibles para entender el impacto actual de la inteligencia artificial desde la perspectiva de las estructuras socio-políticas, culturales, económicas y las cuestiones medioambientales que la rodean.

El primero de los capítulos que la integran, a cargo de Sofia Scasserra (investigadora asociada al Transnational Institute) ofrece una reflexión sobre la reproducción de lógicas coloniales y extractivistas que a lo largo de la historia han marcado el desarrollo de la humanidad y que, hoy en día, adquieren una nueva dimensión neoextractivista relacionada con el despliegue de nuevas tecnologías de la información y de la comunicación como la inteligencia artificial. Se ocupa además, de forma más particular, de las desigualdades y los desequilibrios de poder que provoca el nuevo despliegue de estas lógicas entre los territorios del Norte y del Sur globales, así como entre aquellas regiones consideradas el centro o a las periferias sociales.

Ana Valdivia (profesora e investigadora sobre inteligencia artificial, gobierno y política en el Oxford Internet Institute de la University of Oxford) aborda un aspecto fundamental y, a menudo, poco tratado: el impacto medioambiental del despliegue de la infraestructura material que soporta el desarrollo y la implementación de los sistemas de inteligencia artificial. El aumento de emisiones contaminantes, la explotación de recursos naturales, y el consumo desproporcionado de energía y agua centran esta aportación que desmonta las promesas de que la inteligencia artificial será la solución en la lucha contra el cambio climático, propias del relato predominante promovido por el mercado.

La **segunda parte** del volumen se adentra, de manera más específica, en las implicaciones de este desarrollo tecnológico en diferentes ámbitos de la vida, la salud y el bienestar de las personas.

El capítulo a cargo de No Tech For Apartheid, revisa cómo la inteligencia artificial se relaciona con el mantenimiento (o la destrucción) de la vida y la salud humana en un nuevo contexto de guerras y conflictos armados. Las autoras aportan una explicación sobre cómo se está utilizando la inteligencia artificial para aumentar la eficiencia letal de las armas en guerras, especialmente la que devasta Gaza. El capítulo centra además el debate sobre el rol que desempeñan en todo ello las grandes empresas tecnológicas, movidas por intereses económicos, convirtiendo la guerra en un perverso laboratorio de pruebas tecnológicas.

Sarah Chander (experta en derechos fundamentales y política de igualdad de la Unión Europea, directora y cofundadora de Equinox – Racial Justice Initiative) continúa con un capítulo sobre la discriminación racial, étnica o por razón de origen derivada del uso de la inteligencia artificial. El apartado desgrana así, uno de los principales motivos de preocupación relacionados con el avance tecnológico en el ámbito de la vida individual, familiar y social. En él, Chander ofrece un retrato en profundidad de aquellos elementos de discriminación racial y por motivo del origen de personas en diferentes contextos, poniendo el acento en las derivadas que el racismo estructural presenta en el desarrollo de la inteligencia artificial.

El siguiente capítulo del volumen examina la influencia de la inteligencia artificial en la creación y distribución del conocimiento en la actualidad, así como su accesibilidad por parte del conjunto de la ciudadanía. En él, Pelonomi Moiloa (fundadora y directora de Lelapa AI) desgrana las implicaciones de la inteligencia artificial en la producción de conocimiento y examina la naturaleza de su posible influencia en los procesos de aprendizaje y, por tanto, en la forma en que la ciudadanía entiende el mundo. Por un lado, analiza la imposición de los valores del Norte Global, en relación con las circunstancias en las que se produce la tecnología. Por otro, estudia el efecto de homogeneización que tienen estas prácticas, así como el empobrecimiento que implica la generación de conocimiento a medida. Finalmente, insiste en la necesidad y el impacto de incorporar la diversidad lingüística en este proceso tecnológico.

Impulsado por Oxfam Mexico, el siguiente capítulo se ocupa de cuestiones relativas a la seguridad financiera y el trabajo digno en la llamada “nueva economía de plataforma” o *gig economy* (en inglés). El apartado recoge la experiencia de la organización en el estudio de

las particularidades del impacto de la inteligencia artificial en las condiciones de las personas que trabajan para las plataformas tecnológicas. Paola Villarreal (investigadora independiente) y Ervin Félix (investigador de Oxfam México) explican el papel que desempeñan los algoritmos en la imposición, la pérdida de la negociación y la gestión de estas condiciones, y cómo afectan a la vida de las personas trabajadoras.

Pablo Jiménez Arandia (periodista e investigador especializado en la interacción de la tecnología y la sociedad) analiza el uso de sistemas de decisiones automatizadas por parte de los servicios públicos y, en particular, en relación con el acceso a las prestaciones y a la protección social. En él cuestiona los discursos ligados a la racionalidad y a la mejora de la eficiencia de los servicios sociales, que suelen acompañar a la introducción en ellos de la inteligencia artificial. El autor pone de manifiesto que, en la práctica, se han evidenciado serios problemas derivados de los usos actuales de la tecnología en este ámbito, relacionados con la transparencia, con la gobernanza o con la posibilidad de auditar los sistemas algorítmicos. Hace notar, además, que se han demostrado importantes efectos indeseables, como una pérdida del acceso a servicios básicos por parte de personas que los necesitaban, o como la construcción de arquitecturas de gestión de esos servicios que se han demostrado como discriminatorias, se apoyan en la sospecha y afectan especialmente a los grupos sociales en situaciones más vulnerables.

El siguiente capítulo se centra en uno de los temas estrella que, desde escándalos como el de Cambridge Analytica durante las campañas de 2016 a favor de Donald Trump (en Estados Unidos) o del Bréxit (en Reino Unido), ha tenido un gran protagonismo en la esfera pública: la capacidad de las herramientas de inteligencia artificial para influir o condicionar el voto de la ciudadanía en las elecciones democráticas. Salvatore Romano y Thomas Wright (ambos de AI Forensics) relatan ejemplos de la influencia que una información política sesgada puede tener en el voto a partir de varias investigaciones realizadas por su organización. Se centran en el papel de chatbots y herramientas de IA generativa juegan en el futuro de la integridad democrática.

Cierra la segunda parte del informe el capítulo que se dedica a las intersecciones entre la discriminación y la violencia de género con la inteligencia artificial centrado en el contexto geográfico de Medio Oriente y el norte de África (MENA por sus siglas en inglés, Middle East and North Africa). En particular, Afef Abrougui (investigadora de Fair Tech) identifica los usos de estas tecnologías en los territorios de esta región, las estrategias que despliegan con ellas las grandes empresas tecnológicas y analiza cómo afecta todo ello al activismo y a las políticas desplegadas en materia de género.

Dada la naturaleza poliédrica del objeto de estudio del volumen, estos diez capítulos, incluida esta introducción, tratan temas inevitablemente interrelacionados, cuya interpretación conjunta y situada en el contexto actual es imprescindible. Por este motivo, el documento concluye con un apartado de recapitulación y discusión a cargo de Oxfam Intermón, que relaciona las principales aportaciones y ofrece un marco de interpretación. Partiendo de ello, señala algunas ideas que pueden ayudarnos a enmendar algunos de los problemas señalados y a construir un nuevo contexto para el desarrollo y la implementación de una tecnología que, en definitiva, debe contribuir a avanzar y favorecer las necesidades y los intereses del conjunto de la población global, y no los de unos pocos.

Referencias

Acebo Pérez, Mónica (2022). *Estudi sobre la bretxa digital*. Generalitat de Catalunya. Departament d'Empresa i Treball y KPMG. Disponible en: https://politiquesdigitals.gencat.cat/web/.content/00-arbre/ciutadania/Pla-de-xoc-contra-la-bretxa-digital/Estudi-sobre-la-bretxa-digital-a-Catalunya_DEF.pdf.

Ayala Cañón, Luis; Laparra Navarro, Miguel; Rodrígez Cabrero, Gregorio (coords.) (2022). *Evolución de la cohesion social y consecuencias de la Covid-19 en España*. Madrid: Fundación FOESSA y Cáritas Española Editores. Disponible en: https://www.caritas.es/main-files/uploads/sites/31/2022/01/Informe-FOESSA-2022.pdf.

Calleja-López, Antonio; Cancela, Ekaitz; Cambronero, Marta (2022). *Desplazar los ejes: alternativas tecnológicas, derechos humanos y sociedad civil a principios del siglo XXI,* Oxfam Intermón. Disponible en: https://www.oxfamintermon.org/es/publicacion/desplazar-los-ejes-xxi-oxfam.

Eubanks, Virginia (2018). *Automating Inequality. How high-tech tools profile, police, and punish the poor*. Nueva York: St. Martin's Press.

Fernández-Ardèvol, Mireia; Suárez-Gonzalo, Sara; Sáenz-Hernández, Isabel (2024). *Desigualtat digital i vellesa: la bretxa digital que encara cal tancar*, Consell Assessor del Parlament sobre Ciència i Tecnologia. Disponible en: https://www.parlament.cat/document/intrade/406609846.

Fortune. «*Global 500*» [Datos de 2023], 2023. Disponible en: https://fortune.com/ranking/global500/search/.

Fundació Ferrer i Guàrdia. (2023). *La brecha digital en España*. *Conocimiento clave para la promoción de la inclusión digital*. Disponible en: https://www.ferrerguardia.org/blog/publicaciones-3/encuesta-brecha-digital-en-espana-2022-73

García López, Ernesto (2024). "La percepción social de la desigualdad en España: una aproximación", En: Salido Cortés, Olga; Ruiz-Huerta Carbonell, Jesús (2024). *VI Informe sobre la Desigualdad en España 2024*, Fundación Alternativas. Disponible en: https://fundacionalternativas.org/publicaciones/vi-informe-sobre-la-desigualdad-en-espana/.

McChesney, Robert W. (2013). *Digital Disconnect. How Capitalism is Turning the Internet Against Democracy*, Nueva York: The New Press.

Murphy, Andrea; Schifrin, Matt (2024). *The Global 2000. 2024*. Forbes. Disponible a: https://www.forbes.com/lists/global2000/.

Oxfam Intermón (s.f.). *Multidimensional Inequality Framework*. Disponible en: https://inequalitytoolkit.org/.

Pons, Aleix; Gordo, Ignacio (2024). "Los posibles efectos sobre la desigualdad territorial de la transición digital". En: Salido Cortés, Olga; Ruiz-Huerta Carbonell, Jesús (2024). *VI Informe sobre la Desigualdad en España 2024*, Fundación Alternativas. Disponible en: https://fundacionalternativas.org/publicaciones/vi-informe-sobre-la-desigualdad-en-espana/.

Riddell, Rebecca; Ahmed, Nabil; Maitland, Alex; Lawson, Max; Taneja, Anjela (2024). *Desigualdad S.A. El poder empresarial y la fractura global: la urgencia de una acción pública transformadora*. Oxfam Intermón. Disponible en: https://lac.oxfam.org/en/latest/publications/desigualdad-sa.

Statista (2024). *Ranking de las empresas líderes en el mundo en 2024, por valor de marca (en millones de dólares)*. Disponible a: https://es.statista.com/estadisticas/539088/ranking-de-las-25-principales-marcas-en-el-mundo-por-valor-de-marca/.

Tufekci, Zeynep (2017). *Twitter and the tear gas. The power and fragility of networked protest*, New Haven: Yale University Press.

Zuboff, Shoshana (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Nueva York: Public Affairs.

2. Código colonizado: desentrañando las desigualdades económicas en la era de la IA

Sofía Scasserra

En otras épocas de la historia, la división internacional del trabajo ha reservado a los países de la periferia o del Sur Global el papel de proveedores de materias primas. Históricamente, los países de América Latina y de otras regiones encuadradas en la misma categoría geopolítica, han sido los encargados de aportar los recursos básicos que han sido el motor del crecimiento de la economía mundial: minerales, alimentos, y productos de los más variados con bajo valor agregado. Siguiendo esta lógica de división internacional del trabajo, esos productos se procesaban en el exterior y regresaban como bienes terminados, provocando un desequilibrio estructural en la balanza de pagos que ha marcado profundamente la distribución global del poder. La digitalización y, más recientemente, el despliegue de la inteligencia artificial no ha cambiado esta estructura, es más, se ha apoyado en ella, adaptándola para propiciar un modelo de desarrollo desigual que sigue los mismos parámetros.

Las cadenas de producción de la inteligencia artificial continúan alimentándose de recursos de los países periféricos, esta vez incorporando la extracción de una materia prima muy particular, los datos; además de la mano de obra cualificada. Sin embargo, preservan muy bien que los beneficios se mantengan en unas pocas manos a través de las capacidades de producción que se concentran en el Norte Global y que condicionan las posibilidades de desarrollos en este ámbito, pero también de distribución de los rendimientos.

¿Qué es y qué no es la inteligencia artificial?

Escuchamos a menudo que la culpa no es de la herramienta sino del que la utiliza[1]. Noticias falsas, videos e imágenes no consentidas, discriminación, xenofobia o exclusión son solo algunos de los problemas que se evidencian con la utilización de la inteligencia artificial (IA). Pero, ¿esos son los problemas realmente? ¿o esos son sólo síntomas de problemas más profundos enraizados desde la génesis de estas herramientas?

El minimizar la IA al estado de una mera herramienta tiene un trasfondo político ineludible. Es comparar una tecnología poderosa, que ha cambiado (y sigue transformando) empleos, educación, salud, acceso a la información o la democracia, por mencionar algunas esferas, a una simple botella de agua, una silla o un tenedor, por mencionar alguna que otra tecnología que utilizamos a diario.

La tecnología en general, y las tecnologías digitales en particular, tuvieron y tienen una capacidad transformadora en la sociedad. Toda tecnología tiene impreso en su diseño y producción la intención y el objetivo del fabricante o el diseñador de la misma. Esa intención y ese objetivo puede ser altruista, puede ser el de satisfacer una necesidad puntual, o puede ser, dado que vivimos en un sistema capitalista, el de extraer el máximo valor posible. Y si la fabricación de IA persigue el afán de lucro, puede no mirar las consecuencias sociales que tiene, solo busca "vender". Así, "El automatismo y su utilización bajo la forma de organización industrial denominada automation, posee una significación económica o social más que una significación técnica" (Simondon, *El modo de existencia de los objetos* técnicos, 2007, pág. 33). Gilbert Simondon en esta simple frase nos dice que el proceso de automatización (en este caso expresado a través de la IA) conlleva un problema intrínsecamente social. Dado que toda herramienta que automatiza conlleva en su diseño y fabricación una idea de mundo, de persona, nos determina como deberían ser las cosas de acuerdo a esa idea. Por ende, toda herramienta automatizada, más que una herramienta, es un espacio de disputa política. En efecto, si la IA trae problemas sociales, y esos problemas son principalmente porque el fabricante que tiene el poder de producir la tecnología lo hace de determinada manera y con determinadas intenciones, regular esa producción y disputar ese poder económico, es la nueva lucha colectiva del siglo XXI.

En definitiva, interpretar la IA dentro de las luchas sociales, nos ayuda a comprender cómo disputar ese espacio para que el futuro sea con equidad, igualdad, y justicia social.
Este texto no pretende abordar las luchas de género, medioambientales o laborales, entre

1 "La inteligencia artificial no es peligrosa; nosotros, sí". *La Nación*, 3 de enero de 2015. https://www.lanacion.com.ar/tecnologia/la-inteligencia-artificial-no-es-peligrosa-nosotros-si-nid1757065/

otras muchas que se están dando al cuestionar la IA. El mismo trata de ser un pequeño aporte a una lucha histórica de los pueblos del Sur Global que ya lleva siglos: entender la IA como una nueva forma de extractivismo y colonialismo tecnológico que lleva a renovar esfuerzos para no transformarnos en un nuevo Potosí.

¿Dónde se encuadra la periferia?

Históricamente, los países periféricos han tenido un rol definido en la división internacional del trabajo. En efecto, la economía global se ha organizado asumiendo que las ventajas comparativas son variables relativamente estáticas y designadas de acuerdo a la dotación de recursos que cada economía tiene en stock al momento de definir sus límites. Así, los países periféricos en general, y América Latina en particular, han sido proveedores de materias primas que han sido el motor del crecimiento de la economía mundial: minerales, alimentos, y productos de los más variados con bajo valor agregado. Estos productos se procesan en el exterior, y reingresan en nuestras economías como bienes terminados, generando una crisis en la balanza de pagos estructural que ha pesado en desequilibrios económicos desde antaño.

La teoría de la dependencia económica y el nombre de Raul Prebisch[2], economista argentino que describió estos procesos y dilucidó estrategias para salir de la trampa del debilitamiento de los términos de intercambio, han sido de gran renombre de forma histórica en espacios como la UNCTAD, donde se investiga el desarrollo de las naciones periféricas, buscando respuestas a estas dinámicas estructurales que parecen inamovibles.

El asombro por la tecnología deslumbra. Siempre lo hizo. El asombro nos atrapa ante herramientas nuevas que son más parecidas a la magia que a la realidad. Sucedió ya con la electricidad, con las telecomunicaciones, y con tantos inventos que marcaron épocas en la historia de la humanidad. La tecnología nos sorprende y nos supera constantemente. Nadie inventa un objeto para que realice una actividad peor de lo que la hacen las personas. ¿Quién se pondría un par de anteojos para ver peor? Toda tecnología nos mejora a niveles inimaginados, y por ende, sorprende. Ese asombro nos lleva a pensar que otro mundo es posible, que las cosas pueden cambiar. Y nos imaginamos vendiendo tecnología en un mundo integrado, siendo capaces de responder a demandas nuevas y dinámicas que resuelvan por fin la crisis de balanza de pagos y nos posicione a nivel global.

Pero, ¿es eso lo que está ocurriendo? Veamos un poco.

La producción de IA: artesanal vs. industrial

Si pensamos en una industria, vemos un proceso donde se obtiene materia prima, que normalmente es heterogénea en su composición. La misma se lleva a una fábrica, y atraviesa diversos procesos, se modifica, se le agrega valor y se la transforma en un producto vendible en el mercado. Luego se realiza control de calidad al producto. El mismo debe ser homogéneo y cumplir con determinado estándar de calidad. Más tarde se sale al mercado para la venta masiva del producto, se coloca en diversos mercados y se busca que el producto se posicione por sus prestaciones, calidad y bajo costo. A grandes rasgos, podríamos asegurar que esto que acabo de describir aplica a casi cualquier proceso industrial o producto industrializado. Cuando los países del Norte Global se industrializaron, pensaban en dominar este proceso a lo largo y ancho de la economía.

2 Un perfil de Raul Prebisch se puede encontrar aquí: https://unctad.org/osg/former-secretaries-general-and-officers-charge/raul-prebisch

Si pensamos en la IA por un momento, este proceso también aplica: Se extraen datos, millones de ellos, diversos y heterogéneos, y hasta de mala calidad. Comienza el proceso de industrialización: esos datos se seleccionan, se mejoran mediante la moderación de contenido y el procesamiento en lo que a mí me gusta llamar la "fábrica algorítmica": un entramado de algoritmos que procesan dichos datos y los transforman en información valiosa. Luego se realiza control de calidad de dicha información a través de moderación de contenido y testeos en el mercado digital: ¿es efectivamente la información que proveyó el chat GPT lo que el consumidor está buscando? Cuando respondemos esa pregunta al hacer consultas, estamos colaborando con ese proceso de control de calidad. Finalmente se vende en el mercado masivo un producto estandarizado y homogéneo, a un precio accesible, posicionándolo en el mercado global.

Esto que acabo de describir, es lo que me gusta llamar "IA industrial", o las industrias digitales de gran escala. La IA generativa recientemente lanzada al mercado busca dominar la IA industrial, ofreciendo una plataforma sobre la cual montar otras herramientas basadas en IA de menor escala. El acceso a su capacidad de producción es casi imposible para países periféricos, ya que no solamente se precisa cantidades realmente exorbitantes de datos, sino que también el costo que posee entrenar estos modelos en términos de energía y hardware, es inalcanzable. Se estima que, por ejemplo, Chat GPT consume más de medio millón de kilovatios-hora (kWh) de electricidad al día, mientras que el hogar medio de España consume unos 9 kWh diarios[3]. A su vez, también tienen un costo en términos de inversión en *hardware*. El costo de una GPU A100 de Nvidia se estima entre 10 mil y 15 mil dólares. Teniendo en cuenta que Chat GPT utilizó 25 mil de estas GPU, el costo de adquisición del hardware para entrenar al modelo superó los 250 millones de dólares[4].

Ahora bien, hoy día vemos soluciones basadas en IA, cuya base se sustenta en estos grandes modelos. Es decir, han surgido industrias digitales menores, en mercados de nicho (IA para estudiantes, IA para periodistas, IA para diseño, etc) pero todas utilizan uno de estos grandes modelos como herramienta de diseño de fondo. Copilot, Chat GPT, Gemini, Claude, y algunos otros buscan dominar el mercado de grandes modelos industrializados diciéndole al resto del mundo "este mercado ya está ocupado, ustedes vean cómo pueden usar nuestro modelo para construir algo más de nicho".

Así surgen herramientas que ocupan mercados menores, pero igualmente importantes. Empresas del Norte Global (Europa, EE.UU., y economías industrializadas) que logran armar modelos que sirven a colectivos enteros (medicina, educación, etc). ¿Y la periferia? A la periferia nos dejan lo que yo llamo "IA artesanal": soluciones basadas en IA para empresas puntuales y necesidades puntuales en nuestros territorios, cuando no las diseñan empresas del mismo orte Global. Si miramos la producción e investigación en IA, lo números muestran esta realidad: mientras que EE.UU. tiene el 60% de las nuevas inversiones en IA, países como China, Gran Bretaña, Israel, Canadá, Francia, Japón o Alemania le siguen[5], donde América latina no logra llegar a niveles que merezcan rastrear estadísticas. De hecho, se ha lanzado en un esfuerzo mancomunado de diversas organizaciones, un índice de adopción de IA para la región[6] que trata de mostrar cómo la IA se va utilizando en la región y cómo lograr mejorar los ecosistemas nacientes pero poco se habla de producción a gran escala. Si miramos las patentes inscritas en OMPI[7] (Organización Mundial de la Propiedad Intelectual), la situación es parecida con una leve diferencia: en este caso China lidera las inversiones a nivel global, seguidos por los mismos países ya mencionados más India.

3 "ChatGPT consume 55.000 veces más electricidad que la media de los hogares de España". *Business Insider*, 11 de marzo de 2024. https://www.businessinsider.es/chatgpt-consume-55000-veces-electricidad-hogar-medio-1371358

4 *Diez preguntas frecuentes y urgentes sobre Inteligencia Artificial.* Fundación Sadosky, 2024. https://program.ar/wp-content/uploads/2024/08/Diez-preguntas-frecuentes-y-urgentes-sobre-Inteligencia-Artificial.pdf

5 Top 10 Countries Leading in AI Research & Technology in 2025. *Techopedia*, 2024. https://www.techopedia.com/top-10-countries-leading-in-ai-research-technology

6 Espacio web del Índice Latinoamericano de Inteligencia Artificial (ILIA): https://indicelatam.cl/

7 *Patent Landscape Report - Generative Artificial Intelligence (GenAI).* World Intellectual Property Organization (WIPO), 2024. https://www.wipo.int/web-publications/patent-landscape-report-generative-artificial-intelligence-genai/en/key-findings-and-insights.html

Es decir, el mercado de producción de IA está dominado por grandes empresas y países que realizan las investigaciones para el desarrollo de IA y los modelos a gran escala de base, sobre los que luego hay que montarse para diseñar soluciones menores, que pueden ser vendidas en mercados más pequeños. El Norte dominando la industria, el Sur viendo cómo logra insertarse. Nada nuevo bajo el sol.

Pero, ¿es ese el único rol que jugamos en la periferia? Definitivamente no.

Extractivismo y colonialismo de datos

Al desmembrar la cadena de valor de la IA, como hicimos anteriormente, tenemos en primer lugar, la extracción de materias primas: los datos. Luego, su mejora y procesamiento para finalmente, llegar a su control de calidad y venta masiva. Veamos qué sucede en cada una de estas etapas y cómo están situados los pueblos periféricos.

Los datos

Al comienzo del milenio ya había teléfonos celulares en las manos de la gente, pero estos teléfonos no eran inteligentes ni poseían la variedad de usos y aplicaciones que vemos hoy día. Los teléfonos inteligentes aparecieron en la segunda mitad de la década del 2000 y ahí comenzó el extractivismo de datos. Sí, fue extractivismo a escala global. Se extrajeron datos, se exportaron a los grandes centros de producción tecnológica de forma masiva, sin permisos, habilitaciones o contemplación del efecto que tendría en la sociedad. Y a cambio de ese extractivismo no quedó nada para aquellos que produjeron dichos datos, ni siquiera un ingreso fiscal, dado que las empresas tecnológicas responsables de dicho extractivismo están situadas en otros países, y los datos no pagan impuestos al atravesar las fronteras.

Esto es así porque en el año 1998, mucho antes de que pensáramos que un dato podía tener valor, en la Organización Mundial de Comercio (OMC) se aprobó algo que tiene un nombre muy confuso, pero que tiene consecuencias muy reales: la moratoria en impuestos aduaneros a las transmisiones electrónicas[8]. Dicho de otra forma, se pueden extraer datos, exportarlos a otras latitudes, y no pagar impuestos en fronteras por los mismos. Esta moratoria se ha renovado sin interrupción cada dos años desde que se impuso. En la última conferencia Ministerial de la OMC en Abu Dhabi en febrero del 2024, 8 países levantaron su voz pidiendo revisar este principio. Se hace cada vez más evidente lo injusto de esta norma.

Así, los datos son extraídos, pero no solo del Sur Global, sino de la totalidad de la población conectada a Internet a escala mundial, sin pagar siquiera un impuesto que, de alguna manera, regrese a la población de la que fueron extraídos los datos algún tipo de ganancia.

La mejora de los datos para entrenar sistemas

Ahora bien, esa materia prima hay que refinarla, y ahí sí comienza la fase más despiadada con el Sur Global. No alcanza con el extractivismo transnacional, sino que además se segmenta y profundiza la división internacional del trabajo en la producción tecnológica en favor de las naciones más poderosas.

8 *Declaración sobre el comercio electrónico mundial.* OMC, 1998. https://www.wto.org/spanish/tratop_s/ecom_s/mindec1_s.htm

Una parte importante del proceso de creación de la IA es la de seleccionar información, "limpiarla" y entrenar sistemas. Esta función se conoce como "moderación de contenido", cuyas condiciones de trabajo son paupérrimas, situadas en lugares estratégicos dentro del globo en países periféricos. Salarios bajos[9], nada de protección social, contratos de hora cero[10] y serios problemas de salud mental[11] producto de la hiperproductividad y del contenido que se ven forzados a mirar, son sólo algunas de las terribles consecuencias de estos trabajos que sirven a la industria de IA para mejorar la calidad de la materia prima que se utiliza y entrenar sistemas de forma más eficiente.

Este tema está cobrando cada vez más visibilidad en el mundo tecnológico y se comienzan a acumular las demandas a las empresas[12]. En Kenia ya existe un primer sindicato de moderadores de contenido y la resistencia empieza a expandirse. Más allá de los esfuerzos por cambiar la realidad de cientos de miles de trabajadores en el mundo, lo cierto es que las empresas tecnológicas en su estrategia de división global del trabajo parecen habernos asignado tareas que jamás se harían en el Norte global sin condiciones de trabajo dignas, seguro social y protección de salud mental, entre otras cosas. La idea de que la IA se produce en Silicon Valley con oficinas modernas y condiciones de trabajo ideales, olvida esta parte de la cadena productiva basada en la explotación y el extractivismo.

El procesamiento de los datos

El procesamiento es donde más se agrega valor en el proceso de fabricación industrial de la IA. Definitivamente, tener programadores e ingenieros trabajando en transformar esos datos en información útil y vendible, que diseñe servicios de los más variados, desde un chat de IA generativa hasta un algoritmo que sugiere música, pasando por un sistema que detecte enfermedades a partir de imágenes. Tener el conocimiento para poder realizar esa tarea y lograrlo es agregar valor de forma indiscutida.

Ahora bien, ¿dónde se realizan esas tareas?, ¿quiénes toman la decisión?, y, sobre todo, ¿quiénes son dueños de ese diseño y, por ende, se quedan con la propiedad intelectual del servicio que diseñaron? La respuesta puede parecer obvia, pero tiene matices.

Si bien es cierto que existe Silicon Valley y sus *megaempresas* llenas de trabajadores altamente cualificados, no son éstos los únicos que están al servicio de las *big tech*. El Sur Global provee mano de obra a costos mucho más bajos y por proyecto, no contratados en relación permanente. Es común que países como la India, Argentina o Brasil, tengan una enorme cantidad de trabajadores, ingenieros que han estudiado en sus propias universidades (en muchos casos públicas, solventadas por el Estado como el caso de Argentina y Brasil) que trabajan agregando valor a los productos de las *big tech* a salarios más bajos que aquellos que pagan en EE.UU. Esto es conveniente para ese trabajador que cobra salarios dolarizados en países de monedas más inestables, pero, ¿es bueno para el país?

9 "Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic". *Time*, 18 de enero de 2023. https://time.com/6247678/openai-chatgpt-kenya-workers/

10 "Subterranean Moderators". *Pulitzer Center*, 5 de junio de 2024. https://pulitzercenter.org/stories/subterranean-moderators

11 "Mental trauma: African content moderators push Big Tech on rights". *The Economic Times*, 16 de octubre de 2023. https://economictimes.indiatimes.com/tech/technology/mental-trauma-african-content-moderators-push-big-tech-on-rights/articleshow/104457622.cms

12 "The occupational hazards of cleaning the internet". *Coda*, 28 de febrero de 2023. https://www.codastory.com/authoritarian-tech/reddit-content-moderation-lawsuit/

En el caso de Argentina, por ejemplo, se pueden ver tres problemas:

1. Por un lado, ese trabajador agrega valor a una industria que no es nacional, sino que es foránea y que luego reimporta ese valor agregado al país. Es decir, la vieja historia de la dependencia económica que nos mostraba Raúl Prebisch de vender materias primas para que reingrese enlatado otro producto con alto valor agregado, debilitando los términos de intercambio, parece repetirse en el circuito tecnológico. Se vende línea de código u hora de programador, para que reingrese eso enlatado en un software, en una aplicación, en una nueva herramienta de IA.

2. Por otro lado, ese programador ejerce presión en los salarios en el mercado doméstico[13]. Es común ver empresas nacionales desesperadas buscando talento y trabajadores dispuestos a trabajar por salarios en pesos (dado que es lo que pueden y deben pagar) compitiendo con un mercado gigante que demanda mano de obra a salario dolarizados. La fuga de cerebros hace que sea muy difícil para las pymes tecnológicas locales poder competir[14], complicando el desarrollo tecnológico nacional.

3. Finalmente, estos trabajadores muchas veces buscan cobrar sus salarios dolarizados fuera del país, dado que si los ingresan, se los trasladan a moneda local, evadiendo al fisco. Muchos cobran por el sistema de criptomonedas, siendo muy difícil encontrarlos[15].

   Es decir, la industria de agregar valor para la producción de IA no parece hacer la diferencia estructural que se necesita para saltar al desarrollo, tan ansiado y necesario. El punto clave es que la propiedad intelectual de esos desarrollos es de las empresas norteamericanas, siendo los trabajadores meros ejecutores de proyectos. Una vez más se hace extractivismo de recursos del Sur Global sin dejar valor ni transferencia tecnológica que nos ayude a crear nuestro propio desarrollo. Esta historia relatada desde la experiencia argentina se repite con matices en países como Chile, Costa Rica y Brasil, entre otros.

América latina y la lógica de bienes comunes

El lugar que ocupa la región en la producción y desarrollo de IA no parece ser protagónico. Se nos ha relegado a ser, como antaño, proveedores de recursos para que otros desarrollen la tecnología, se apropien del valor, y expandan sus formas de sentir y pensar en el mundo. En efecto, la IA es una herramienta que es fuertemente influenciada por las creencias y los preconceptos que se tuvieron no solo al recabar los datos, sino también al programarla.

Existen numerosos ejemplos de cómo la IA discrimina[16], es sexista, racista, y, sobre todo, no tiene en cuenta al Sur Global. De más está decir que la mayoría de los servicios ofrecidos funcionan mejor en inglés[17] que en cualquier otro idioma y, si bien en español funcionan bastante bien, en las lenguas menos habladas como el guaraní o el quechua, directamente no funcionan como deben.

13 "¿Cuánto gana un programador en Argentina?". *BAE Negocios, 12 de septiembre de 2024.* https://www.baenegocios.com/sociedad/Cuanto-gana-un-programador-en-Argentina-20240912-0058.html

14 *Nuevas dinámicas de comportamiento en el sector de software y servicios informáticos.* Centro Interdisciplinario de Estudios de Ciencia, Tecnología e Innovación (CIECTI), 2023. https://www.ciecti.org.ar/wp-content/uploads/2023/09/RI-Deslocalizacion-RI3.pdf

15 "Cobrar el sueldo en Bitcoins, una alternativa que gana espacio en 2022". *Forbes Argentina*, 5 de enero de 2022. https://www.forbesargentina.com/innovacion/cobrar-sueldo-bitcoins-una-alternativa-gana-espacio-2022-n11523

16 Espacio web de la Algorithmic Justice League, una organización basada en EE.UU. que visibiliza la discriminación, especialmente, por motivos de raza, etnia o lugar de origen en las herramientas algorítmicas: https://www.ajl.org/

17 "AI's language gap". *Axios*, 2023. https://www.axios.com/2023/09/08/ai-language-gap-chatgpt

Ejemplos de tecnologías que no fueron pensadas localmente hay miles. Uno que me gusta mencionar son los *emojis* de Whatsapp. Para el cono sur el mate es una bebida trascendental. Invitar a tomar mates es algo cotidiano, símbolo de amistad, amor, camaradería. Su *emoji* no existía y tuvieron que presentar varias decenas de páginas[18] justificando que se agregue al listado de Unicode. Si eso sucede con cosas tan visibles como esas, imaginémonos, por un momento, lo que ocurre en la caja negra de los sistemas. *Softwares* que no tienen en cuenta costumbres locales, sistemas de detección de enfermedades que fueron hechos con datos de otras poblaciones con dietas, horarios y climas muy distintosa los de las poblaciones donde son utilizados, son solo algunos de los ejemplos que podemos encontrar.

En este sentido América Latina tiene mucho para aportar si se dieran las condiciones para que sus lógicas, sus costumbres y sus luchas se plasmaran en los sistemas que son utilizados a diario.

La región está atravesada por el debate de los bienes comunes en torno a sus recursos naturales. ¿Se puede ser tan ignorante como para reclamar propiedad sobre un río y contaminarlo cuando es de todos y todos nos beneficiamos del agua dulce que corre por su cauce? Lo cierto es que los bienes comunes merecen una protección especial. Son de todos y valen para todos, y todos debemos cuidarlos. Los datos son bienes comunes. Para empezar, son bienes no rivales[19]. Es decir, que al consumirlos no se agotan. El arte por ejemplo es un bien no rival: que una persona mire un cuadro o escuche una canción, no evita que otra persona pueda hacer lo mismo. Lo consumo, pero no se agota. Es el caso de la educación, la seguridad, el transporte público, todos servicios públicos que no deben ser privatizados. Los bienes comunes van un paso más allá: no son siquiera del Estado, pero éste sí debe protegerlos por el bien de todos.

“Tú no puedes comprar el viento, Tú no puedes comprar al sol, Tú no puedes comprar la lluvia, Tú no puedes comprar el calor” dice el poeta en su canción dedicada a Latinoamérica[20]. Probablemente “Tú no puedes comprar mis relaciones interpersonales, mis gustos, mis deseos, mis saberes” debe ser la lucha colectiva que debemos escribir.

Si no se permite este debate, nos estaremos perdiendo de poder usar los datos y la información que pueden producir, en favor de los pueblos, buscando crear tecnologías no rivales que beneficien a toda la sociedad.

¿Quién hace las reglas?

Estas lógicas de dependencia han sido sostenidas por décadas en acuerdos legales que conforman la arquitectura de un sistema internacional diseñado para que gane el más poderoso y evitar que los países periféricos accedan al tan ansiado desarrollo económico, pudiendo el Norte abastecerse de forma continua de sus recursos con un extractivismo obsceno.

Esto sucede con todos los acuerdos que regulan comercio de bienes y servicios, inversiones, y flujos financieros, entre otros.

Pero con las cuestiones digitales no existían reglas. Y fue tal el descalabro que generó la IA que se hizo evidente la necesidad de gobernanza y regulación. En efecto, no podemos estar todos sujetos a los caprichos de diseño de EE.UU. o China, se necesitan reglas que pongan las cosas en su lugar y obliguen a diseñar la tecnología en concordancia con las

18 “Argentina gana la lucha por el emoji del mate, el primer ícono sudamericano”. *El País* / *Verne*, 12 de julio de 2019. https://verne.elpais.com/verne/2019/07/12/mexico/1562944754_939569.html

19 Explicación sencilla del concepto de “rivalidad” en economía, a través de Wikipedia: https://es.wikipedia.org/wiki/Rivalidad_(econom%C3%ADa)

20 *Latinoamérica*. Calle 13, 2011. [Vídeo musical en Youtube]. https://www.youtube.com/watch?v=DkFJE8ZdeG8

regulaciones existentes. El debate es urgente. Sin embargo, la intención de regular y dar gobernanza global parece no haberse materializado en una propuesta que lleve a esa dirección.

Pensando la gobernanza basándose en el libre comercio

El primer intento de gobernanza tecnológica que existió en el mundo fue diseñado[21] por los abogados de las grandes corporaciones tecnológicas norteamericanas. No parece ser la mejor idea poner a los zorros a vigilar el gallinero. Este intento no es más que el acuerdo de libre comercio electrónico que apareció por primera vez a la luz en la negociación del Trade in Services Agreement (TISA)[22], y luego se replicó en el Tratado TransPacífico (TPP)[23] y hoy por hoy está presente en el Acuerdo de Declaración Conjunta de la OMC[24] y otros tratados de libre comercio bilaterales o subregionales. Este acuerdo lo que hace es liberalizar de una vez y para siempre (dado su carácter supra nacional) lo que ocurre en la esfera digital.

El texto (que se va reproduciendo de acuerdo en acuerdo con algunos matices), liberaliza el traspaso transfronterizo de datos, no permitiendo a los Estados regular ni su transferencia, ni su almacenaje, ni su procesamiento, ni su localización. Es decir, confirma y ancla el saqueo al infinito, prohibiendo el cobro de impuestos en frontera. Lo que han hecho hasta el día de hoy, pero fijado por ley internacional de modo tal que ningún Estado pueda cambiarlo.

A su vez, prohíbe la apertura o transferencia del código fuente de un algoritmo, complicando su auditoría, pero también, sobre todo, bloqueando cualquier posibilidad de que un Estado exija transferencia tecnológica como condición de acceso a mercados, herramienta legal que se puede usar, sobre todo, para combatir el desastre medioambiental que sufrimos y hacer efectiva una transición justa a tecnologías más sustentables. En Sudáfrica, por ejemplo, debido a la crisis de agua[25], se necesita de forma desesperada la tintura de telas sin utilización de agua. Dicha tecnología no logra esparcirse en la industria debido a regulaciones de propiedad intelectual que impiden la transferencia tecnológica. Otro caso a resaltar es el avance de las regulaciones y estándares medioambientales y cómo operan como barreras al comercio. En efecto, es cada vez más difícil para los países en vías de desarrollo exportar cumpliendo con los estándares internacionales en temas medioambientales exigidos por los países centrales. Este es un debate sumamente vigente en la OMC donde en este año 2024 se siguen discutiendo los CBAM[26].

El acuerdo de comercio electrónico a su vez, exime de responsabilidad a las plataformas digitales por los contenidos que publican. Esto no sólo trae problemas al fomentar la difusión de noticias falsas, sino que además promueven contenido violento en Internet. Los moderadores de contenido para entrenar sistemas y limpiar información se ven forzados a contemplar dichas imágenes, exacerbando las condiciones precarias de trabajo de dichos trabajadores que sufren problemas de salud mental, como ya se ha mencionado anteriormente.

21 *EU Digital Trade Rules: Undermining attempts to rein in Big Tech.* The Left in the European Parliament, 2023. https://left.eu/app/uploads/2023/03/Summary-Digital-Trade-EN.pdf

22 El Trade in Services Agreement (TISA) fue un acuerdo de libre comercio de servicios que se comenzó a negociar a mediados de la década del 2010-2020. El acuerdo se negociaba en secreto absoluto entre 23 países. Sus textos fueron filtrados por el portal Wikileaks. El acuerdo finalmente no llegó a terminarse y se diluyó luego que Donald Trump llegara a la primera presidencia de los EE.UU.

23 Este acuerdo se negoció también en la misma década que el TISA. A diferencia del primero, este se concluyó aunque sin incluir a los EE.UU., que había sido su principal impulsor. Hoy lo integran 12 países de la región del pacífico.

24 El acuerdo de Declaración Conjunta de Comercio Electrónico de la OMC, es un acuerdo de libre comercio digital que se encuentra en proceso de negociación. El acuerdo no está publicado y pueden leerse filtraciones en la página web https://www.bilaterals.org/?wto-electronic-commerce-agreement&lang=en.

25 *Why does the Cape Town water crisis impact the textile industry?* O Ecotextiles, 2018. https://oecotextiles.blog/2018/03/07/why-does-the-cape-town-water-crisis-impact-the-textile-industry/

26 Este es un mecanismo propuesta por la Unión Europea (UE) para algunos productos industriales, argumentando que no pueden ingresar a la UE productos que hayan tenido determinada cantidad de emisiones de carbono al ser producidas. Esto deja fuera del mercado a muchísimas industrias de países periféricos. https://taxation-customs.ec.europa.eu/carbon-border-adjustment-mechanism_en

Es decir, todo el acuerdo está pensado con lógicas liberales (y extractivistas), buscando aumentar la rentabilidad de las empresas de tecnología, sin importar el impacto en los pueblos.

El Pacto Digital Mundial

El Pacto Digital Mundial (GDC, por sus siglas en inglés)[27] es un marco para la cooperación digital internacional que se está negociando actualmente como un anexo al Pacto por el Futuro. Este pacto intergubernamental busca "construir un sistema multilateral que funcione para todos, en todas partes"[28] con acciones concretas para asegurar un futuro mejor para "toda la humanidad". Dicho en otras palabras, intenta ser un marco de gobernanza global en temas digitales. El mismo, se basa en los tres pilares del sistema de las Naciones Unidas (ONU): desarrollo, paz y seguridad, y derechos humanos. A simple vista, parecería ir en la dirección correcta. Pero, algunas críticas surgen por parte de la sociedad civil[29]:

> **1.** No aborda el control corporativo de la infraestructura digital: El PGD "reduce la idea de infraestructura digital pública a 'sistemas digitales compartidos, sin cuestionar cómo se transferirá a manos públicas la naturaleza esencialmente pública de los aspectos clave de la infraestructura digital"[30].
>
> **2.** Falta de compromiso con la regulación de las empresas: "no contiene compromisos concretos para que los estados regulen a las empresas para que cumplan con los derechos humanos digitales"[31].
>
> **3.** No se aborda el desarrollo de capacidades locales a largo plazo: "Si bien la localización de los modelos de IA dominantes puede ser una solución pragmática a corto plazo, se necesitan estrategias políticas para desarrollar la capacidad a largo plazo para crear modelos de IA locales, contextualizados y culturalmente apropiados"[32].
>
> **4.** Se necesitan mecanismos para garantizar una innovación inclusiva: "los bienes públicos digitales no pueden considerarse automáticamente como facilitadores de la innovación inclusiva. Su mérito funcional radica únicamente en la gobernanza de interés público "[33].

El PGD no enfrenta el problema de la concentración de poder en la economía digital y, en cambio, propone soluciones superficiales que no abordan las causas profundas de la desigualdad digital. Se necesita un enfoque más contundente que promueva el desarrollo de capacidades locales, la gobernanza pública de la infraestructura digital y la regulación efectiva de las empresas para garantizar un futuro digital más justo y equitativo, lejos de prácticas extractivistas y colonialistas.

¿Y el regionalismo? Falta de capacidad de capacidades

Frente a esta realidad desesperanzadora respecto a los instrumentos de gobernanza global, surge el interrogante: ¿y si creamos una gobernanza regional para tener IA con nuestros valores y principios?

27 Espacio web del Pacto Digital Mundial: https://www.un.org/digital-emerging-technologies/es/global-digital-compact

28 "The Global Digital Compact We Need for People and the Planet", *IT for Change*, 2024: https://itforchange.net/index.php/global-digital-compact-we-need-for-people-and-planet

29 Espacio del Global Digital Justice Forum: https://globaldigitaljusticeforum.net/

30 "The Global Digital Compact We Need for People and the Planet", *IT for Change*, 2024: https://itforchange.net/index.php/global-digital-compact-we-need-for-people-and-planet

31 *Ídem*.

32 *Ídem*.

33 *Ídem*.

En principio parece una estrategia correcta e inteligente frente a un mundo que parece no dar respuesta a las desigualdades estructurales que enfrentamos de antaño. Es más, parece querer exacerbarlas.

En América latina poseemos espacios de integración como el Mercosur o la Alianza del Pacifico, entre otros. Pero allí los debates en torno a la IA parecen quedarse en temas superficiales de lo que se espera de la IA: no discriminar, proteger la privacidad, asegurar acceso, entre otros principios. Si bien estos temas son importantes, no logran dar con debates profundos respecto a la desigualdad, la concentración de poder y la industrialización digital. La pregunta sería entonces, cómo lograr que las voces de la sociedad civil se oigan en una burocracia gubernamental que parece haber olvidado para quiénes deberían gobernar. Los espacios de participación son limitados y las agendas parecen estar cooptadas por los lobbies corporativos.

Por otro lado, la construcción de alternativas soberanas se hace necesaria y urgente: entender la tecnología con nuestras propias lógicas y salir de la dependencia es también producir tecnología que sirva a nuestras poblaciones. Sobre todo, Infraestructuras Públicas Digitales Regionales que presenten servicios y sean alternativas a las *big tech*.

Y en este sentido, cabe preguntarse: ¿tenemos esa capacidad? la respuesta es ambigua y heterogénea. Existen ideas, existen recursos humanos, existen investigadores pensando alternativas y trayendo la tradición latinoamericana de pensamiento al desarrollo de tecnologías digitales alternativas. Parece haber capacidad de pensar un futuro mejor al que se plantea. Pero, ¿tiene esa capacidad la oportunidad de materializarse en oportunidades? ¿Se puede tener capacidad si existe fuga de cerebros?, ¿si falta financiación?, ¿si existen crisis estructurales de balanza de pagos?, ¿si hay acuerdos comerciales firmados que impiden el accionar del Estado para promover políticas de desarrollo digital nacional?

Creemos que aun con toda esta adversidad, es posible. La región ha encontrado respuestas a problemas adversos más de una vez. Y la integración productiva Sur-Sur puede ser la salida. En efecto, la construcción de tecnologías soberanas sólo es posible si pensamos en integrar esfuerzos y salir a vender productos a nuestros mercados diseñados por y para nuestra propia gente.

Conclusión: Hacia una Justicia Económica digital.

A lo largo de este texto hemos visto desde diversos ángulos cómo el Sur Global se encuentra enmarcado en una lógica extractivista y colonialista que se repite desde antaño hasta la actualidad, ahora en la fabricación de la IA a nivel global.

En la disputa político-económica, parece que hemos perdido la batalla por lograr aprovechar una nueva tecnología para cambiar las lógicas, salir de la dependencia estructural y lograr conquistar nuevas ventajas comparativas con valor agregado.

Pero no todo está perdido. Existe hoy día una batalla geopolítica por el dominio del mercado tecnológico entre China y EE.UU., y si aun conservamos nuestra soberanía regulatoria y somos astutos económicamente, construyendo estrategias de integración productiva, podemos lograr posicionarnos en mercados globales con insumos intermedios para la fabricación tecnológica. En la disputa económica, si nos aliamos en integración productiva, tenemos capacidad de generación de tecnologías alternativas que traigan innovación al mercado. La pregunta es si vamos a tener gobiernos que permitan la venta de desarrollos con participación público-privado, siendo los capitales transnacionales los nuevos dueños de nuestra propiedad intelectual, o si van a proteger la soberanía industrial y la creación de valor.

Se necesita urgente justicia económica y digital. Otra tecnología es posible. Pero para lograrla, necesitamos que cada rincón del planeta pueda crearla, generando mercados competitivos e innovadores que se beneficien de la diversidad.

No puede ser un destino inevitable. La humanidad puede construir algo mejor si no permitimos la concentración de poder inigualable que supone la IA. El debate crece a nivel global. Hay que estar a la altura de las circunstancias. Si el Sur Global se une en esta pelea, somos imparables. No podemos permitir un nuevo siglo de saqueo de recursos. No volvamos a ser Potosí.

3. El impacto medioambiental de la Inteligencia Artificial: No es una nube, sino una nave industrial

Ana Valdivia
Oxford Internet Institute (University of Oxford)
Centre for Capitalism Studies (University College London)

En un contexto de emergencia climática es necesario y prioritario reflexionar sobre qué tecnologías van a ser útiles y eficaces para cuidar vidas tanto humanas, como no humanas. La Inteligencia Artificial se ha propuesto como una solución a este problema debido a su capacidad de analizar patrones en datos. De hecho, las grandes tecnológicas están ofreciendo soluciones algorítmicas para librarnos de los impactos del cambio climático. No obstante, el creciente consumo energético de esta tecnología ha puesto en tela de juicio estas narrativas, convirtiendo la Inteligencia Artificial en parte del problema, más que de la solución. Varios estudios académicos y periodísticos han demostrado que la reciente Inteligencia Artificial generativa consume una gran cantidad de recursos naturales, como minerales, agua y tierra, una exigencia que genera conflictos políticos y medioambientales. Este capítulo realiza una revisión crítica de cómo la inteligencia artificial y su infraestructura contribuyen al cambio climático creando resistencia por parte de comunidades locales que se ven enfrentadas al extractivismo originado para seguir entrenando modelos de Inteligencia Artificial. Además, invita a repensar juntas qué tipo de propuestas son una oportunidad para cambiar de rumbo esta tecnología e imaginar un mundo donde el algoritmo realmente nos beneficie a todas, si es que esto es posible.

Introducción

Climática, un medio independiente especializado en clima y biodiversidad, anunció que el año 2024 será el más caluroso de la historia y el primero que se superen los 1.5°C de calentamiento (Robaina, 2024). La temperatura media global ha sido de 1,59°C y esto conlleva graves consecuencias para la vida en la Tierra tal y como nos avisan tantas científicas climáticas. De hecho, Valencia (España) se ha convertido estos días en la muestra de tantos años quemando combustibles fósiles: un Mar Mediterráneo demasiado caliente por la actividad capitalista que alimenta tormentas e inundaciones jamás vividas antes. Y es que las científicas nos lo vienen anunciando desde 1856, cuando Eunice Foote probó que ciertos gases crearían el efecto invernadero en la atmósfera (Foote, 1856). Otro científico, Guy Callendar, concluyó que la temperatura de la Tierra había aumentado 0,3°C en tan solo 50 años. Desde entonces, se han propuesto diferentes modelos físicos para entender mejor el efecto de los combustibles fósiles en la Tierra, dejando evidencias de cómo la ciencia y la tecnología pueden emplearse para mejorar nuestro entendimiento sobre el impacto que tiene este modelo de producción y consumo.

En los últimos años, los avances tecnológicos también han mejorado el conocimiento científico de la emergencia climática y su relación con los combustibles fósiles. Modelos de computación más sofisticados y sensores que miden con mejor exactitud nos permiten diseñar modelos climatológicos más exactos. Dentro de este contexto tecnológico, la Inteligencia Artificial se ha presentado como una herramienta clave para luchar contra el cambio climático. Esta tecnología es capaz de analizar patrones en grandes cantidades de datos. Cowls et al. (2023) explican cuáles son las oportunidades que ofrece esta tecnología para mitigar el cambio climático en diferentes sectores como el transporte, la energía, el urbanismo, predicción del clima, etc. Por ejemplo, se han propuesto ciertos algoritmos basados en técnicas de aprendizaje automático que mejoran la eficiencia energética de un edificio a partir de analizar los patrones de consumo de energía y optimizar su uso (Zekić-Sušac, Mitrović y Has, 2021).

Dado este optimismo tecnológico en el que la Inteligencia Artificial podrá ser empleada para afrontar la emergencia climática, las grandes corporaciones tecnológicas no han dudado en proponer y financiar sus propias soluciones a través de sus compromisos filantrópicos. Un ejemplo es la compañía de Jeff Bezos: Amazon. Esta firma tecnológica, que tiene un beneficio económico de 36.852 millones dólares estadounidenses según Wikipedia, fundó en 2020 la Bezos *Earth Fund* con 10 billones de dólares para combatir el cambio climático. Esta fundación anunció el 16 de abril de 2024 que iba a invertir 100 millones de dólares del total de los 10 billones en proyectos de Inteligencia Artificial que "protejan nuestro planeta". Así lo anunciaba la vicepresidenta de esta fundación en su canal de Youtube:

> "Cuando escuchas sobre Inteligencia Artificial, ¿qué te viene a la cabeza? Para mí es el potencial para la humanidad de solucionar problemas, grandes problemas como el cambio climático o la pérdida de naturaleza. La pregunta es ¿cómo? Eso es lo que estamos explorando en la Bezos Earth Fund y estoy emocionada de anunciar nuestro programa 'El gran reto de la Inteligencia Artificial para el clima y la naturaleza', un compromiso de 100 millones de dólares para proteger nuestro planeta."
>
> Lauren Sánchez (2024)

Las áreas que propone la *Bezos Earth Fund* para mitigar el cambio climático son: proteínas sostenibles, red de suministro eléctrico y conservación de la biodiversidad. Algunas de las preguntas que se plantean en el vídeo promocional son, ¿puede la Inteligencia Artificial analizar combinaciones de proteínas que creen carne artificial? O, ¿puede la Inteligencia Artificial traer energías renovables a comunidades con falta de electricidad? No obstante,

una curiosidad que nos planteamos es que esta fundación no se pregunta si la Inteligencia Artificial puede ayudar a Amazon a contabilizar o mitigar su huella de carbono. Y es que Amazon fue acusada en 2022 de subestimar drásticamente sus emisiones de gases de efecto invernadero.

El periodista Will Evans realizó una investigación en la que expuso como Amazon no quiso compartir públicamente los datos de su huella de carbono con organizaciones independientes y sin ánimo de lucro como el *Carbon Disclosure Project* para hacer público su huella de carbono y compararla con otras compañías. Además, Amazon en esas cuentas solo tenía contabilizaba las emisiones de carbono de los productos que propiamente fabrica e ignoraban el cómputo de las emisiones de los productos que no fabrica, los cuáles se estimaban en un 60% de sus ventas. Como resultado, algunas trabajadoras de Amazon están presionando a la empresa para que incorpore prácticas realmente sostenibles. De hecho, han publicado recientemente el informe *Burns Trust: The Amazon Unsustainability Report* (Quemando la Verdad: El informe de la insostenibilidad de Amazon) que expone críticamente cómo la gran empresa tecnológica contribuye a la emergencia climática al vender herramientas de Inteligencia Artificial a empresas de combustibles fósiles (Amazon Employees for Climate Justice, 2024).

Además, poco se sabe del impacto medioambiental de la infraestructura digital de Amazon. Esta empresa provee de servicios “en la nube” a otras empresas, como almacenamiento de datos o ejecución de algoritmos. Pero estos servicios “en la nube”, no se desarrollan en nuestro cielo azul, sino en una nave industrial en tierra firme. Estos servicios que Amazon ofrece transcurren en centros de datos, también conocidos como granjas de servidores, que necesitan energía y agua para mantener encendidos y enfriados los servidores en los que se almacenan y gestionan datos o se entrenan productos de Inteligencia Artificial como ChatGPT. Debido a que los centros de datos funcionan las 24h, los 7 días de la semana, los 365 días del año, su impacto medioambiental es mayúsculo y ha llamado la atención de académicas, periodistas y activistas en los últimos años. Por ejemplo, comunidades locales situadas cerca de los centros de datos de Amazon en Estados Unidos, el país con mayor número de centros de datos a nivel mundial, se han manifestado debido al ruido que generan y la cantidad de agua que gastan. En el estado de Virginia, una movilización se opuso al centro de datos de Amazon debido a que iba a construir más de 40km de torres de alta tensión en una zona rural (Smolaks, 2015). Este caso, de entre otros casos que veremos en este capítulo, muestra cómo la Inteligencia Artificial y su creciente infraestructura supone un problema para la emergencia climática exacerbando desigualdades y vulnerando derechos medioambientales.

Este capítulo titulado “El impacto medioambiental de la Inteligencia Artificial: no es una nube, sino una nave industrial” que se encuentra dentro de este libro *Los engranajes de la máquina. Poder y desigualdades en la inteligencia artificial* presenta una mirada crítica a la industria y la infraestructura de la Inteligencia Artificial. En la primera parte analizaremos la materialidad de la Inteligencia Artificial haciendo un repaso a su infraestructura y las cadenas de suministro. En la segunda parte mostraremos casos reales de qué resistencias han emergido debido a las desigualdades sociales que crea la infraestructura de la Inteligencia Artificial. En la última parte, presentaremos una valoración sobre la sostenibilidad del actual modelo de desarrollo de esta tecnología, así como qué tipo de herramientas existen para resistir la desigualdad de la industria tecnológica.

La infraestructura de la IA y sus cadenas de suministro

Durante varias décadas se ha invisibilizado la infraestructura física del mundo digital y de Internet. Las grandes tecnológicas han utilizado la metáfora de la nube para hablar de aquel lugar dónde se almacenan nuestras conversaciones, imágenes y otros materiales digitales, invisibilizando grandes centros de datos y cables submarinos. De hecho, hasta se sigue utilizando el icono de una nube azul en nuestros ordenadores para referirse a ese lugar. Esta metáfora de la nube da una sensación de inmaterialidad y nos propone una imagen etérea de la digitalización, sin que ello tenga ninguna consecuencia o impacto medioambiental (Jacobson y Hogan, 2019; Wiig, 2015). Aunque algunas académicas, activistas y periodistas ya nos alertaban hace unos años de la materialidad del mundo digital y nos invitaban a estar alerta a la cantidad de agua y electricidad que consumían (Hogan, 2015; Velkova, 2016; Brevini, 2021; Peña, 2023), no ha sido hasta los últimos años que se ha dejado de invisibilizar la materialidad digital.

La Inteligencia Artificial se ha anunciado a bombo y platillo como una tecnología que nos podría librar del cambio climático mediante algoritmos que identifican la deforestación o predicen la subida del nivel del mar (Boston Consulting Group, 2022). No obstante, la Inteligencia Artificial generativa, entrenada para *generar* textos o imágenes a través de otros textos o imágenes, ha puesto en cuestión los beneficios que esta tecnología podría tener en la lucha contra el cambio climático debido a la cantidad de recursos que necesita. Algunos cálculos estiman que Llama 3, el gran modelo del lenguaje entrenado por Meta, gastó 22 millones de litros de agua en tan solo 97 días, lo que una ciudadana del barrio de Sant Andreu de Barcelona gastaría en 643 años (Li et al. 2023). Y es que la cantidad no solo de agua, sino también de electricidad y otros recursos naturales que necesita este tipo de algoritmos para ser entrenados ha ido creciendo a lo largo de los años, debido a que el tamaño del algoritmo ha crecido y se ha vuelto más sofisticada, por lo que necesita más capacidad de cómputo y consume más recursos.

Pero no es solo la cantidad de recursos que se necesitan para entrenar estos algoritmos. Como recientemente se ha investigado, las cadenas de suministro de la Inteligencia Artificial nos ayudan a visibilizar los recursos y la mano de obra que se necesita para fabricar un chip de Inteligencia Artificial que luego sirva para entrenar el algoritmo (Valdivia, 2024). Estas cadenas de suministro se pueden estructurar en 3 fases: extracción de recursos minerales, centros de datos y vertederos electrónicos.

Minas, fábricas y chips para la Inteligencia Artificial

El activista medioambiental y doctor en ciencias políticas Jorge Riechmann denunciaba recientemente que en tan solo ocho años (2016-2023) la humanidad ha consumido más toneladas de materiales que en todo el siglo XX (MaterialFlows.net, 2024). Esta cantidad de materiales demuestra que nuestras sociedades están consumiendo más productos hechos de materiales que deben ser extraídos, transportados y fabricados. La industria textil o alimentaria contribuyen a esa extracción de materiales que es necesaria para fabricar nuestra ropa o alimentos. No obstante, debido a la transformación digital, la industria electrónica también está contribuyendo a ese extractivismo a través de los materiales necesarios para fabricar móviles, ordenadores y chips.

Dentro de esta transformación digital, la atención se ha volcado sobre la industria de la telefonía móvil o portátiles, analizando qué tipo de materiales son necesarios para fabricar nuestros teléfonos y cuál es su impacto medioambiental (Rodriguez, 2024); pero se ha dedicado muy poca atención sobre la industria de los chips necesarios para la Inteligencia Artificial. Y es que, de hecho, las empresas que fabrican productos electrónicos para esta tecnología se han convertido en las más poderosas teniendo en cuenta la variable de la capitalización del mercado, que es un indicador económico que recoge el valor total de

las acciones de una empresa. Así, la firma estadounidense NVIDIA, que es la empresa tecnológica que distribuye el 80% de chips utilizados para la Inteligencia Artificial—también conocidos como GPUs por sus siglas en inglés *Graphics Processing Unit*—sobrepasó el pasado 15 de noviembre de 2024 a otras empresas como Apple, Microsoft o Amazon, convirtiéndose en la compañía líder mundial en términos de capitalización del mercado, acumulando 3.600 billones de dólares estadounidenses (Statista, 2024).

A pesar de que NVIDIA es la empresa que distribuye los chips, TSMC con sede en Taiwán es la encargada de fabricar las GPU. Esta empresa utiliza maquinaria especializada holandesa para fabricar la electrónica de la Inteligencia Artificial, lo que también tiene serias consecuencias medioambientales que ponen en tela de juicio cómo esta tecnología puede librarnos del cambio climático. Por ejemplo, el último informe de sostenibilidad de TSMC estima que el agua utilizada por sus fábricas ha ido incrementándose año tras año, gastando 260.000 metros cúbicos diarios de agua en 2022 (S&P Global Ratings, 2024). Debido a la cantidad de agua que gasta TSMC, el gobierno taiwanes se vió envuelto en un escándalo cuando en 2021 anunció a sus agricultoras que debían de priorizar el agua para fabricar chips, antes que cultivar arroz (Zhong and Chang Chien, 2021). Pero no es solo agua. Las GPU están formadas en su 99.9% de silicio, es decir arena, cuya extracción también lleva a serios problemas medioambientales relacionados con las operaciones de depósitos mineros (Missouri Coalition for the Environment, 2023). También se utilizan otros materiales como el tántalo, el oro o la plata cuya extracción tiene, igualmente, impactos negativos aunque se utilice menor cantidad (Valdivia, 2024).

Centros de datos

Los algoritmos de la Inteligencia Artificial se han vuelto cada vez más sofisticados, lo que requiere también de una gran cantidad de datos para entrenarlos. Si analizamos el tamaño de un algoritmo en función de la cantidad de parámetros—esto son variables de configuración necesarias para guiar al algoritmo en su entrenamiento—vemos que ha ido aumentando en los últimos años. Por ejemplo, mientras BERT, un modelo del lenguaje publicado por Google en 2018 contiene 340 millones de parámetros, el algoritmo de ChatGPT diseñado por OpenAI en 2020, GPT-3, tiene 175 billones. Este incremento en el número de parámetros ha transformado la manera en la que se programan los algoritmos, ya que ya no se pueden entrenar en "local", es decir, en tu propio ordenador. Estos algoritmos deben ser entrenados en "la nube", es decir en un centro de datos, debido no solo al tamaño del algoritmo, sino también a la cantidad de datos que necesita. Es por esta razón que la atención sobre el impacto medioambiental de la Inteligencia Artificial ha aumentado en los últimos años, ya que la "nube", como decía en la sección anterior, no es una "nube" sino una nave industrial que ingiere grandes cantidades de agua y energía, además de suelo.

Se estima que los centros de datos consumen un 1% de la electricidad mundial (Spencer and Singh, 2024). Aunque no todos los centros de datos sirven para entrenar Inteligencia Artificial, esta tecnología ha aumentado el consumo de la factura de los centros que sí la utilizan. De hecho, Microsoft y Google declararon en su informe de sostenibilidad de 2024 que sus emisiones de carbono habían aumentado debido a la energía que se necesitaba para entrenar sus algoritmos de Inteligencia Artificial. Y es que muchos centros de datos utilizan combustibles fósiles para generar energía. Debido a los compromisos de emisiones cero que estas empresas tienen para 2030, Google Amazon y Microsoft han anunciado inversiones en energía nuclear. Así, en un futuro podrán decir que no emiten carbono, pero consumirán uranio (Penn and Weise, 2024).

Los centros de datos también consumen agua. Esta infraestructura necesita enfriar sus salas de servidores y el agua es lo más eficiente para realizar dicha tarea. Los centros de datos son edificios que funcionan los 365 días del año, por lo que su necesidad de este fluido es constante. Por lo tanto, el impacto medioambiental de estas instalaciones está también

relacionado con la cantidad de agua que consumen. Pero es difícil conocer la cantidad exacta que emplean ya que existe bastante opacidad respecto a su consumo de agua. Las empresas se esconden tras la confidencialidad corporativa para no publicar estas cifras. No obstante, ha habido escándalos relacionados con el gasto de agua de los centros de datos. Microsoft se vió envuelto en uno de ellos en Holanda en 2022. Mientras la tecnológica fundada por Bill Gates prometía que solo necesitaba de 12 a 20 millones de litros de agua para su construcción, un periódico local destapó en mitad de una alerta por sequía durante el verano que realmente el centro de datos consumió 4 veces más durante su fase de construcción: 84 millones de litros anuales (Vuijk, 2022). Aunque según Microsoft 36 millones de litros se devolvían al suministro, esto provocó el rechazo de la comunidad local que ya se había opuesto al proyecto antes de que se construyera. El mencionado centro de datos de Microsoft había sido una de las dos excepciones a una moratoria aprobada por el gobierno holandés ante la campaña de oposición ciudadana.

Basura electrónica

El impacto medioambiental de la Inteligencia Artificial no acaba en un centro de datos, sino en un vertedero electrónico. Aunque mucho se ha hablado de la cantidad de residuos electrónicos que el Norte Global envía al Sur Global y de los daños medioambientales que esa basura genera, el análisis de cómo la Inteligencia Artificial y su infraestructura pueden contribuir a ello está muy poco documentado. En una reciente investigación sobre el impacto medioambiental de la Inteligencia Artificial y su cadena de suministro se establece que las GPUs tienen un ciclo de vida de entre 3 a 5 años (Valdivia, 2024). Eso significa que antes de 5 años los centros de datos desechan sus chips e instalan nuevos, lo que puede contribuir a aumentar la basura electrónica. De hecho, un reciente estudio de Nature estimaba que, si no se toman medidas, la basura electrónica generada por los últimos modelos de la Inteligencia Artificial podría incrementarse entre 1,2 a 5 millones de toneladas durante el periodo 2020 a 2030 (Wang *et al*., 2024).

El impacto medioambiental de la basura electrónica está relacionado con la contaminación del suelo y el agua debido a los químicos que se desprenden de los circuitos electrónicos y los contaminan. Por ejemplo, se ha demostrado que existen concentraciones más elevadas de mercurio, tanto en el suelo como en el agua subterránea o en superficie cerca de dos vertederos electrónicos en Ghana (Amponsah et al., 2022)

El impacto en las comunidades locales: el caso de meta en Talavera de la Reina (España)

Talavera de la Reina es una ciudad situada en Castilla-La Mancha, una región española ampliamente conocida por su legado rural. Sin embargo, esta zona ha sufrido una grave despoblación debido a los flujos migratorios del ámbito rural hacia las ciudades, un fenómeno conocido como la “España vaciada” (Taibo, 2021). Talavera de la Reina también se ha hecho conocida recientemente por albergar el primer centro de datos de hiperescala de Meta en España. Este proyecto ha sido considerado de Interés Singular por el gobierno regional de Castilla-La Mancha, junto con un casino, un aeropuerto y un campo de golf, lo que facilita la “urbanización mediante la reclasificación de suelo, descuidando los ecosistemas existentes y permitiendo la construcción en áreas protegidas” (Escudero-Gómez, 2023). El proyecto de este centro de datos ocupa 191 hectáreas para construir 130.000 metros cuadrados para almacenar servidores. Este ambicioso proyecto promete un plan sostenible con acciones como “revertir la pérdida de biodiversidad” y “restaurar más volumen de agua del que se consume” (Meta y Zarza Networks, 2023).

El informe ambiental, al cual tuvimos acceso, estima que el centro de datos tendrá una capacidad eléctrica instalada de 248 MW, equivalente al consumo eléctrico anual de 71 hogares españoles. En términos de agua potable, el consumo total se estima en 327.000.000 litros anuales, lo que equivale al consumo anual de 7.000 hogares. Además, en su informe ambiental se declara que el proyecto se encuentra dentro de la zona relevante para el Plan de Recuperación del Águila Imperial Ibérica y el Buitre Negro, pero el terreno que ocupará el centro de datos representa solo el 0,004 y 0,009% de esta zona. De hecho, se evaluó el impacto ambiental sobre estas especies protegidas contando el número de aves observadas en seis visitas de campo realizadas entre diciembre de 2021 y junio de 2022.

Sin embargo, el informe ambiental de Meta recibió críticas por parte de organizaciones ambientalistas, académicos y periodistas. Como en el caso de Holanda, uno de los principales puntos de crítica fue el consumo de agua, un tema sensible dado que España sufre sequías muy severas que se van a ir acentuando con la emergencia climática. Un punto importante de objeción, planteado formalmente por SEO/Birdlife, fue el consumo de agua de esta infraestructura. No obstante, esta alegación fue desestimada. El centro de datos de hiperescala de Meta declaró que su consumo de agua potable se redujo de 327.000.000 a 40.000.000 litros anuales, reevaluando la demanda máxima de agua de 37 a 10 litros por segundo. A pesar de ello, el centro de datos de Meta consumirá el 8% del consumo total de Talavera de la Reina. Meta planea reducir dramáticamente el consumo de agua mediante el uso de sistemas de enfriamiento por aire seco, es decir, sistemas de flujo de aire frío que no utilizan agua. Sin embargo, la literatura sobre enfriamiento de centros de datos sugiere que este método es efectivo en zonas de baja temperatura y ese no es el caso de Talavera de la Reina durante el verano. Además, la Sociedad Española de Ornitología (SEO/Birdlife) reclamó que el análisis de monitoreo de aves debería mejorarse para evaluar mejor el impacto del centro de datos sobre la fauna local. Esta organización especificó que el estudio debería extenderse al menos un año para evaluar de manera más precisa el impacto del centro de datos en la región. El informe ambiental tampoco consideró cómo el centro de datos de Zuckerberg podría afectar la pérdida de áreas de alimentación y forrajeo. El informe no menciona que la zona estará ubicada a solo 3 kilómetros de la zona crítica del Águila Imperial Ibérica. Nuevamente, estas alegaciones sobre el impacto de los centros de datos en la fauna fueron desestimadas.

En respuesta, Meta ha provocado la creación de la primera organización de base contra los centros de datos en España: Tu Nube Seca Mi Río. Esta organización se define como una agrupación tecnoambiental surgida en abril de 2023 con el propósito de:

> "[S]ensibilizar sobre el impacto ambiental de los centros de datos, especialmente en cuanto al uso de recursos hídricos. Nos unimos como resultado de la iniciativa de Meta/Facebook de crear un gran centro de datos en Talavera de la Reina (Madrid, España)."
>
> *Tu Nube Seca Mi Río.*

En un evento organizado en la Universidad de Cambridge (Reino Unido) que reunió a diversas comunidades locales en Europa resistiendo los impactos de centros de datos, una militante de Tu Nube Seca Mi Río declaró que Mark Zuckerberg la "motivó a convertirse en activista contra los centros de datos". Su familia había resistido anteriormente la construcción del aeropuerto, otro de los Proyectos Singulares apoyados por el gobierno local, y ahora investigaba cómo este centro de datos también podría impactar negativamente en el territorio. Los agricultores ya están luchando para regar sus campos debido a las severas sequías que afectan a esta zona. Aunque no sepamos, a día de hoy, si este centro de datos en Talavera de la Reina tenga relación directa con la Inteligencia Artificial, sí sabemos que Mark Zuckberg planea hospedar virtualmente el Metaverso y productos relacionados con él, algunos de los cuales están diseñados con Inteligencia Artificial (Nix, 2023). En la actual emergencia climática, surge una pregunta urgente: ¿priorizaremos el agua para sostener la vida o para sostener los sueños de Silicon Valley?

Una propuesta al posible desarrollo de la Inteligencia Artificial para un contexto de emergencia climática

Existen soluciones tecnológicas que pueden ser beneficiosas para prevenir y mitigar los impactos climáticos. Modelos físicos y de ecuaciones diferenciales que ayudan a predecir el rumbo de un huracán, una ola de calor o una DANA han demostrado ser eficaces para avisar a la población con cierta antelación de los riesgos climáticos (siempre y cuando la clase política esté sensibilizada sobre los peligros medioambientales que supone la emergencia climática). Sabemos que estos impactos, que cada vez serán más severos y pondrán en riesgo tanto a la vida humana como la no humana, son consecuencia de un sistema económico capitalista, neoliberal y extractivista que busca quemar combustible fósil para acumular capital. La Inteligencia Artificial, si no se aplica con cierto sentido crítico, solo servirá para seguir alimentando a ese sistema económico que destroza nuestros mares, tierras y montañas. Hay que poner en tela de juicio cuál es el beneficio qué un algoritmo puede traer a la sociedad, comparándolo y analizándolo críticamente con cuáles pueden ser sus limitaciones, riesgos y consecuencias negativas, además de la cantidad de recursos naturales y financieros que supone la puesta en funcionamiento de dicho algoritmo.

A pesar de que mucho se ha dicho sobre los riesgos de los algoritmos, como el sesgo discriminatorio que amplifican (Valdivia y Sánchez Monedero, 2022), hoy en día la infraestructura también se ha convertido en un eje opresor. Centros de datos que esconden sus impactos medioambientales o no muestran de manera transparente la cantidad de agua que gastan, cadenas de suministro de chips para entrenar algoritmos de Inteligencia Artificial que extraen los minerales creando graves daños medioambientales sobre el suelo, y basura electrónica creada por esos chips al final de su vida útil que siguen contaminando el suelo, ríos y acuíferos (Taffel, 2021). Estos procesos de extractivismo y desposesión resuenan con las teorías críticas de colonialismo de datos elaborado por los académicos Ulises A. Mejias y Nick Couldry, la cual traza una genealogía entre el colonialismo que está intrínseco en el extractivismo de datos por parte de la gran industria tecnológica y los recursos necesarios para fabricar dispositivos digitales que se extraen de la Mayoría Global para provecho del Norte Global (Mejias y Couldry, 2024). En este contexto de emergencia climática, debemos permanecer atentas, no solo a si la solución tecnológica va a ser realmente útil, sino también a las injusticias sociales y medioambientales que van a emerger debido a la creciente industria de la infraestructura digital, ya que se están instalando más centros de datos, lo que contribuye a fabricar más electrónica. Todo ese crecimiento tiene un impacto sobre el suelo, los insectos, las aves y las comunidades locales de los lugares en los que se instala el centro de datos, la mina de dónde se extrae el mineral para fabricar el chip o el vertedero electrónico dónde se deposita ese chip al final de su ciclo.

Debido a las consecuencias negativas y al impacto discriminatorio de varias soluciones tecnológicas, instituciones gubernamentales como la Comisión Europea decidieron establecer un marco regulatorio de reglas del juego para esta tecnología: el *Artificial Intelligence Act*. Aunque organizaciones de derechos digitales quedaron satisfechas con el resultado final del texto, uno de los temas que esta regulación no toma en cuenta es el daño medioambiental que la Inteligencia Artificial crea. De hecho, solo Estados Unidos ha publicado un texto en el Senado durante el mandato de Biden en 2024 en el que propone a la agencia de medioambiente estadounidense analizar los impactos medioambientales de la Inteligencia Artificial y proporcionar un sistema que informe de dichos daños. Aunque el marco legal europeo de la Inteligencia Artificial no toma en cuenta su impacto medioambiental, es cierto que existen otras herramientas legales que pueden ser útiles para esclarecer cuántos recursos consume esta tecnología. Por ejemplo, el Parlamento Europeo revisó una ley de 2012, *Energy Efficiency Directive*, que ahora obliga a los centros de datos de más de 50 MW a reportar sus métricas de consumo, tales como energía o agua consumida. No obstante, existe una oportunidad regulatoria para proteger a las comunidades locales delante de la creciente infraestructura de la Inteligencia Artificial, no solo de los centros de datos, sino también de las minas, fábricas de chips y basureros electrónicos.

Durante años se hablaba de la caja negra algorítmica, pero hoy en día deberíamos hablar también de la caja negra de la infraestructura algorítmica y digital. Necesitamos transparencia y rendición de cuentas sobre las cadenas de suministro de la Inteligencia Artificial y sus impactos. Por ejemplo, es imposible trazar todos los proveedores de corporaciones como NVIDIA o Microsoft, lo que hace realmente imposible cuantificar sus emisiones de carbono. Aunque existen marcos regulatorios que exigen más transparencia sobre las cadenas de suministro, muchos de los informes de sostenibilidad publicados por la industria tecnológica no tienen en cuenta las emisiones de sus proveedores. Esto hace que hoy en día no sepamos con certeza cuál es la huella de carbono de la industria tecnológica y digital. Necesitamos una regulación que promueva el bien común y una toma de decisión realmente democrática para la infraestructura digital, es decir, que permita a las comunidades decidir si quieren que se abra una mina para extraer silicio o la instalación de un centro de datos, informándoles transparentemente de las consecuencias de esa decisión. Informar sobre qué daños medioambientales se perpetúan en la cadena de suministro tecnológica, y también conocer cuáles son esos actores privados que están dañando nuestros ecosistemas.

Pero no solo necesitamos regulación. De hecho, también podemos reapropiarnos de la tecnología y, en especial, de la Inteligencia Artificial, orientarla hacia el bien común y alejarla de la acumulación de capital. Yásnaya Elena Aguilar Gil, que es una lingüista y activista mixe, introdujo el concepto de *tequiología*, una perspectiva que busca alejar a la tecnología de su función habitual de servir al mercado, la competencia y los intereses privados, para situarla al servicio de lo común y la cooperación. Esto implica compartir datos, hacer el código accesible y construir infraestructuras digitales públicas que realmente sirvan para cuidarnos en este contexto de emergencia climática. La idea es empezar desde lo pequeño y lo local, inspirándose en la histórica resistencia de los pueblos oprimidos que, organizados en comunidades pequeñas, tejieron redes de apoyo mutuo para evitar tributos o rebelarse contra los abusos de la Corona española.

En este marco, la Inteligencia Artificial aplicada a proyectos menos ambiciosos y de alcance local puede convertirse en una herramienta poderosa para el bien común, permitiendo organizar resistencias frente a injusticias como la explotación de recursos naturales o la opacidad de la industria de los centros de datos analizando grandes cantidades de información pública—algo que a la Inteligencia Artificial se le da muy bien— para que proporcione transparencia. Se trata de reapropiarnos de la Inteligencia Artificial para ponerla al servicio de las personas y de intereses colectivos. En palabras de la lingüista ay üüjk:

> “Si el mundo adoptara esta visión tequiológica, quizá podríamos rescatar el potencial creativo de las nuevas tecnologías, apartándolas de un sistema que las devora y pone en peligro la vida humana.”

Referencias

Amazon Employees for Climate Justice (2024). *Burns Trust: The Amazon Unsustainability Report*. Accesible en: https://static1.squarespace.com/static/65681f099d7c3d48feb86a5f/t/668ebf702516716ca72bbf98/1720631157044/unsustainability-report.pdf (Visitado el 10 de noviembre de 2024).

Amponsah, Lydia Otoo, et al. (2023). Mercury contamination of two e-waste recycling sites in Ghana: an investigation into mercury pollution at Dagomba Line (Kumasi) and Agbogbloshie (Accra). *Environ Geochem Health* 45, 1723–1737. https://doi.org/10.1007/s10653-022-01295-9

Brevini, Benedetta (2021). *Is AI Good for the Planet?* Polity, Cambridge (UK).

Boston Consulting Group (2022) *How AI Can Be a Powerful Tool in the Fight Against Climate Change*. Accesible en: https://web-assets.bcg.com/ff/d7/90b70d9f405fa2b67c8498ed39f3/ai-for-the-planet-bcg-report-july-2022.pdf (Visitado el 14 de noviembre de 2024).

Callendar, Guy Stewart (1938). The artificial production of carbon dioxide and its influence on temperature. *Quarterly Journal of the Royal Meteorological Society,* 64(275), 223-240. https://doi.org/10.1002/qj.49706427503

Cowls, Josh; Tsamados, Andreas; Taddeo, Mariarosaria y Floridi, Luciano (2023) The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. *AI & Society*, 1-25. https://doi.org/10.1007/s00146-021-01294-x

Escudero-Gómez, Luis Alfonso (2023) Construir cualquier cosa en cualquier lugar: los Proyectos de Singular Interés en la región de Castilla-La Mancha (España). *EURE* (Santiago) 49, no. 147: 1-24. http://dx.doi.org/10.7764/eure.49.147.08

Evans, Will (2023) Private Report Shows How Amazon Drastically Undercounts Its Carbon Footprint. *Reveal News.* Accesible en: https://revealnews.org/article/private-report-shows-how-amazon-drastically-undercounts-its-carbon-footprint/ (Visitado el 10 de noviembre de 2024).

Foote, Eunice (1856) ART. XXXI. Circumstances affecting the heat of the sun's rays. *American Journal of Science and Arts* (1820-1879), 22(66), 382.

Hogan, Mél (2015). Data flows and water woes: The Utah data center. *Big Data & Society,* 2(2). https://doi.org/10.1177/2053951715592429.

Jacobson, Kate y Hogan, Mél (2019) Retrofitted data centres: A new world in the shell of the old. *Work Organisation, Labour & Globalisation*, 13(2), 78-94. https://doi.org/10.13169/workorgalaboglob.13.2.0078

Li, Pengfei, et al (2023) Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models. ArXiv https://doi.org/10.48550/arXiv.2304.03271

Materialflows.net (2024). *Global trends of material use*. Accesible en: https://www.materialflows.net/global-trends-of-material-use/ (Visitado el 18 de noviembre 2024).

Mejias, Ulises A. & Couldry Nick (2024) *Data Grab: The New Colonialism of Big Tech (and How to Fight Back)*. WH Allen, Dublin (Ireland).

Meta y Zarza Networks (2023) *Proyecto de Singular Interés "Meta Data Center Campus".*

Missouri Coalition for the Environment (2023) *Impacts of Silica Mining*. Accesible en: https://moenvironment.org/wp-content/uploads/sites/370/2023/01/Impacts-of-Silica-Mining-1.3.2023-1.pdf (Visitado el 18 de noviembre de 2024).

Nix, Naomi (2023). Facebook pivoted to the metaverse. Now it wants to show off its AI. *The Washington Post*. Accesible en: https://www.washingtonpost.com/technology/2023/05/14/meta-generative-ai-metaverse/ (Visitado el 18 de noviembre de 2024).

Penn, Ivan y Weise, Karen (2024). Hungry for Energy, Amazon, Google and Microsoft Turn to Nuclear Power. *The New York Times.* Accesible en: https://www.nytimes.com/2024/10/16/business/energy-environment/amazon-google-microsoft-nuclear-energy.html (Visitado el 18 de noviembre de 2024).

Peña, Paz (2023). *Tecnologías para un planeta en llamas.* Paidós, Santiago de Chile (Chile).

Robaina, Eduardo (2024). Un 2024 para la historia: el año más caluroso y el primero por encima de 1,5 °C. *Climática*. Accesible en: https://climatica.coop/2024-ano-mas-caluroso-y-por-encima-15-oc-copernicus/ (Visitado el 10 de noviembre de 2024).

Rodriguez, Helena (2024). Los dos "Mobiles": radiografía del impacto ecosocial del sector de los 'smartphones'. *Climática*. Accesible en: https://climatica.coop/mobile-impacto-ecosocial-smartphones/ (Visitado el 18 de noviembre de 2024).

https://climatica.coop/mobile-impacto-ecosocial-smartphones/

S&P Global Ratings (2024). TSMC And Water: A Case Study Of How Climate Is Becoming A Credit-Risk Factor. S&P Global. Accesible en: https://www.spglobal.com/ratings/en/research/articles/240226-sustainability-insights-tsmc-and-water-a-case-study-of-how-climate-is-becoming-a-credit-risk-factor-12992283

Sánchez, Lauren (2024). AI for Climate and Nature: The Bezos Earth Fund Announces $100M Grand Challenge. Bezos Earth Fund. Accesible en: https://www.youtube.com/watch?v=nyP3yU5Qu9Y&ab_channel=BezosEarthFund (Visitado el 10 de noviembre de 2024).

Smolaks, Max (2015). Amazon faces Virginia protest over power lines. Data Center Dynamics, Accesible en: https://www.datacenterdynamics.com/en/news/amazon-faces-virginia-protest-over-power-lines/ (Visitado el 10 de noviembre de 2024).

Spencer, Thomas y Singh Siddharth (2024) What the data centre and AI boom could mean for the energy sector. *International Energy Agency*. Accesible en: https://www.iea.org/commentaries/what-the-data-centre-and-ai-boom-could-mean-for-the-energy-sector (Visitado el 18 de noviembre de 2024).

Statista (2024) Leading tech companies worldwide as of November 15, 2024, by market capitalization (in billion U.S. dollars). *Statista*. Accesible en: https://www.statista.com/statistics/1350976/leading-tech-companies-worldwide-by-market-cap/ (Visitado el 18 de noviembre de 2024).

Taffel, Sy (2019) *Digital Media Ecologies: Entanglements of content, code and hardware.* Bloomsbury, Ireland (Dublin).

Taibo, Carlos (2021) *Iberia vaciada: Despoblación, decrecimiento, colapso*. Los Libros de la Catarata.

TSMC (2023) *TSMC 2022 Sustainability Report*.

Valdivia, Ana y Sánchez-Monedero, Javier (2022). *Una introducción a la IA y la discriminación algorítmica para movimientos sociales.* AlgoRace. Accesible en: https://www.algorace.org/wp-content/uploads/2024/09/2022-11-informe-algorace.pdf (Visitado el 19 de noviembre de 2024).

Valdivia, Ana (2024). The supply chain capitalism of AI: a call to (re) think algorithmic harms and resistance through environmental lens. *Information, Communication & Society*, 1-17. https://doi.org/10.1080/1369118X.2024.2420021

Velkova, Julia (2021). Data that warms: Waste heat, infrastructural convergence and the computation traffic commodity. *Big Data & Society*, 3(2). https://doi.org/10.1177/2053951716684144.

Vuijk, Bart (2022). Datacenter Microsoft Wieringermeer slurpte vorig jaar 84 miljoen liter drinkwater. *Noordhollandsdagblad*. Accesible en: https://www.noordhollandsdagblad.nl/regio/noordkop/datacenter-microsoft-wieringermeer-slurpte-vorig-jaar-84-miljoen-liter-drinkwater/11414775.html (Visitado el 18 de noviembre de 2024).

Wang, Peng, et al (2024). E-waste challenges of generative artificial intelligence. *Nature Computational Science*, p. 1-6. https://doi.org/10.1038/s43588-024-00712-6

Wiig, Alan (2015). The Urban, Infrastructural Geography Of 'The Cloud'. Looking at where data moves, where it *lives*. *Medium*. Accesible en: https://medium.com/vantage/the-urban-infrastructural-geography-of-the-cloud-1b076cf9b06e (Visitado el 14 de noviembre de 2024).

Zekić-Sušac, Marijana; Mitrović, Saša y Has, Adela (2021). Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities." *International journal of information management* 58. https://doi.org/10.1016/j.ijinfomgt.2020.102074.

Zhong, Raymond and Chang Chien, Amy (2021). Drought in Taiwan Pits Chip Makers Against Farmers. *The New York Times.* Accesible en: https://www.nytimes.com/2021/04/08/technology/taiwan-drought-tsmc-semiconductors.html (Visitado el 18 de noviembre de 2024).

4. Aceleracionismo militar: Inteligencia Artificial, *big tech* y el genocidio en Gaza

NoTechForApartheid
Traducción: Camino Villanueva Rodríguez

La intersección entre la inteligencia artificial y la industria militar es cada vez más intensa como demuestra el hecho de que la inversión en este ámbito está creciendo a un ritmo acelerado. Una buena parte de ese volumen de inversión aparece etiquetado comúnmente como inversión en el desarrollo de inteligencia artificial, pero a menudo se obvia que se enmarca en contratos relacionados con el ámbito de la defensa. Uno de los efectos que ha tenido esta alianza entre inteligencia artificial e industria militar ha sido la deshumanización de la guerra, con las consecuencias que eso tiene también a efectos, incluso, legales. Dos circunstancias marcan el progreso de esta relación: por un lado, la introducción de la inteligencia artificial en el propio sector bélico industrial; pero, por otro, también la incorporación de la tecnología de consumo, es decir, la que no está desarrollada expresamente para un uso militar, a su utilización por parte de los ejércitos.

Precisamente, la transferencia de la tecnología de consumo hacia usos relacionados con el contexto bélico es especialmente problemático. El empleo de datos de origen

civil en el entrenamiento de unos modelos que acaban en labores de defensa, o la participación de corporaciones de esa tecnología de consumo en el desarrollo de herramientas dedicadas a fines militares perfilan algunos de esos conflictos. Junto a ellos cuestiones relacionadas con una pretendida infalibilidad que no es tal o una supuesta objetividad plagada de sesgos, construyen enormes mecanismos de vigilancia, control y represión, para los que es necesario configurar nuevos marcos regulatorios y de respuesta social.

Introducción

El ataque genocida contra Gaza que comenzó en 2023 es una de las campañas militares modernas más mortíferas y destructivas. La cifra de muertes diarias ha superado la de todos los demás grandes conflictos del siglo XXI,[1] incluidas mujeres, niñas y niños.[2] Miles de personas más han quedado mutiladas por los bombardeos israelíes o han muerto de forma indirecta por enfermedad e inanición. Al menos el 90 % de la población de Gaza ha sido desplazada y más del 80 % de los edificios han quedado destruidos.[3]

Dos procesos han sido determinantes para este genocidio: la militarización de la inteligencia artificial (IA) en todos los sistemas militares, tácticos y de vigilancia; y la incorporación de las capacidades de la tecnología de consumo en las operaciones militares.

La incorporación de la tecnología de consumo en las fuerzas armadas israelíes ha permitido que sus operaciones bélicas y técnicas se intensifiquen más allá de lo que los sistemas internos podían manejar. Aunque las fuerzas militares israelíes mantienen desde hace tiempo vínculos con las grandes empresas tecnológicas, así como con los sectores militar y de la vigilancia, el contrato del Proyecto Nimbus —firmado en 2021— ha ampliado estas relaciones y ha involucrado directamente a Google y Amazon en la vigilancia masiva y el genocidio del pueblo palestino.

El Proyecto Nimbus proporciona servicios de Amazon Web Services (AWS) y Google Cloud a las fuerzas armadas y el gobierno israelíes. Las fuerzas militares israelíes utilizan profusamente los servicios en la nube.[4] La coronel Racheli Dembinsky, comandante del Centro de Computación y Sistemas de Información de las Fuerzas de Defensa de Israel (FDI), ha declarado al respecto:

> (C)on el inicio de la invasión terrestre de Gaza por parte del ejército israelí a finales de octubre de 2023, (...) los sistemas militares internos se sobrecargaron rápidamente (...) los servicios en la nube ofrecidos por las principales empresas tecnológicas permitían al ejército adquirir servidores de almacenamiento y procesamiento ilimitados con sólo pulsar un botón (...). Sin embargo, la ventaja 'más importante' que aportaban las

1 https://www.aljazeera.com/news/2024/1/11/gaza-daily-deaths-exceed-all-other-major-conflicts-in-21st-century-oxfam
2 https://www.oxfam.org/es/letters-and-statements/oxfam-afirma-que-el-ejercito-israeli-ha-asesinado-en-un-ano-mas-mujeres-y
3 https://www.pbs.org/newshour/world/90-of-gaza-residents-have-been-displaced-by-israels-evacuation-orders-un-says, https://www.middleeastmonitor.com/20240820-un-over-80-of-gazas-buildings-destroyed/
4 https://www.972mag.com/cloud-israeli-army-gaza-amazon-google-microsoft/ (traducción propia)

> empresas de la nube (...) eran sus funcionalidades avanzadas en materia de inteligencia artificial.

Ori Givati, veterano de las fuerzas armadas israelíes que ahora es activista, afirma que la capacidad de Israel para procesar grandes cantidades almacenadas de datos a fin de vigilar a la población palestina forma parte de la ocupación.[5] Este potencial incluye desde hace tiempo la IA, al menos en alguna de sus formas. El conjunto de sistemas de vigilancia Wolf Pack utiliza el reconocimiento facial para censar e identificar a las personas palestinas, a menudo sin su conocimiento ni consentimiento.[6] Éste y otros sistemas de vigilancia invasiva alimentan las redes de almacenamiento masivo de datos a largo plazo que proporcionan los datos de entrenamiento de los sistemas de inteligencia operativa basados en la IA. Desde 2023 han surgido múltiples sistemas de selección de objetivos, seguimiento y combate basados en la IA:

- Habsora (El Evangelio) es un sistema de IA que genera objetivos de ataque y facilita el funcionamiento de una "fábrica de asesinato masivo"[7]
- Lavender (Lavanda) es un listado de objetivos humanos (*kill list*) automatizada que utiliza la IA para analizar los datos recopilados, por medio de la vigilancia masiva, sobre la mayoría de quienes viven en Gaza y señalar a personas como blancos.[8]
- Where's Daddy (Dónde está papá) es un sistema automatizado utilizado para hacer un seguimiento de los objetivos (incluidos los generados por Lavender) y bombardearlos cuando entran en su domicilio familiar, con lo que a veces se mata a familias enteras.[9]
- Drones y robots exploradores con funciones automatizadas de identifi cación y seguimiento de los objetivos.[10]

Estas tecnologías intensifican la deshumanización, complican la rendición de cuentas y dirigen la guerra a una masacre totalmente automatizada. Un mando lo expresaba de esta manera:[11]

> "Luchas desde el interior del ordenador portátil". Antes, "veías el color de ojos del enemigo, mirabas por los prismáticos y lo veías explotar". Ahora, sin embargo, cuando aparece un objetivo, "les dices (a los soldados) a través del ordenador portátil: 'Disparen con el tanque'".

En este capítulo, analizamos el hecho de que las armas de IA no responden a calificativos tales como "precisas" y "objetivas", examinamos en detalle el verdadero carácter militar del Proyecto Nimbus y las mentiras que lo rodean y mostramos con ejemplos el modo en que el personal de las empresas tecnológicas ha expresado su preocupación y ha sufrido represalias y represión por pronunciarse. Por último, analizamos críticamente los listados de objetivos humanos elaborados por la IA y sostenemos que su uso constituye un crimen de guerra.

5 https://theintercept.com/2022/07/24/google-israel-artificial-intelligence-project-nimbus/
6 https://www.amnesty.org/es/documents/mde15/6701/2023/es/, https://www.972mag.com/isdef-surveillance-tech-israel-army/
7 https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/
8 https://www.972mag.com/lavender-ai-israeli-army-gaza/
9 *Ibid*.
10 https://www.businessinsider.com/israel-drone-that-can-fire-a-sniper-rifle-while-flying-developed-2022-1, https://www.idf.il/en/mini-sites/technology-and-innovation/jaguar-the-idf-s-newest-most-advanced-robot/
11 https://www.972mag.com/cloud-israeli-army-gaza-amazon-google-microsoft/ (traducción propia)

Guerra de IA

El uso de la IA para la guerra viene extendiéndose, sobre todo en el último decenio. En 2015, Siemens calculaba que el gasto militar mundial en robótica había aumentado a 7.500 millones de dólares estadounidenses, y preveía un incremento acelerado del mismo hasta alcanzar los 16.500 millones en 2025.[12] El gasto en IA del gobierno estadounidense es mayoritariamente militar: en 2022, casi el 90 % del valor de los contratos correspondía al Departamento de Defensa.[13]

El auge de las aplicaciones militares de la IA tiene su origen directo en la expansión de los aparatos de vigilancia en todo el mundo y es posible gracias a las ingentes cantidades de datos generados por el sector de la tecnología de consumo. La IA es el medio elegido por las fuerzas armadas para dar operatividad a sus datos de vigilancia, por ejemplo analizando las imágenes de vigilancia por medio de drones para encontrar personas o marcando las publicaciones en las redes sociales según los criterios que deciden. Las armas de IA proporcionan invulnerabilidad al quienes las manejan, son un medio de guerra psicológica y se apropian de la tecnología de consumo para reducir los costos.

Información general sobre IA y aprendizaje automático

La IA desempeña muchas funciones en el imaginario popular y en el discurso público. En esencia, no obstante, la IA se deriva de la estadística, utilizando grandes conjuntos de datos y mediante la optimización de parámetros a partir de una inicialización aleatoria. El modelo más utilizado es el discriminativo, llamado así porque distingue entre tipos de entradas, como adivinar si una flor es un lirio o una orquídea o qué palabra es más probable que se escriba a continuación. Estos ejemplos se representan en el modelo como un conjunto de números que describen el objeto en cuestión. Puede tratarse de "características", es decir, aspectos calculables del objeto, como el número de pétalos de una flor, o de representaciones más complejas tales como los píxeles de una imagen. A partir de muestras de datos suficientemente grandes y emparejadas con etiquetas que indican lo que debe predecir, el modelo "aprende" una función que puede distinguir con cierta precisión entre los objetos de una etiqueta concreta a partir de los datos disponibles."

Para relacionar esto con los usos militares de la IA, el comandante de la Unidad 8200 en el momento de escribir este texto describía las "características" siguientes como datos de entrada de un sistema de creación de listados de objetivos humanos:[14]

> (E)star en un grupo de WhatsApp con una persona militante conocida, cambiar de teléfono móvil cada pocos meses y cambiar de dirección a menudo.

12 https://web.archive.org/web/20180207122319/https://www.siemens.com/innovation/en/home/pictures-of-the-future/digitalization-and-software/autonomous-systems-infographic.html
13 https://www.brookings.edu/articles/the-evolution-of-artificial-intelligence-ai-spending-by-the-u-s-government/
14 https://www.972mag.com/lavender-ai-israeli-army-gaza/ (traducción propia)

Los sistemas de inteligencia artificial actuales suelen utilizar modelos grandes de lenguaje (LLM, por sus siglas en inglés), multimodales o no. Los LLM multimodales son, de hecho, modelos grandes de lenguaje que admiten archivos de imagen, audio o vídeo, además de texto, como datos de entrada o de salida: ChatGPT de OpenAI, Gemini de Google y Bedrock de Amazon son algunos ejemplos. Los LLM han ampliado enormemente el uso de la IA en la industria, pero siguen siendo sólo modelos estadísticos a gran escala que se entrenan con cantidades ingentes de datos. Gemini de Google, por ejemplo, se entrena a partir de los datos que las personas usuarias introducen en Gemini,[15] los datos públicos de Internet y posiblemente otros conjuntos privados de datos. Estos modelos aprenden en primer lugar una representación de patrones lingüísticos, y luego se afinan con datos etiquetados para dar respuestas a preguntas, seguir instrucciones básicas e incluso realizar tareas complejas como la transcripción de audio. Los conjuntos de datos necesarios para el ajuste suelen externalizarse a personal subcontratado.[16] No obstante, como crean respuestas a partir de modelos estadísticos de datos, estos modelos no ofrecen forzosamente respuestas objetivas, razonables ni fácticas.

Los sistemas de inteligencia artificial actuales suelen utilizar modelos grandes de lenguaje (LLM, por sus siglas en inglés), multimodales o no. Los LLM multimodales son, de hecho, modelos grandes de lenguaje que admiten archivos de imagen, audio o vídeo, además de texto, como datos de entrada o de salida: ChatGPT de OpenAI, Gemini de Google y Bedrock de Amazon son algunos ejemplos. Los LLM han ampliado enormemente el uso de la IA en la industria, pero siguen siendo sólo modelos estadísticos a gran escala que se entrenan con cantidades ingentes de datos. Gemini de Google, por ejemplo, se entrena a partir de los datos que las personas usuarias introducen en Gemini,[17] los datos públicos de Internet y posiblemente otros conjuntos privados de datos. Estos modelos aprenden en primer lugar una representación de patrones lingüísticos, y luego se afinan con datos etiquetados para dar respuestas a preguntas, seguir instrucciones básicas e incluso realizar tareas complejas como la transcripción de audio. Los conjuntos de datos necesarios para el ajuste suelen externalizarse a personal subcontratado.[18] No obstante, como crean respuestas a partir de modelos estadísticos de datos, estos modelos no ofrecen forzosamente respuestas objetivas, razonables ni fácticas.

A continuación presentamos dos casos importantes de fallos relacionados con los listados de objetivos humanos elaborados por la IA, que esta información general permite comprender.

Etiquetas erróneas en los datos de entrenamiento

Una fuente que trabajó con el equipo de ciencia de datos militares en Lavender declaró lo siguiente:[19]

> "Me preocupó el hecho de que cuando se entrenó Lavender, utilizaron el término 'agente de Hamás' de forma poco precisa e incluyeron en el conjunto de datos de entrenamiento a personas que trabajaban en la defensa civil".

15 https://www.searchenginejournal.com/google-gemini-privacy-warning/507818/
16 https://www.engadget.com/ai/google-accused-of-using-novices-to-fact-check-geminis-ai-answers-143044552.html
17 https://www.searchenginejournal.com/google-gemini-privacy-warning/507818/
18 https://www.engadget.com/ai/google-accused-of-using-novices-to-fact-check-geminis-ai-answers-143044552.html
19 https://www.972mag.com/lavender-ai-israeli-army-gaza/ (traducción propia)

En este caso, por "trabajar en la defensa civil" se entiende rescatar a personas y cadáveres de entre los escombros tras los bombardeos.[20] Un modelo entrenado a partir de estos datos propaga esos errores en la inferencia, lo que podría considerarse una violación del principio de distinción (es decir, determinaría la comisión de un crimen de guerra).

De forma más sutil, el conjunto de datos de entrenamiento puede estar sesgado en cuanto a la recopilación de los datos (por ejemplo, la muestra que se elige), la metodología de etiquetado y el tratamiento de las características de los datos de entrada.

Errores de distribución del conjunto de datos

Los modelos sólo pueden funcionar a partir de los datos que se les proporcionan: no pueden interpretar el contexto fuera de los datos de entrenamiento y, por tanto, no pueden tomar en cuenta circunstancias o información nuevas. Prueba de ello es que casi siempre obtienen peores resultados en situaciones reales que en las pruebas. Los modelos se entrenan para reducir al mínimo los errores existentes en la información de entrenamiento mediante la búsqueda de correlaciones de datos. Cuando el entorno en el que operan cambia, las correlaciones subyacentes se modifican y pierden precisión.

En el caso concreto de Lavender, el hecho de que la gente sufra bombardeos y se vea obligada a huir de sus casas supone un cambio de los datos de entrada del modelo. Una fuente israelí lo expresaba así: "En la guerra, los palestinos cambian de teléfono continuamente".[21]

Aplicaciones militares y aplicaciones de consumo

En contextos de consumo, la IA ha encontrado el mayor grado de aceptación en productos en los que equivocarse en la respuesta no tiene consecuencias demasiado graves y los aciertos son muy preciados, como la investigación, el plegamiento de proteínas y el descubrimiento de fármacos. Por poner un ejemplo de contraste, la ley sobre el suicidio asistido de California (Estados Unidos) exige que dos profesionales de la medicina determinen que la persona en cuestión sufre una enfermedad terminal con un pronóstico de seis meses o menos de vida y está en posesión de sus facultades mentales. Aunque existen debates legítimos sobre el grado de libre determinación que se permite, está claro que es sumamente importante que el diagnóstico de enfermedad terminal sea preciso y, por tanto, en general se admite que ésta no es una aplicación ética de la IA. Los listados de objetivos humanos elaborados por la IA deben entrar en esta categoría. No debemos presumir que estas listas son más sofisticadas que, por ejemplo, los anuncios que recibimos: posiblemente lo son menos.

Caracterización equívoca de las armas de IA

Las nuevas tecnologías relacionadas con las armas de IA se implantan a menudo con un barniz de modernidad y se presentan como de vanguardia, lo que refleja el lenguaje utilizado por las marcas de la tecnología de consumo para dibujar una visión del carácter irrefutable de los avances tecnológicos.

20 https://www.euronews.com/2024/07/13/civil-defence-workers-recover-60-bodies-from-rubble-in-two-districts-of-gaza-city
21 https://www.972mag.com/lavender-ai-israeli-army-gaza/ (traducción propia)

El ejército israelí[22] (como anteriormente el ejército estadounidense[23]) suele describir los homicidios automatizados como "precisos" o "quirúrgicos", algo que la realidad contradice:[24]

> "A la hora de atacar a presuntos militantes de bajo rango señalados por Lavender, el ejército prefería utilizar únicamente misiles no guiados, conocidos comúnmente como bombas 'tontas' (por oposición a las bombas 'inteligentes' de precisión), que pueden destruir edificios enteros sobre sus ocupantes y causar un número considerable de víctimas mortales. 'No se quieren malgastar bombas caras con gente sin importancia; le sale muy caro al país y hay escasez (de esas bombas)', señaló C., uno de los mandos de inteligencia".

Tras los ataques del 7 de octubre de 2023, el ejército israelí decidió que "por cada agente de bajo rango de Hamás que Lavender señalaba, se permitía a matar a hasta 15 o 20 civiles". El ejército estadounidense, que hizo un uso considerable de drones por primera vez en la guerra de Irak, también permitió cifras altas de víctimas mortales civiles en Pakistán,[25] Afganistán,[26] Somalia,[27] y Yemen.[28]

La política de aumentar el número de víctimas mortales civiles admisibles obedece al tipo de tecnología utilizada y, en el contexto militar, pone de manifiesto la profunda deshumanización que estos sistemas aplican a sus víctimas.

La automatización de los homicidios también se anuncia como más "objetiva". Un alto mando afirmó sobre Lavender:[29]

> "Confío mucho más en un mecanismo estadístico que en un soldado que ha perdido a un amigo dos días antes. Todo el mundo allí, incluido yo, había perdido a alguien el 7 de octubre. La máquina lo hizo con frialdad, y eso hizo que resultara más fácil".

Los modelos de IA siguen empleándose en un sistema más amplio dirigido por personas. En el caso de Gaza, la operación está impulsada por la venganza, el odio y, posiblemente, el deseo de apoderarse de tierras o recursos:[30] por ejemplo, en Lavender se redujeron los umbrales para maximizar la destrucción; así lo explicó un mando:[31]

> "En los días que no había objetivos (cuyo índice de características fuera suficiente para autorizar un ataque), atacábamos con un umbral más bajo. Nos presionaban constantemente: 'Traednos más objetivos'. Nos gritaban de verdad. Terminábamos (matando) a nuestros objetivos muy rápido".

En la guerra de drones estadounidense, los mandos militares también se han centrado en el "recuento de objetivos humanos" como medida del éxito operativo, y de forma similar se menciona que la mayoría de los ataques se dirigen contra agentes de bajo grado.[32]

22 https://www.cbsnews.com/news/israel-military-ground-operation-al-shifa-hospital-gaza-hamas/
23 https://journals.sagepub.com/doi/abs/10.1177/0263276411423027
24 https://www.972mag.com/lavender-ai-israeli-army-gaza/ (traducción propia)
25 https://www.newamerica.org/future-security/reports/americas-counterterrorism-wars/the-drone-war-in-pakistan/
26 https://web.archive.org/web/20180624010404/https://www.thebureauinvestigates.com/projects/drone-war/afghanistan
27 https://www.newamerica.org/future-security/reports/americas-counterterrorism-wars/the-war-in-somalia/
28 https://www.newamerica.org/future-security/reports/americas-counterterrorism-wars/the-war-in-yemen/
29 https://www.972mag.com/lavender-ai-israeli-army-gaza/ (traducción propia)
30 https://www.amnesty.org/en/wp-content/uploads/sites/4/2024/12/MDE1587442024SPANISH.pdf
31 https://www.972mag.com/lavender-ai-israeli-army-gaza/ (traducción propia)
32 https://theintercept.com/2021/10/24/drone-war-books-neil-renic-wayne-phelps/

De estos ejemplos se desprende que, en lugar de proporcionar precisión u objetividad, la principal consecuencia de los sistemas de selección de objetivos basados en la IA es maximizar la destrucción e impedir que se vea el papel que desempeñan los prejuicios y la toma de decisiones humana en las operaciones de bombardeo.

Razones del uso de armas de IA

Las fuerzas armadas tratan de sacar provecho de los sistemas de armas de IA de diversas maneras. Concretamente, los drones controlados por IA permiten a quienes los manejan lanzar ataques sin poner en peligro su propia vida. Esto puede reducir notablemente el costo social de la guerra, y facilitar que se mantenga el apoyo a las acciones bélicas.

De forma un poco menos evidente, el personal estadounidense que maneja drones se familiariza en cierta medida con sus objetivos, puesto que los siguen durante un tiempo prolongado.[33] Los drones cuentan con dispositivos de vigilancia de alta resolución, que permiten a quienes los manejan:[34]

> "(V)er al objetivo de cerca, ver lo que le ocurre durante la explosión y las secuelas (...) se está más lejos físicamente, pero se ve más".

Por la descripción de Lavender y Where's Daddy,[35] parece que al personal israelí que maneja drones le lleva menos tiempo, ya que la selección y el seguimiento de los objetivos están más automatizados. Sin embargo, las funcionalidades de vigilancia de estos sistemas no son menores, por lo que quienes los manejan tienen la facultad de matar a sus víctimas con un esfuerzo mínimo o con el grado de implicación que deseen.

Como analizaremos más adelante, las armas de IA también permiten aprovechar eficazmente la tecnología de consumo.

Vigilancia y terror

La tecnología de vigilancia es la base de los listados de objetivos humanos elaborados por la IA y de las actividades de ocupación utilizadas para mantener las condiciones de apartheid, como las paradas en los puestos de control.

En congresos tecnológicos tales como la Exposición de Defensa de Israel suelen participar empresas emergentes de vigilancia[36] que prometen transformar la ocupación y la guerra por medio de una mayor eficacia. Las autoridades israelíes han tratado de presentar la ocupación que llevan a cabo como "ilustrada", gracias a las soluciones técnicas que podrían "reducir el conflicto". En la práctica, el número de paradas aumentó con los sistemas de AI debido a su dependencia de grandes cantidades de datos:[37]

33 *Ibid.*
34 https://journals.sagepub.com/doi/abs/10.1177/0263276411423027 (traducción propia)
35 https://www.972mag.com/lavender-ai-israeli-army-gaza/
36 https://www.972mag.com/isdef-surveillance-tech-israel-army/
37 https://www.cambridge.org/core/journals/international-journal-of-middle-east-studies/article/algorithmic-state-violence-automated-surveillance-and-palestinian-dispossession-in-hebrons-old-city/80B9C5192057ACEA17089E488CFC1486 (traducción propia)

> "(C)on Blue Wolf, esos altercados (las paradas) se intensificaron. Según expresaban los soldados, el ejército utilizaba pequeños concursos para incentivar a las brigadas a recopilar la mayor cantidad de datos posible".

Esta forma de proceder revela la dependencia recíproca de la IA y la recopilación de datos de vigilancia. Una mayor vigilancia necesita la IA para procesar los datos de forma eficaz, mientras que las ventajas percibidas de la IA sirven para justificar una labor de vigilancia más intensa. El objetivo de estos sistemas gemelos es proporcionar una sensación de poder y control a las fuerzas de ocupación por el aumento de su capacidad para apreciar y predecir posibles hostilidades, al tiempo que funcionan como una forma de guerra psicológica contra las víctimas de la violencia del Estado. Los sistemas automatizados no son totalmente aleatorios —es decir, se adaptan al menos en parte a nuestros actos— y son también lo suficientemente imprecisos como para provocar detenciones indebidas e incluso casos de detención arbitraria o tortura.[38] Esto infunde terror a la población oprimida: lo que se hace puede activar los sistemas automatizados, pero no está claro exactamente de qué manera. Cualquier comportamiento, incluidas las publicaciones en las redes sociales, puede aumentar el peligro de detención, lo que provoca un efecto disuasorio. En Israel, estos sistemas de vigilancia encuentran una justificación adicional en sus efectos sobre la extracción de valor de la fuerza laboral y la ciudadanía palestinas. La fuerza laboral palestina constituye una parte sustancial de la economía israelí[39], y el alcance de los sistemas de vigilancia establecidos genera una mano de obra ya disciplinada para la ocupación militar; y la información de vigilancia recopilada enriquece de manera simultánea a la parte ocupante al proporcionar los datos de entrenamiento de los sistemas de IA que pueden venderse en todo el mundo.

Tecnología de consumo

La vigilancia por medio de IA y las armas de IA hacen un uso más directo de la tecnología de consumo que muchos otros tipos de tecnologías militares. Por ello, los ejércitos de Occidente dependen cada vez más de establecer contratos con las grandes empresas de tecnología de consumo. Las principales compañías proveedoras de servicios en la nube podrían ser las únicas organizaciones que disponen de la infraestructura, la capacidad y los datos necesarios para entrenar modelos de IA de última generación que pueden adaptarse para uso militar; y la facilidad para adquirir servicios, almacenamiento y recursos informáticos de alta gama ya existentes es una gran ventaja para las fuerzas armadas que amplían sus actividades. Así lo expresaba una fuente de inteligencia:[40]

> "(Las empresas de la nube) también tienen sus propios mecanismos de conversión de voz a texto. Y son buenos, tienen muchas funcionalidades. ¿Qué sentido tiene desarrollarlo todo en la unidad del ejército si esas funcionalidades ya existen?".

Las empresas tecnológicas estadounidenses tienen una larga trayectoria de interdependencia con los ejércitos de Occidente. In-Q-Tel, fundada en 1999, es una empresa de capital riesgo sin ánimo de lucro creada para derivar fondos de la CIA al sector privado y transferir de nuevo estas tecnologías a los organismos de defensa e inteligencia estadounidenses. Cada vez hay más de estas empresas que se dedican a la vigilancia y minería de las redes sociales, como Dataminr, Geofeedia, PATHAR y TransVoyant.[41]

38 https://reliefweb.int/report/occupied-palestinian-territory/welcome-hell-israeli-prison-system-network-torture-camps-enhe
39 https://www.nbcnews.com/news/world/israel-hamas-india-labor-shortage-migrant-workers-rcna135603
40 https://www.972mag.com/cloud-israeli-army-gaza-amazon-google-microsoft/ (traducción propia)
41 https://theintercept.com/2016/04/14/in-undisclosed-cia-investments-social-media-mining-looms-large/

El Proyecto Maven, iniciado en 2017, supuso un cambio en el modo en que el Departamento de Defensa de Estados Unidos se integraba con las empresas de tecnología de consumo e indica un giro pronunciado hacia la priorización de la IA. El contrato de este proyecto utilizó el trabajo de profesionales de la ingeniería del sector privado para entrenar la IA en la identificación de objetos a partir de datos de vigilancia militar, y en él han participado al menos 21 empresas privadas.[42] No obstante, en 2018, Google optó por no renovar el contrato debido a las protestas llevadas a cabo por su personal. Tras esta situación problemática para la venta de productos del sector privado, aparecieron dos tendencias destacables en el sector tecnológico: las empresas tecnológicas hacían declaraciones públicas sobre la ética de la IA, y comenzaron a aplicar medidas estrictas para limitar la influencia de su personal en las decisiones de la compañía.[43]

Tres años después de que el personal protestara contra el Proyecto Maven se firmó el Proyecto Nimbus. Al año siguiente se anunció el contrato Capacidad Bélica Conjunta en la Nube (JWCC, por sus siglas en inglés), por valor de 9.000 millones de dólares estadounidenses para soluciones militares de IA, que se adjudicó a Amazon Web Services, Google, Microsoft y Oracle.[44] Este contrato es muy similar al Proyecto Nimbus en el sentido de que proporciona servicios en la nube con condiciones específicas que permiten utilizar las tecnologías para el combate bélico. Mientras que las extracciones más directas de trabajo de ingeniería se estrellaron contra la resistencia del personal y el daño a la reputación, esta transformación del contrato ha dado muestras de resistencia: se hace pasar como una simple venta de servicios de consumo, que otorga a las fuerzas armadas acceso al trabajo acumulado de la ingeniería del sector privado, y permite a las empresas tecnológicas acceder a los presupuestos de gasto militar manteniendo la apariencia de imparcialidad.

Además, la percepción de que el Proyecto Nimbus es un contrato común de informática en la nube e inteligencia artificial de consumo ha permitido a Google y Amazon encubrir las implicaciones militares del acuerdo y evitar un mayor control, sin dejar de facilitar el uso bélico avanzado de su tecnología.

Proyecto Nimbus

El Proyecto Nimbus es un proyecto emblemático del gobierno israelí para proporcionar infraestructura y servicios en la nube. Google y Amazon fueron escogidas en 2021 como empresas proveedoras de la nube en un contrato valorado en 1.200 millones de dólares. Pese a las numerosas declaraciones engañosas publicadas por Google, el acuerdo es principalmente de carácter militar.

Recientemente se ha revelado que al menos el 70 % de los ingresos que Google prevé obtener del Proyecto Nimbus proceden del ejército israelí:[45]

> "Según las condiciones del acuerdo, Google esperaba obtener la mayor parte del dinero del Ministerio de Defensa de Israel, alrededor de 525 millones de dólares de 2021 a 2028, cifra muy superior a los 208 millones que esperaba recibir del resto del gobierno central del país".

Nota: del Ministerio de Defensa dependen las FDI, Israel Military Industries e Israel Aerospace Industries.

42 https://en.wikipedia.org/wiki/Project_Maven
43 https://www.businessinsider.com/google-thanksgiving-four-trial-protest-2021-8
44 https://harpers.org/archive/2024/03/the-pentagons-silicon-valley-problem-andrew-cockburn/
45 https://www.nytimes.com/2024/12/03/technology/google-israel-contract-project-nimbus.html (traducción propia)

Además, diversos miembros de las FDI han confirmado el valor militar del Proyecto Nimbus, en particular:

- Varias fuentes militares afirman que "(la información de) la vigilancia de toda la población palestina de Gaza es tan amplia que no puede almacenarse sólo en los servidores militares".[46] Yossi Sariel, comandante de la Unidad 8200 de las fuerzas armadas israelíes (que desarrolla Lavender) en el momento de elaboración de este informe, declaró que tal cantidad de información puede almacenarse "únicamente en empresas como Amazon, Google o Microsoft".[47]

- Dos de las principales empresas estatales fabricantes de armas de Israel, Israel Aerospace Industries y Rafael Advanced Defense Systems, tienen la obligación de utilizar Amazon y Google, a través de Nimbus, para sus necesidades informáticas en la nube.[48]

- Gaby Portnoy, jefe de la Dirección Cibernética Nacional de Israel, declaró: "Están ocurriendo cosas extraordinarias en el campo de batalla gracias a la nube pública Nimbus, cosas que influyen mucho en la victoria (...)".[49]

- En la conferencia Tecnologías de la Información para las FDI 2024, la coronel Racheli Dembinsky afirmó que la informática en la nube es "un arma en todos los sentidos de la palabra",[50] y ofreció una descripción explícita del uso que las FDI hacen de las nubes de Google, Amazon y Microsoft.[51]

- Thomas Kurian, director general de Google Cloud, anunció la integración de Vertex AI con Palantir,[52] "la traficante de armas de IA del siglo XXI".[53] Google también ofrece acceso al software Foundry de Palantir a quienes utilizan Nimbus.[54]

Aunque no tenemos pruebas que vinculen directamente el Proyecto Nimbus con Lavender, esta relación tiene encaje técnico, en el sentido de que Lavender utiliza datos de vigilancia masiva, que a menudo se procesan con IA en la nube. El gobierno israelí pretende migrar los proyectos gubernamentales a la nube;[55] los proyectos más secretos se mantienen en servidores militares, pero algunos proyectos de inteligencia se alojan en ella.[56] El factor limitativo hasta el momento es la falta de centros de datos con medidas de seguridad totales para ejecutar los sistemas operativos más confidenciales; sin embargo, hay pendientes varios contratos para posibilitar este tipo de procesamiento mediante acuerdos similares con empresas de tecnología de consumo. Google también está ofreciendo su LLM Gemini a la policía israelí y al funcionariado de la seguridad nacional.[57] Entre otros usos, los modelos de texto pueden determinar que se cause daño en forma de vigilancia de las redes sociales, que ya se utiliza ampliamente como base para detener a personas palestinas.[58]

46 https://www.972mag.com/cloud-israeli-army-gaza-amazon-google-microsoft/ (traducción propia)
47 *Ibid.*
48 https://theintercept.com/2024/05/01/google-amazon-nimbus-israel-weapons-arms-gaza/
49 https://www.wired.com/story/amazon-google-project-nimbus-israel-idf/
50 https://www.404media.co/google-cloud-listed-then-removed-as-sponsor-of-israeli-military-tech-conference/
51 https://www.youtube.com/watch?v=qLBDfnZJrC8 (traducción propia)
52 https://jackpoulson.substack.com/p/microsoft-and-google-have-been-working, https://cloud.google.com/blog/topics/partners/google-cloud-and-palantir-announce-annalytics-partnership
53 https://www.theverge.com/2024/8/8/24216215/palantir-microsoft-azure-ai-defense-partnership-surveillance (traducción propia)
54 https://theintercept.com/2024/05/01/google-amazon-nimbus-israel-weapons-arms-gaza/
55 https://www.gov.il/en/pages/press_24052021
56 https://www.972mag.com/cloud-israeli-army-gaza-amazon-google-microsoft/
57 https://www.wired.com/story/amazon-google-project-nimbus-israel-idf/
58 https://www.adalah.org/en/content/view/10959

Encubrir la complicidad empresarial

El Proyecto Nimbus representó un hito importante para el negocio de Google Cloud, y brindó a Google la posibilidad de aumentar la cuota de mercado de Cloud gracias al elevado porcentaje de los presupuestos gubernamentales que se destinan a la IA militar. También supone un punto de inflexión en la forma en que las plataformas prestadoras de servicios en la nube de las grandes empresas tecnológicas en general pueden convertirse en contratistas de las fuerzas armadas, aunque con un costo para su reputación y cultura laboral.
En su afán de lucro, Google ha recurrido a una dura represión de su personal y al engaño público para eludir su responsabilidad por los daños que sus productos facilitan. Cuando se le ha preguntado por el Proyecto Nimbus, Google ha repetido dos alegaciones concretas:[59]

> "Este trabajo no está orientado a tareas militares, clasificadas ni muy confidenciales que tengan relevancia con respecto a las armas o los servicios de inteligencia",

y[60]

> "(e)l contrato de Nimbus tiene por objeto las tareas que los ministerios gubernamentales israelíes, que han aceptado cumplir nuestras condiciones de servicio y nuestra política de uso aceptable, ejecutan en nuestra nube comercial".

En una declaración enviada por correo electrónico, Google afirmaba que se mencionaban de forma explícita las condiciones del servicio y la política de uso aceptable habituales de Cloud.[61] El examen llevado a cabo de la licitación de Nimbus revela que esta información es falsa; en realidad, las personas usuarias de Nimbus sólo tienen que cumplir unas condiciones del servicio adaptadas.[62] En la licitación se afirma que todos los servicios, incluida la IA avanzada, deben ponerse a disposición de todos los poderes públicos. El Ministerio de Hacienda israelí se negó a hacer públicas estas condiciones de servicio ajustadas.[63]

Además, Google ha rehusado hacer cumplir sus condiciones del servicio. La inteligencia militar israelí creó una "lista negra" de presuntos militantes tras los atentados del 7 de octubre de 2023, utilizando el reconocimiento facial a través de Corsight (empresa que contrata a personas del entorno de Amazon y Google Cloud[64]) y Google Photos. Los sospechosos, incluidos los identificados erróneamente, fueron detenidos y sometidos a malos tratos.[65] Pese a que este daño constituía un incumplimiento de las políticas de uso aceptable de Google Photos, Google rehusó hacer cumplir sus políticas.[66]

59 https://www.wired.com/story/amazon-google-project-nimbus-israel-idf/, https://theintercept.com/2024/05/01/google-amazon-nimbus-israel-weapons-arms-gaza/, https://time.com/7013685/google-ai-deepmind-military-contracts-israel/ (traducción propia)
60 *Ibid.,* https://theintercept.com/2024/05/01/google-amazon-nimbus-israel-weapons-arms-gaza/, https://www.nytimes.com/2024/12/03/technology/google-israel-contract-project-nimbus.html (traducción propia)
61 https://www.wired.com/story/amazon-google-project-nimbus-israel-idf/
62 https://theintercept.com/2024/12/02/google-project-nimbus-ai-israel/
63 https://bsky.app/profile/sambiddle.com/post/3lcxu3ty7xc2h
64 https://archive.ph/HvcJ6#selection-957.0-957.71
65 https://www.nytimes.com/2024/03/27/technology/israel-facial-recognition-gaza.html, https://mondoweiss.net/2024/01/the-shocking-inhumanity-of-israels-crimes-in-gaza/
66 https://theintercept.com/2024/04/05/google-photos-israel-gaza-facial-recognition/

Nimbus no es ni mucho menos el primer caso en que Google miente a su personal. La empresa afirmó que la tecnología de su Proyecto Maven no se utilizaba para identificar a personas y que no se habían adaptado ningún servicio de IA. Sin embargo, el código reveló que, en realidad, se etiquetaban personas y vehículos. Se pagaba a personal subcontratado para que etiquetara las imágenes satelitales, sin informarle de su finalidad militar.[67] Google mintió también al afirmar que su proyecto del buscador Dragonfly se encontraba "en fase de prospección".[68] Y Thomas Kurian mintió al decir que Google Cloud no se utilizaría en la frontera sur estadounidense;[69] posteriormente se descubrió que la Oficina de Aduanas y Protección Fronteriza de Estados Unidos utilizaba esa plataforma para procesar las imágenes de las torres de vigilancia fronteriza.[70]

El encubrimiento de los contratos militares de tecnología informática en la nube y del uso de la IA de consumo permite a las empresas alegar que respetan los compromisos éticos mientras siguen facilitando el uso bélico avanzado de su tecnología. Los principios de Google en materia de IA se implantaron en 2018 tras las protestas contra el Proyecto Maven (probablemente para calmar a su personal)[71] y han proliferado en el sector, como en el caso de Amazon[72] y Microsoft.[73]

Estas normas funcionan como medidas de relaciones públicas, pero están concebidas para ser inoperantes a la hora de impedir realmente que sus productos faciliten causar daño. Google tiene contratado al personal que lleva a cabo los exámenes sobre la aplicación de los principios de la IA y puede despedirlo si supone un obstáculo. En 2022, una persona portavoz de Google dijo a DefenseOne que los principios de IA de la empresa:[74]

> "(S)e aplican a la actividad de la IA adaptada, no al uso general de los servicios de Google Cloud (...) Significa que las fuerzas armadas pueden utilizar nuestra tecnología de forma bastante amplia".

Muchos productos de infraestructura de IA de consumo pueden utilizarse o reutilizarse como parte de una "cadena de objetivos humanos" de las fuerzas armadas, a veces de forma aún más encubierta si se emplea una empresa proveedora externa.

La mayoría de quienes se incorporan a una empresa como Google creen que trabajan en un producto de consumo como Google Search, Google Photos o Google Cloud. Lamentablemente, las decisiones de la dirección han implicado al personal de la empresa en la ocupación militar y el genocidio que lleva a cabo Israel. Aunque el mayor grado de responsabilidad y culpabilidad recae en la dirección, la pura verdad es que el trabajo del equipo de ingeniería se ha convertido en un arma. Dado que Google proporciona al ejército israelí la gama completa de productos para empresas de Google Cloud mediante el Proyecto Nimbus, muchas personas que trabajan en Google tienen motivos justificados para temer que su trabajo sea una pieza del ataque genocida de Israel contra Gaza.

El personal de Google ha intentado expresar sus motivos de preocupación por muchas vías, especialmente cauces "adecuados" tales como las herramientas internas de presentación de informes, el traslado de la información a la dirección, el envío de correos electrónicos al equipo directivo, la formulación de preguntas durante las reuniones generales[75] y las peticiones internas, dirigidas incluso al programa de derechos humanos de Google.

67 https://theintercept.com/2019/02/04/google-ai-project-maven-figure-eight/
68 https://theintercept.com/2018/08/17/internal-meeting-reveals-how-google-bosses-misled-staff-on-their-china-censorship-plan-here-are-the-questions-they-must-answer/ (traducción propia)
69 https://www.cnbc.com/2020/10/30/google-cloud-ceo-kurian-to-employees-not-working-on-border-wall.html
70 https://theintercept.com/2022/07/24/google-israel-artificial-intelligence-project-nimbus/
71 https://www.wired.com/beyond-the-beyond/2018/06/googles-ai-principles/
72 https://sustainability.aboutamazon.com/human-rights/principles
73 https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl
74 https://www.defenseone.com/technology/2022/06/new-google-division-will-take-aim-pentagon-battle-network-contracts/368691/
75 https://time.com/7013685/google-ai-deepmind-military-contracts-israel/

Pese a estos esfuerzos, la dirección se ha negado sistemáticamente a colaborar, minimizando y negando las preocupaciones del personal,[76] mintiendo sobre las conexiones militares del contrato y mostrando un sesgo claro en contra de las voces palestinas y otras aliadas en la empresa.[77]

Ante la reacción hostil de la empresa, el personal constituyó la campaña No al Uso de la Tecnología en el Apartheid,[78] difundió peticiones firmadas por más de 1.000 personas empleadas,[79] encabezó varias protestas en 2022[80] y se manifestó contra las conferencias Cloud Next[81] y Mind the Tech, esta segunda centrada en Israel.[82] En abril de 2024 se organizaron sentadas simultáneas en las oficinas de Nueva York y Sunnyvale, así como concentraciones públicas.[83]

La dirección de Google hizo que se detuviera a nueve personas y despidió al menos a 51.[84] Además, hubo dimisiones en protesta por los despidos.[85] Se trata de un caso especialmente notorio de una constante que se refleja en todo el sector tecnológico, como Microsoft, que también vende tecnología de la nube al ejército israelí y adoptó medidas contra el personal que había organizado una vigilia por la población palestina martirizada.[86]

Del mismo modo que proporciona tecnología para la labor de vigilancia y las operaciones militares israelíes, Google favorece un entorno laboral hostil para las personas musulmanas, árabes y palestinas.[87] El despido en particular de las voces palestinas constituye un incumplimiento de las directrices sobre diligencia debida en materia de derechos humanos establecidas por la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos.[88]

La represión de las opiniones del personal, el engaño a la prensa y la ciudadanía y los continuos negocios militares secretos de Google evidencian todo lo que las grandes empresas tecnológicas están dispuestas a hacer para sacar provecho de la guerra y el genocidio. Sacan provecho de la percepción de que su IA es tecnología de vanguardia, aunque su aplicación práctica suponga matar a niños y niñas.

Quienes trabajan en estas empresas han defendido y siguen defendiendo su derecho a que su trabajo no se emplee en el genocidio, pero son objeto de una fuerte represión. Su labor de protesta puede tener mucha fuerza, porque aprovecha su propia participación en las repercusiones del trabajo que realizan, y aumenta la posibilidad de cambio en el lugar donde se produce la complicidad. Las vías para mitigar el daño que causa el armamento basado en la IA deben incluir un aumento de las medidas de protección laboral, de modo que las personas empleadas puedan solidarizarse entre sí.

76 https://www.middleeasteye.net/news/project-nimbus-israel-apartheid-google-amazon-protests
77 https://theintercept.com/2023/11/15/google-israel-gaza-nimbus-protest/
78 https://www.instagram.com/notechforapartheid/
79 https://www.instagram.com/jewishvoiceforpeace/p/CVQb8CWpMj7/?img_index=1
80 https://www.forbes.com/sites/richardnieva/2022/09/09/google-and-amazon-protest-project-nimbus-ai-contract-israel/?sh=45d147e5d162
81 https://www.latimes.com/business/story/2023-08-29/google-cloud-employees-protest-israeli-military-contract
82 https://time.com/6964364/exclusive-no-tech-for-apartheid-google-workers-protest-project-nimbus-1-2-billion-contract-with-israel/
83 https://www.latimes.com/business/story/2024-04-16/google-israel-sit-ins-project-nimbus
84 https://time.com/6964364/exclusive-no-tech-for-apartheid-google-workers-protest-project-nimbus-1-2-billion-contract-with-israel/, https://apnews.com/article/google-israel-protest-workers-gaza-palestinians-96d2871f1340cb84c953118b7ef88b3f
85 https://www.jpost.com/arab-israeli-conflict/gaza-news/guardian-of-the-walls-the-first-ai-war-669371
86 https://apnews.com/article/microsoft-fired-workers-israel-palestinians-gaza-72de6fe1f35db9398e3b6785203c6bbf
87 https://medium.com/@notechforapartheid/googleopenletter-868f0c4477db
88 https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_SP.pdf

Moralidad y legalidad

Sistemas tales como Habsora, Lavender y Where's Daddy provocan aversión, y debemos mantener esa repulsión fundamental a la matanza sin sentido. Sin embargo, estas cuestiones se despachan también a veces con dichos del tipo "la guerra es un infierno" y, para tener una visión más clara, debemos analizar el sistema más a fondo. En este apartado, intentamos situar las armas de IA en un marco bélico más general, y estudiar si estas tecnologías, aunque encubran la rendición de cuentas, siguen vulnerando la legislación vigente.

Situar las armas de la IA

Los listados de objetivos humanos elaborados por la IA son una manifestación de la asimetría de poder y la dinámica racial. Todas las guerras son terribles, pero en otros casos, una de las facciones podría alegar que representa a la población o esperar integrarla. En el contexto del genocidio de Gaza, esta dinámica queda ilustrada por el comentario de Yoav Gallant: "Luchamos contra animales humanos".[89] La segregación racial y las condiciones de *apartheid* que aplica Israel afianzan la condición de inferioridad de las personas palestinas. La facilidad y la rapidez para matar que ofrece la IA son expresiones intencionales de este carácter desechable de la vida, y consideramos que el uso de este tipo de tecnologías es una vulneración de la legislación actual y un ámbito que merece nuevas normas.

Podemos considerar estos aspectos relacionados con las armas de IA:

- Lavender clasifica con eficacia a toda la población de Gaza, mediante la vigilancia masiva, para establecer prioridades acerca de a quién conviene matar primero.
- Los vehículos aéreos no tripulados utilizados en la guerra de drones no son punteros en cuanto a velocidad, alcance ni furtividad.[90]

Podría decirse que estos elementos encarnan un enfoque brutal de la lucha antiterrorista, considerando que las propias fuerzas israelíes aterrorizan a una población en su mayoría civil durante un largo periodo. Las guerras actuales de drones son conflictos asimétricos, en los que se utilizan armas avanzadas con un efecto devastador contra personas que no tienen acceso a ellas.

Aunque algunas armas basadas en IA, como los drones, pueden ser inevitables en el marco de las "carreras armamentísticas" y su regulación inviable, existe una diferencia clara entre los sistemas dirigidos por la inteligencia humana o que son verosímilmente necesarios para evitar la derrota en el campo de batalla, y los sistemas como Lavender, que sustituyen la toma de decisiones humana. Los sistemas como Lavender pueden y deben prohibirse.

Comparación con el trabajo de inteligencia no realizado por la IA

Cabe preguntarse si los listados de objetivos humanos elaborados por la IA son peores que las decisiones tomadas por los mandos de inteligencia. Ambas situaciones pueden ser inmorales o, desde un punto de vista jurídico, pueden implicar crímenes de guerra tales como la violación de los principios de distinción y proporcionalidad. En este caso, no obstante, parece que el sistema de IA es realmente peor en cuanto a precisión, estructura y daño general causado. Una de las principales formas en que aumenta la destrucción es porque es más rápido. Un ex jefe del Estado Mayor de las FDI dijo acerca de Habsora:[91]

89 https://www.youtube.com/watch?v=ZbPdR3E4hCk
90 https://es.wikipedia.org/wiki/General_Atomics_MQ-1_Predator#Specifications, https://es.wikipedia.org/wiki/General_Atomics_MQ-9_Reaper
91 https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/ (traducción propia)

> “Antes, en Gaza llegábamos a generar hasta 50 objetivos al año. Ahora, la máquina producía 100 objetivos al día”.

Los mandos de Habsora y de Lavanda son conminados a identificar y matar a los objetivos lo más rápidamente posible. Parte de esta velocidad y magnitud proviene de la rebaja de los umbrales respecto a quién se considera objetivo, para responder a las exigencias. Esta medida y el eficaz grado de aleatoriedad aplicado para ajustarse a un “presupuesto de asesinatos” o “presupuesto de daños colaterales” deben constituir crímenes de guerra.

Exponemos esta crítica no para eximir de responsabilidad a los mandos de inteligencia, sino para sostener que los sistemas de IA plantean nuevas violaciones de los derechos humanos. Lo contrario es igual de importante: los crímenes de guerra relacionados con la IA no deben ser absueltos por la amenaza de que Estados Unidos o Israel cometan crímenes de guerra no relacionados con la IA si se ven privados en algún sentido del acceso a los recursos informáticos.

Rendición de cuentas y explicabilidad de la IA militar

Los listados de objetivos humanos elaborados por la IA también plantean nuevos retos en materia de rendición de cuentas y explicabilidad. Son producto de un conjunto más diverso de factores, que incluyen:

- El funcionariado de ciencia de datos militares, que etiqueta a las personas como militantes o civiles.
- Los mandos de inteligencia de los ejércitos, que toman la decisión final de matar, pero a quienes se ordena confiar en la IA.
- Los altos cargos empresariales, que ofrecen tecnologías civiles para uso militar.
- El personal de las empresas tecnológicas, que desarrollan las tecnologías de vigilancia subyacentes, aunque a menudo con otro propósito. Los datos de las personas usuarias suelen utilizarse, sin consentimiento informado, para entrenar o mejorar estos modelos.

Podría decirse que cada una de estas partes está menos implicada que quienes trabajaban en los servicios de inteligencia antes de que existiera la IA, que revisaban toda la información sobre alguien y decidían si era militante. Las predicciones de la IA son difíciles de interpretar, sobre todo a la hora de aclarar el carácter de sus fallos. Consideramos que las normas jurídicas existentes deben seguir aplicándose a algunas de estas partes, pero sostenemos que también está justificado que se promulgue nueva legislación. Las decisiones tomadas por todas estas partes influyen en la determinación de a quién se mata, y el alcance de la responsabilidad jurídica sobre el conjunto de ellas influirá en la intensidad de la presión que se ejerza en contra de este sistema.

Las armas de IA concentran el poder

La automatización eficaz del trabajo de inteligencia y de la creación de listas de objetivos humanos supone una concentración de poder y una banalización del proceso de matar aún mayores. En palabras del personal estadounidense que maneja de drones:[92]

> "(...) Ver al hijo de la persona a la que acabo de aniquilar con un misil Hellfire recoger los pedazos de su padre. No me centré en el acto de matar; el hecho de ver la cara del niño y la interacción con el resto de su familia es lo que sigue atormentándome".

En respuesta, desde el interior de las fuerzas armadas se ha pedido establecer una mayor distancia social entre el personal operador y el objetivo. Los listados de objetivos humanos elaborados por la IA pueden crear esta distancia y hacer que las guerras sean aún más mortíferas. No debemos magnificar el valor de la conciencia culpable de los mandos humanos, pero sí nos preocupa la eliminación relativamente repentina y completa de ese sentido moral.

Los listados de objetivos humanos elaborados por la IA también podrían concentrar aún más el poder: resultaría sencillo concebirlos de modo que cambiar los umbrales para determinar quién se considera militante requiera poca supervisión.

Distinción y proporcionalidad

Los listados de objetivos humanos elaborados por la IA son contrarios a la figura de la distinción, que exige que las partes en un conflicto armado distingan entre civiles y objetivos militares. Hay muchos elementos que constituyen limitaciones fundamentales de los modelos de IA:

Normas jurídicas

Aunque no son ni mucho menos las únicas normas jurídicas pertinentes, hay dos principios del derecho internacional humanitario (DIH) que destacan especialmente en el caso de los sistemas de armas de IA:

- La distinción establece que las partes en un conflicto armado tienen la obligación de distinguir entre civiles y objetivos militares y de dirigir ataques únicamente contra objetivos militares.
- La proporcionalidad es la prohibición de los ataques en los conflictos que exponen a la población civil a un riesgo que sería excesivo en relación con la ventaja militar obtenida.
-

Los sistemas como Lavender vulneran estas normas de varias maneras.

92 https://theintercept.com/2021/10/24/drone-war-books-neil-renic-wayne-phelps/

Mecánica de los listados de objetivos humanos elaborados por la IA

En los listados de objetivos humanos elaborados por la IA básicamente se combinan señales de entrada "difusas", ninguna de las cuales identifica de forma positiva ni concluyente a los militantes. En concreto, estas señales de entrada contienen los nombres y patrones de comportamiento de Lavender:[93]

> "Las fuentes explicaron que la máquina de Lavender a veces marcaba erróneamente a quienes se comunicaban de manera similar a los agentes de Hamás o de la Yihad Islámica Palestina conocidos, como policías y trabajadores de la defensa civil, familiares de militantes, habitantes que casualmente tenían un nombre y apodo idénticos a los de algún agente, y gazatíes que utilizaba dispositivos que en su día habían pertenecido a agentes de Hamás".

No se puede racionalizar de modo alguno que la combinación de características justifique matar a alguien.

Supervisión mínima

Quienes aprueban los homicidios con drones tienen órdenes de asumir los listados de objetivos humanos elaborados por Lavender, y sólo comprueban que los objetivos señalados por Lavender sean varones:94

> Una fuente afirmó que el personal humano solía servir solamente para "aprobar automáticamente" las decisiones de la máquina y añadió que, normalmente, dedicaba personalmente sólo unos "20 segundos" a cada objetivo antes de autorizar un bombardeo, simplemente para asegurarse de que el objetivo señalado por Lavender era varón.

Es evidente que el dato de ser hombre no basta para separar los objetivos militares de la población civil. Esto, sumado a los errores de los listados de objetivos humanos elaborados por la IA que analizamos a continuación, quebranta la distinción.

Homicidio semiarbitrario

Hechos como la masacre de My Lai son ampliamente reconocidos como violaciones del principio de distinción. La guerra de alta tecnología debe recibir una denominación similar. Supongamos que un agente de inteligencia identificara 35 objetivos como "sin duda militares" y luego otros 50 de los que no estuviera seguro (quizá hubiera algún indicio, pero no suficientes). Si incluyera los 50, no parecería trabajo de inteligencia, así que lanza un dado por cada uno de los 50 e incluye aquellos en los que sale un seis. ¿Hubo una violación del principio de distinción, incluso si resulta que la mayoría de los objetivos eran militares? Diríamos que sí.

Los listados de objetivos humanos elaborados por la IA funcionan de forma parecida: que no sean *totalmente* aleatorios no significa que no haya algo de aleatoriedad. Cualquier modelo no trivial tiene errores y, si estudiamos el proceso de entrenamiento, esos errores se asemejan más a un "ajuste a un presupuesto de asesinatos" que a una conclusión equivocada a la que llega un mando humano. El hecho de que se clasifique correctamente a algunos militantes y civiles no debe confundirnos para reconocer que etiquetar como militantes a un conjunto arbitrario de objetivos de los que "no se tiene seguridad", cuando no hay pruebas fiables, es un crimen de guerra.

93 https://www.972mag.com/lavender-ai-israeli-army-gaza/ (traducción propia)
94 *Ibid.*

Poner en entredicho las cifras de exactitud declaradas

Cuando los mandos de inteligencia israelíes describen Lavender como "preciso en un 90 %", se refieren a la proporción de clasificaciones correctas sobre un conjunto de datos de evaluación. En el caso de Lavender, lo describen así:[95]

> "(C)omprobábamos 'manualmente' la precisión de una muestra aleatoria de varios centenares de objetivos seleccionados por el sistema de IA".

Esa cifra no procede de una investigación póstuma sobre las personas fallecidas y no debe interpretarse como indicativa de la precisión en el mundo real. Está demostrado que, debido a los cambios en la distribución de los conjuntos de datos, la precisión medida se desploma en las situaciones del mundo real. Este conjunto de datos de evaluación está en manos de las fuerzas armadas, por lo que no puede ser validado por ninguna fuente neutral ni fiable.

Consideremos la cita del personal de inteligencia que mencionamos antes "En la guerra, las personas palestinas cambian de teléfono continuamente": cambiar más de teléfono está probablemente correlacionado con ser militante, lo que significa que los fallos del modelo provocan homicidios de civiles. Los modelos de IA no pueden incorporar datos fundamentales de contextualización ni ninguna información que no esté representada en sus datos de entrada como, por ejemplo, "la zona X es un punto de destino de las personas refugiadas, así que, naturalmente, esperamos que esa población se haya ido a otro lugar". Esta es una limitación fundamental de cualquier listado de asesinatos elaborado por la IA que irremediablemente ocasiona incumplimiento del principio de distinción.

Reducir los umbrales para maximizar la matanza

Ya hemos explicado que los umbrales se reducen para maximizar la muerte y la destrucción. La mecánica de reducir los umbrales cada vez que las autoridades israelíes exigen más objetivos tiene el efecto de clasificar a toda la población de Gaza para el bombardeo. Where's Daddy hace deliberadamente un seguimiento de las personas objetivo hasta que llegan a su casa, lo que a menudo provoca bombardeos que matan a familias enteras. Las autoridades israelíes alegan que resulta más fácil bombardearlas cuando están en casa. Como hemos mencionado antes, los objetivos suelen ser presuntos agentes de bajo rango, lo que significaría que la ventaja militar de matarlos es mínima. Esta violación clara de la proporcionalidad está directamente propiciada por el uso de sistemas de IA.

Encubrimiento complejo

Como ya hemos mencionado, el equipo de ciencia de datos militares etiquetó erróneamente a trabajadores de la defensa civil como agentes de Hamás. La tecnología ha encubierto en cierto modo esta vulneración, puesto que una de las partes ha etiquetado erróneamente a personas —evidentemente para matarlas—, pero no ha tomado literalmente esa decisión. Quienes aprobaron los ataques se basaron en la información errónea de esa otra parte, por medio de la lista de asesinatos elaborada por la IA. En última instancia, la consecuencia son ataques dirigidos contra civiles.

95 *Ibid.*

Conclusión

Los sistemas de IA se utilizan para posibilitar unos niveles espantosos de destrucción en Gaza. Cuando las naciones poderosas atacan a grupos que consideran desechables, la precisión nunca es la prioridad. Los motivos tienen que ver con matar a la gente de forma rápida y barata, y debemos rechazar los argumentos absurdos de que un mayor desarrollo cambiará en lo fundamental esas prioridades.

Las organizaciones de derechos humanos ya han expresado su preocupación por la "deficiencia en materia de rendición de cuentas" con respecto a las armas autónomas letales.[96] El DIH atribuye a personas la responsabilidad jurídica de los crímenes de guerra y las infracciones de los Convenios de Ginebra, pero, cuando las decisiones se basan en la IA, no está claro que las entidades desarrolladoras y proveedoras deban rendir cuentas.

Este problema se intensifica aún más en el caso de la tecnología de consumo. La culpa de las empresas creadoras de aplicaciones de IA explícitamente militarizadas puede resultar más fácil identificar, pero el alcance del uso de tecnologías de consumo a través de los servicios en la nube está encubierto de forma intencionada, como lo demuestran las mentiras de Google. No obstante, es evidente que estas empresas venden con deliberación su tecnología a fuerzas armadas que cometen crímenes de guerra, y que las tecnologías desempeñan un papel clave en facilitar esos crímenes. Los equipos directivos de Google, Amazon y Microsoft están profundamente implicados y, sin embargo, hasta ahora pueden, con creces, obtener ganancias y también evitar la culpabilidad jurídica. Sus decisiones implican al personal tecnológico en este negocio mortífero, lo que genera complicidad sectorial y señala al mismo tiempo un lugar para la resistencia.

Las armas de IA no son en realidad "precisas" ni "objetivas", y su utilización debe considerarse un crimen de guerra. El personal de las empresas tecnológicas debe tener el derecho legal a oponerse a su complicidad en los crímenes de guerra, así como en otras violaciones de derechos humanos tales como la vigilancia masiva y el apartheid. La intensificación de la represión por parte de estas empresas justifica la adopción de más medidas de protección laboral.

96 https://www.hrw.org/news/2020/06/01/need-and-elements-new-treaty-fully-autonomous-weapons

5. Raza y resistencia: desentrañar las economías políticas de la inteligencia artificial

Sarah Chander
Equinox Initiative for Racial Justice
Traducción: Arantxa Albiol Benito

Aunque casi siempre son celebrados como avances progresistas, la implantación de la IA y los procesos de digitalización interactúan con sistemas preexistentes de racismo estructural, sistemas de opresión interconectados, capitalismo, segurización, militarismo y extracción. En este capítulo se examina la manera en que la IA interactúa con el racismo estructural, comenzando por las manifestaciones de discriminación racial en la vigilancia policial, el control migratorio y la asistencia social, y extendiéndose posteriormente a las economías políticas de extracción, explotación, criminalización y guerra digital como características centrales del sector de la IA. Aunque no existe una "solución rápida" que pueda deshacer siglos de racismo y discriminación sistémicos, este capítulo recorre diversas impugnaciones de los sistemas de IA racializados y termina ofreciendo algunas vías de resistencia significativa, como el fortalecimiento del poder entre las comunidades afectadas y los enfoques disruptivos y abolicionistas, pero también la redistribución de los recursos, alejándolos de los sistemas de vigilancia y control y dirigiéndolos hacia la atención comunitaria y social.

Introducción

La inteligencia artificial (IA) y los sistemas automatizados de toma de decisiones se desarrollan, prueban, procesan e implantan en cada vez más esferas de la vida pública. Aunque casi siempre son celebrados como avances progresistas, el despliegue de la IA y los procesos de digitalización interactúan en general con contextos preexistentes de racismo estructural, sistemas de opresión interconectados, capitalismo, segurización, militarismo y extracción. En este capítulo se examinan las interacciones de la IA con el racismo estructural y las distintas formas en que podemos combatir y mostrar resistencia.

Concretamente, la IA se utiliza cada vez más en diversos sectores que ya comportan daños, discriminación y violencia desproporcionados para las personas racializadas. Desde la vigilancia y la toma de decisiones discriminatorias en la actuación policial, pasando por el testeo de nuevas herramientas de gestión de la población con las personas migrantes por parte de las autoridades migratorias, hasta los procesos de evaluación de riesgos que perfilan a las personas racializadas como "sospechosas" y criminales, existen infinitas intersecciones entre la implantación de la tecnología y el racismo estructural.

El capítulo distingue de forma generalizada entre las dos formas en que racismo e inteligencia artificial y tecnología se vinculan. La primera, analizada en el apartado I, se refiere a las repercusiones desproporcionadamente perjudiciales del uso de los sistemas de IA para las personas y comunidades racializadas. La segunda, expuesta en el apartado II, describe una economía política de la IA racializada en sentido más amplio, más allá del impacto individual o comunitario, analizando las formas en que los sistemas de IA se integran en sistemas más amplios de opresión racializada, gestión de la población, extracción y producción.

El uso de sistemas basados en los datos para vigilar y dar una lógica a la discriminación no es novedoso. El uso de sistemas de recopilación de datos biométricos, como las huellas dactilares, tiene su origen en los sistemas coloniales de control. El uso de marcadores biométricos para experimentar, discriminar y exterminar también fue una característica del régimen nazi. No obstante, en este capítulo no se profundiza en estas trayectorias raciales, coloniales y militaristas del desarrollo de la tecnología y la IA. Si observamos los precedentes de la IA, desde los orígenes coloniales de la toma de huellas dactilares y las prácticas de vigilancia racializada, como las denominadas *lantern laws*,[1] hasta las "innovaciones" más recientes de la inteligencia artificial moderna en el ejército estadounidense, vemos que la tecnología siempre ha desempeñado un papel en las formaciones raciales, la vigilancia policial y la gestión de la población.

Así pues, la IA es, por su propia naturaleza, un concepto y una construcción racializados, y muchas veces las iniciativas para "solucionar" los daños raciales serán insuficientes si no se integran en estrategias más profundas, estructurales y políticas vinculadas a la justicia racial y social, la despenalización, la redistribución y, en última instancia, la descolonización. Analizaremos esas respuestas, tanto a corto como a largo plazo, reformistas y abolicionistas, institucionales, legislativas y centradas en la resistencia, en el apartado III.

En este capítulo se toma a Europa como punto de referencia, ya que es donde la autora reside y desarrolla su actividad. Sin embargo, el texto establecerá una conexión inherente entre los procesos de extracción mundiales, los avances tecnológicos y el impacto desde el Sur Global hacia el norte.

1 Hace referencia a una legislación impuesta en la ciudad de Nueva York en el S.XVIII, que obligaba a las personas negras, mestizas o indígenas que caminasen de noche y sin ir acompañadas por ninguna persona blanca a llegar un farol para hacerse visibles. Browne, S. (2015) *Dark Matters: On the Surveillance of Blackness*

IA y racismo estructural: en contexto

En Europa y en todo el mundo se utilizan sistemas de IA para supervisarnos y controlarnos en los espacios públicos, predecir nuestra probabilidad de delinquir en el futuro, "prevenir" la migración, predecir nuestras emociones y tomar decisiones trascendentales que determinan nuestro acceso a los servicios públicos, como las prestaciones sociales. El desarrollo y el uso de la IA en estas esferas concretas puede considerarse el elemento central de cualquier análisis sobre la IA y el racismo estructural, ya que es en estas áreas donde éste más se manifiesta, incluso sin un componente digital. En este apartado se demuestra cómo la IA no hará sino alimentar esta realidad de racismo estructural con más herramientas, más poderes legales y menos responsabilidad y transparencia para los agentes públicos con influencia sobre la seguridad y el bienestar de las personas racializadas.

Agencias de vigilancia policial y seguridad

El creciente uso de la IA en los contextos de vigilancia policial y migración tiene enormes implicaciones para la discriminación racial. En el campo de la actuación policial, los sistemas de IA permiten técnicas nuevas y más invasivas de vigilancia y control, "tecnificando" y sistematizando así la discriminación.[2] Desde el uso del reconocimiento facial para identificar a las personas mientras se mueven libremente por lugares públicos hasta los sistemas de vigilancia policial predictiva para decidir quién es un delincuente antes de que cometa delitos, la IA brinda a los Gobiernos la posibilidad de articular la vigilancia vulnerando las libertades de nuevas y perniciosas maneras. La implantación de este tipo de tecnologías expone a las personas de color a una mayor vigilancia, a una toma de decisiones más discriminatoria y a una elaboración de perfiles más perjudicial.

La vigilancia masiva basada en la IA[3] forma parte de un espectro de técnicas que la policía, las autoridades locales y las empresas emplean para identificar a las personas en lugares públicos. Agrupados como sistemas de "identificación biométrica remota", se trata de técnicas mediante las cuales las fuerzas de seguridad utilizan la IA para identificar a las personas, ya sea capturando imágenes faciales ("reconocimiento facial") o utilizando otros métodos, como la forma de andar y la voz, para determinar su identidad. Muchas veces estas técnicas se combinan con sistemas de "categorización biométrica", diseñados para deducir características o comportamientos a partir de la categorización de rasgos biométricos. Por ejemplo, se han implantado sistemas de categorización biométrica para deducir el género, la raza y otras características sensibles, o para caracterizar a las personas como "sospechosas", recurriendo a menudo a variables altamente racializadas.

Las tecnologías de vigilancia masiva, ya sea con fines de identificación, categorización, reconocimiento de emociones u otros, perpetuarán el racismo estructural. En particular, se ha demostrado que el reconocimiento facial y otros sistemas biométricos de vigilancia masiva han facilitado la vigilancia policial excesiva de las comunidades racializadas y las localidades donde residen.[4] Todo ello, unido a ideas preconcebidas altamente racializados y clasistas y datos sesgados sobre dónde y quién comete los delitos, ha derivado en una implantación desproporcionada de tecnologías biométricas en zonas donde viven personas racializadas. Por ejemplo, un estudio encargado por European Digital Rights puso de manifiesto numerosos ejemplos de instalación excesiva de cámaras de vigilancia "biométricas" en zonas cercanas a lugares de trabajo y locales LGBT+. Además, estas técnicas suelen emplearse específicamente para vigilar a las personas migrantes. Según la investigación

2 Williams, Patrick y Kind, Eric (2019). *Data-driven policing: the hardwiring of discriminatory policing practices in Europe* https://www.statewatch.org/media/documents/news/2019/nov/data-driven-profiling-web-final.pdf

3 EDRi (2021). *The rise and rise of biometric mass surveillance in the EU*: https://edri.org/wp-content/uploads/2021/11/EDRI_RISE_REPORT.pdf

4 Agencia de los Derechos Fundamentales de la Unión Europea (2019). *Facial recognition technology fundamental rights considerations in the context of law enforcement*.

del Hermes Center,[5] el Ministerio del Interior italiano adquirió en 2017 el "SARI", un sistema de reconocimiento facial que se utilizará en distintos contextos, incluidas las manifestaciones. Como se descubrió en 2019, el Hermes Center detectó un uso desproporcionado del sistema con personas migrantes, alimentando un clima altamente racializado de sospecha, vigilancia excesiva y criminalización de estas personas.[6]

Otro ejemplo de la infraestructura de vigilancia policial racializada por la IA es el creciente recurso a sistemas de vigilancia policial predictiva en toda Europa. La expresión "vigilancia policial predictiva" designa los sistemas que elaboran perfiles de personas y zonas, predicen supuestas conductas delictivas o la comisión de delitos, y evalúan el supuesto "riesgo" de delincuencia o criminalidad en el futuro.[7] Desde sistemas como el "Top 400" y "ProKid" en Holanda, que llevan a cabo evaluaciones de riesgo de las personas para puntuar la probabilidad de que cometan delitos en el futuro, hasta sistemas como el "Delia" en Italia, que evalúa específicamente el riesgo de las zonas o lugares,[8] los sistemas de vigilancia policial predictiva son intentos de predecir patrones delictivos para fundamentar la asignación de recursos policiales. Los resultados de este tipo de sistemas son un mayor número de interacciones con las fuerzas del orden, que pueden dar lugar a un aumento de la elaboración de perfiles raciales y de las identificaciones y registros, controles migratorios, que pueden conducir a la deportación, detenciones, procesamientos y condenas discriminatorias, e incluso a casos potenciales de brutalidad policial racista.

Los sistemas de vigilancia policial predictiva afectan desproporcionadamente a las personas racializadas, a las personas percibidas como potenciales migrantes, terroristas, personas pobres y de clase trabajadora, en zonas pobres y populares. Como estos sistemas se basan en datos y prácticas policiales existentes que ya muestran patrones discriminatorios contra las personas racializadas, migrantes y pobres, los resultados de implantar sistemas de vigilancia policial predictiva son, por tanto, intrínsecamente discriminatorios. Esto puede manifestarse de muchas maneras. En algunos casos, estos sistemas utilizan explícitamente datos étnicos o sustitutos de los mismos como variables en la predicción algorítmica de la delincuencia. La organización Fair Trials, nos señala varios ejemplos de esto, entre ellos, el ya mencionado sistema Delia en Italia, que utiliza datos étnicos en sus predicciones, y el "RisCANVI", que implantó el Departamento de Justicia de Cataluña y utiliza datos relativos a la nacionalidad.[9]

En muchos casos, los sistemas de vigilancia policial predictiva (independientemente de las categorías específicas de los datos utilizados) dan lugar a una sobrerrepresentación sustancial de personas racializadas, presentadas como de "alto riesgo" de delincuencia futura. Los sistemas holandeses "Top-400" y "Top-600" generaron bases de datos de personas con probabilidades de cometer delitos en el futuro o reincidir, lo que dio lugar a una intensa sobrevigilancia por parte de la policía y los agentes de la seguridad social, como visitas domiciliarias sin previo aviso, control y seguimiento, y otras sanciones informales como informar a los empleadores y otros círculos sociales. Estas listas incluían un número desproporcionado de hombres y jóvenes holandeses-marroquíes, así como de otras minorías raciales, personas sin hogar y personas de familias con bajos ingresos.[10] Sobre todo, es importante recordar que los sistemas de vigilancia policial predictiva están diseñados para "predecir y prevenir" la delincuencia y, por ello, crean una forma de estado de "pre-delincuencia" por el que las personas sienten el impacto de la vigilancia policial, aunque todavía no hayan cometido el delito en cuestión. Más allá de la preocupación que suscitan estos

5 Coluccini, Riccardo (2017). Italian police has acquired a facial recognition system. https://medium.com/@ORARiccardo/italian-police-has-acquired-a-facial-recognition-system-a54016211ff2

6 https://www.wired.it/attualita/tech/2019/04/03/sari-riconoscimento-facciale-stranieri/

7 EDRi (2022). *Prohibit predictive policing and profiling AI systems in law enforcement*: https://edri.org/wp-content/uploads/2022/05/Prohibit-predictive-and-profiling-AI-systems-in-law-enforcement-and-criminal-justice.pdf

8 Fair Trials (2021). *Automating Injustice: the use of artificial intelligence and automated decision making systems in criminal justice in Europe*. https://www.fairtrials.org/app/uploads/2021/11/Automating_Injustice.pdf

9 *Ibíd.*

10 https://controlealtdelete.nl/articles/top400-aanpak-is-discriminerend#gsc.tab=0

sistemas de vigilancia y sospecha racializados y la criminalización de las personas pobres, estos también contribuyen a la erosión generalizada del principio de presunción de inocencia.

Migración

Intrínsecamente ligado al papel que la IA desempeña en la vigilancia policial racializada está el uso de la IA en la "gestión" de la migración y el control de fronteras. Cada vez se desarrollan más sistemas de IA para rastrear, controlar y supervisar a las personas migrantes de formas nuevas y dañinas. Desde los detectores de mentiras y los "perfiles de riesgo" basados en IA, utilizados en un gran número de trámites migratorios, hasta la vigilancia tecnológica en rápida expansión de las fronteras europeas, los sistemas de IA tienen un peso cada vez mayor en el enfoque de la UE respecto a la migración.

Una de las manifestaciones más claras del carácter racializado del uso de la IA en la migración es el recurso a los sistemas predictivos. Desde la evaluación individualizada del riesgo hasta sistemas analíticos predictivos más amplios utilizados para prever patrones migratorios e implantar medidas migratorias en esas zonas. Los sistemas individualizados de evaluación de riesgos, como el Sistema Europeo de Información y Autorización de Viajes (ETIAS), que permite la elaboración de perfiles (y potencialmente el uso de IA) para clasificar a las personas que viajan en perfiles de riesgo predefinidos relacionados con supuestos riesgos migratorios, de seguridad o de salud pública. Esta elaboración de perfiles se lleva a cabo teniendo en cuenta una serie de factores, como los datos históricos sobre las tasas previas de denegación o estancia excesiva y la información facilitada por los Estados miembros sobre los riesgos para la seguridad.[11] Los sistemas de evaluación de riesgos elaboran perfiles de las personas basándose en indicadores de riesgo predeterminados integrados en los sistemas de control, cuyos parámetros suelen ser opacos y no se hacen públicos. Estos sistemas son discriminatorios por naturaleza: codifican los supuestos sobre el vínculo entre los datos personales y las características con riesgos particulares. No se juzga a las personas por su comportamiento individual o por factores que estén bajo su control, sino por ciertas características predeterminadas, como la nacionalidad.[12] Así pues, el creciente uso de las evaluaciones de riesgos automatizadas o basadas en la IA en el control de la migración es una forma de objetivar patrones de sospecha racializada como forma de prevenir y gestionar la migración, sistematizando un proceso de generalización sobre las personas antes de que se hayan desplazado. Aunque no se detalla aquí, el creciente uso de la IA en la gestión de casos individuales para cuestionar la veracidad de las solicitudes de migración también forma parte de esta tendencia, como el uso del reconocimiento de emociones (uso de sistemas de IA para deducir o evaluar emociones, basándose en datos biométricos), los polígrafos con IA y los sistemas de reconocimiento de dialectos basados en IA.

Más allá de la toma de decisiones individuales, estamos viendo la inversión en IA como parte de un dispositivo de vigilancia generalizado en constante expansión. Esto incluye IA para la vigilancia en las fronteras y sistemas analíticos predictivos para prever las tendencias migratorias. Según la red Border Violence Monitoring Network, Gobiernos e instituciones como Frontex pueden estar utilizando tecnologías basadas en la inteligencia artificial para facilitar la expulsión de migrantes, lo que a menudo equivale a desapariciones forzosas. La Red Euromediterránea de Derechos Humanos (Euromed Rights) ha documentado el creciente uso de la inteligencia artificial en las fronteras exteriores de la UE, financiada como parte de las "alianzas" de externalización coloniales (acuerdos entre la UE y Estados no pertenecientes a la UE que exigen técnicas de gestión de la migración, a menudo a cambio de ayuda) para cooptar a Estados no pertenecientes a la UE en el régimen migra-

11 Véase Niovi Vavoula (2020) *The Commission Package for ETIAS Consequential Amendments – Substitute Impact Assessment,* págs. 20–30.

12 Niovi Vavolua (2022) *Immigration and Privacy in the Law of the European Union – The Case of Information Systems.*

torio preventivo de la UE.[13] En este contexto, existe un peligro real de que las previsiones aparentemente inocuas sobre patrones de migración se utilicen para facilitar retrocesos, repliegues y otras formas de impedir que las personas ejerzan su derecho a solicitar asilo. Esta infraestructura también incluye el recurso a sistemas de IA para la aplicación de la legislación migratoria dentro de las fronteras: por ejemplo, en Francia, Alemania, Países Bajos y Suecia, se ha facultado a la policía para tomar las huellas dactilares de las personas a las que paran en la calle para comprobar su situación en materia de migración. Esta infraestructura y el mayor recurso a la tecnología de vigilancia han sido fomentados y facilitados por el recientemente adoptado Pacto sobre Migración y Asilo de la UE [14], que marca el comienzo de una nueva era de vigilancia digital mortífera, ampliando la infraestructura digital para un régimen fronterizo de la UE basado en la criminalización y el castigo de las personas migrantes y las personas racializadas.[15]

Prestaciones y servicios sociales

A menudo presentadas como medidas de reducción de costes y aumento de la eficiencia con un impacto neutro en los derechos de las personas, el creciente recurso a la IA y a los procedimientos de digitalización en general en los servicios sociales está teniendo un impacto demostrable en las personas con rentas bajas, de clase trabajadora y racializadas. Muchas veces vemos que los resultados de estos sistemas tienen efectos discriminatorios y agravan la vigilancia o los resultados negativos que experimentan los grupos marginados. Sin embargo, como se refleja en los casos relativos a la vigilancia policial predictiva, también vemos que la propia decisión de implantar estos sistemas en determinadas zonas o con respecto a determinados grupos de la sociedad, refleja las tendencias racistas, xenófobas y contrarias a las personas pobres de los Gobiernos, que dan prioridad a los denominados daños sociales, como el fraude en las prestaciones, frente a otras cuestiones sociales. Se trata de decisiones intrínsecamente políticas, clasistas y racializadas, cualesquiera que sean los efectos o el funcionamiento de los sistemas concretos.

En Europa, la IA y los sistemas algorítmicos se han utilizado para desempeñar tareas de comprobación de la identidad, asignar y determinar el acceso y la elegibilidad a los servicios de seguridad y asistencia social (como las prestaciones por desempleo), y también para predecir y evaluar el riesgo de fraude en las prestaciones y la asistencia social.

Por ejemplo, en 2014, el Gobierno holandés autorizó y puso en marcha el *Systeem Risico Indicatie*, o SyRI, en barrios predominantemente de bajos ingresos, una iniciativa para predecir y calificar la probabilidad de que las personas incurran en fraude fiscal o relativo a las prestaciones. Para ello, el sistema SyRI enlazó y analizó una gran cantidad de datos sobre los ciudadanos y ciudadanas holandeses, incluidos datos sobre la situación laboral, la propiedad de bienes inmuebles, la educación, la jubilación, la actividad empresarial, los ingresos y el patrimonio, la pensión y las deudas. Se basaba en una colaboración entre los municipios, el Ministerio de Asuntos Sociales y Empleo, la Policía, la fiscalía, los servicios migratorios y las autoridades fiscales y de asistencia social. [16] A continuación, el sistema generó una lista de direcciones de personas que mostraban una mayor probabilidad de fraude en las prestaciones públicas. Aunque el tribunal anuló el sistema por motivos de privacidad y transparencia, el daño infligido a las personas pobres y de clase trabajadora, en su mayoría procedentes de comunidades racializadas y migrantes, fue enorme.[17]
En Austria, como destacó la organización epicenter.works, el Gobierno puso en marcha el

13 Euromed rights (2023). *Artificial Intelligence: the new frontier of the EU's border externalisation strategy*: https://euromedrights.org/wp-content/uploads/2023/07/Euromed_AI-Migration-Report_EN-1.pdf
14 https://home-affairs.ec.europa.eu/policies/migration-and-asylum/pact-migration-and-asylum_en#what-is-the-pact-on-migration-and-asylum
15 Protect Not Surveil Coalition (2024). T*he EU Migration Pact: A deadly regime of migrant surveillance*: https://www.equinox-eu.com/wp-content/uploads/2024/04/The-Migration-Pact-ProtectNotSurveil.pdf
16 Mustafa, Nawal (2023). Article 47: The age of digital inequalities. Digital Rights are Charter Rights del Digital Freedom Fund: https://digitalfreedomfund.org/wp-content/uploads/2023/09/Digirise_V11_Digital.pdf
17 Bits of Freedom (2021) "We want more than symbolic gestures in response to discriminatory algorithms".

"algoritmo AMS", un sistema algorítmico diseñado para predecir las perspectivas de empleo de una persona que busca trabajo basándose en factores como el género, el grupo de edad, la ciudadanía, la salud, la profesión y la experiencia laboral. Al priorizar los servicios basándose en esta información, el algoritmo AMS dio lugar a una reducción de las ayudas a los solicitantes de empleo con escasas y grandes perspectivas de empleo, y también discriminó a las mujeres mayores de 30 años, a las mujeres al cuidado de menores y a las personas migrantes.[18]

En otros casos, hemos visto a los gobiernos emplear sistemas de IA con el propósito aparentemente banal de identificar a las personas como condición previa para acceder a su seguridad social. Aunque constituyen menos un ejercicio de "toma de decisiones" que los ejemplos anteriores, estos casos de uso han tenido graves repercusiones discriminatorias, en la medida en que los fallos e imprecisiones de estos sistemas han dado lugar a falsas determinaciones de fraude de identidad, dejando a personas sin prestaciones. Estos sistemas son una forma tecnológica de aumentar la presión administrativa sobre las personas que necesitan acceder a las prestaciones, lo que perjudica desproporcionadamente a las personas racializadas y a las personas en situación precaria. En su comentario sobre las peligrosas repercusiones de la introducción de la IA en la asistencia social, Human Rights Watch afirmó que esta tendencia hacia la automatización puede "discriminar a las personas que necesitan apoyo de la seguridad social, comprometer su privacidad y dificultar que puedan optar a la ayuda del Gobierno".[19]

Más allá de la discriminación: las economías políticas de la IA racializada

Aunque los desproporcionados efectos negativos experimentados por las personas racializadas son un aspecto de la relación entre la IA y el racismo estructural, es preciso evaluar un marco mucho más amplio. Concretamente, es tanto o más importante evaluar las repercusiones transformadoras del sector de la IA, desde las repercusiones económicas del desarrollo y la producción de IA hasta los grupos de presión y la relación con sectores más amplios, como la seguridad, la defensa y la militarización. Estas economías políticas repercuten en las comunidades racializadas a todos los niveles e interactúan con los sistemas de explotación racista y la sobreexposición sistemática a la muerte prematura a escala mundial. El análisis de economía política de la IA también desvela la conexión entre el sector de la IA y el capitalismo racial, así como el papel de la IA en la "producción y explotación de la vulnerabilidad a la muerte prematura diferenciada por grupos".[20] En este apartado se ofrece una breve descripción introductoria de estas distintas economías políticas.

Extracción, producción y política industrial de la IA

La IA es un negocio multimillonario. Aunque a menudo se elogia en términos políticamente neutros o positivos como un beneficio inequívoco para la sociedad o un símbolo de progreso económico, es imprescindible recordar que los sistemas de IA son una parte fundamental de las iniciativas de las grandes empresas tecnológicas para ampliar su poder de mercado.[21]

18 Epicenter.works (2020). Warum der polnische "AMS"-Algorithmus gescheitert ist. *Epicenter.works*. https://epicenter.works/content/warum-der-polnische-ams-algorithmus-gescheitert-ist

19 Human Rights Watch (2021). *How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net*. https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net

20 Gilmore, Ruth Wilson (2007). Golden gulag: Prisons, surplus, crisis, and opposition in globalizing california, Berkeley, CA: University of California Press.

21 EDRi (2021). *How Big Tech maintains its dominance*: https://edri.org/our-work/how-big-tech-maintains-its-dominance/

La Inteligencia Artificial se ha convertido en un pilar central de la política industrial de los Gobiernos e instituciones de todo el mundo. Como parte de una labor más amplia para generar valor a través de la digitalización de las economías, los Gobiernos han invertido de forma exponencial en discursos tecnosolucionistas que esgrimen los beneficios de la IA y la necesidad de "promover su adopción" para seguir siendo competitivos a escala mundial. Sin embargo, promover la IA en el sector público en su conjunto, sin exigir pruebas científicas que justifiquen la necesidad o el propósito de tales aplicaciones en algunas situaciones potencialmente perjudiciales, podría repercutir más directamente en la vida cotidiana de las personas, sobre todo de los grupos marginados.

En términos más generales, las políticas industriales "pro IA" ocultan una serie de perjuicios sociales más amplios en el centro del sector de la IA y de su concepto en general. En primer lugar, ocultan que el sector de la IA se basa de forma inherente en la extracción y la explotación: de la mano de obra, de los recursos naturales, de las tierras. Cada vez hay más estudios que revelan el grado de explotación laboral subyacente en el sector de la IA, principalmente de trabajadores del Sur Global. Sistemas como Chat GPT –manifestación de cara al consumidor de grandes modelos de lenguaje o sistemas de inteligencia artificial de propósito general– explotan sistemáticamente a las y los trabajadores de diversas maneras, por ejemplo, como parte del proceso de producción y mantenimiento mediante el etiquetado, filtrado y recopilación de datos,[22] la moderación de contenidos para plataformas en línea, o mediante la prestación directa de servicios como trabajadores de plataformas de transporte o mensajería. Todas estas formas de empleo se basan en una mano de obra precaria y muy explotada, generalmente con largas jornadas laborales, bajos salarios y malas condiciones. Prug y Bilic describen este proceso de "dividir y ocultar el trabajo" para garantizar que los productos parezcan mágicamente automatizados, todo ello mientras se amasa el valor del trabajo de las y los trabajadores del Sur Global, o en situaciones precarias en el norte mundial.[23]

Además, el sector de la IA, en su proceso de producción, también se basa en una gran explotación medioambiental y extracción de recursos naturales. Los sistemas de IA y los marcos que los sustentan requieren grandes capacidades de cálculo y, por tanto, de fuentes de energía. Además, la recopilación de datos para la IA se basa en el uso cada vez mayor de dispositivos móviles y redes de sensores, todo lo cual requiere dispositivos que dependen de la extracción de minerales y de materiales, origen de, por ejemplo, la guerra en el Congo. Como ha destacado la organización Generation Lumiere, el sector electrónico mundial ha alimentado la guerra en la República Democrática del Congo, debido a la demanda de cobalto, cobre, coltán y litio, de los que el país tiene grandes reservas.[24] El sector de la inteligencia artificial intenta ocultar las cuestiones relacionadas con el medio ambiente, el clima y los conflictos mediante el marketing, afirmando que la inteligencia artificial nos ayudará a luchar contra el cambio climático (por ejemplo, mediante la modelización de previsiones), una narrativa que, por desgracia, reproducen varias instituciones responsables de la elaboración de políticas.

Segurización digital, militarización y guerra digital

Como se ha destacado en el apartado anterior, cada vez son más las fuerzas del orden y los organismos migratorios que recurren a la IA. En lugar de considerarlos como acontecimientos aislados, debemos saber que estos usos forman parte de tendencias más amplias de segurización y militarización. Aquí la "segurización" se refiere a un marco constituido por recursos, leyes, narrativas y, cada vez más, infraestructuras tecnológicas movilizados

22 Prug, Toni, y Bilić, Paško (2021). Work Now, Profit Later: AI Between Capital, Labour and Regulation. En P. V. Moore y J. Woodcock (Eds.), *Augmented Exploitation: Artificial Intelligence, Automation and Work* (pp. 30–40). Pluto Press.
23 *Ibíd.*
24 Generation Lumière (2024). La consommation en métaux des Européens génère des massacres – *Reporterre*: https://reporterre.net/L-appetit-en-metaux-des-Europeens-genere-des-massacres#:~:text=%C2%AB%20La%20consommation%20en%20m%C3%A9taux%20des%20Europ%C3%A9ens%20g%C3%A9n%C3%A8re%20des%20massacres%20%C2%BB,-Des%20mineurs%20en&text=Le%20conflit%20au%20Kivu%2C%20en,et%20de%20l'Union%20europ%C3%A9enne.

en pos de una visión de la "seguridad" que gira en torno a soluciones a los problemas sociales basadas en la militarización, la punición y la vigilancia. En el marco de la securización de la UE, las instituciones combinan irrevocablemente el concepto de seguridad pública con la policía, las fronteras y el ejército.

La securización digital —la integración de la digitalización en las tendencias de la securización[25]— implica proyectos que amplían la base jurídica de las infraestructuras tecnológicas para aumentar la vigilancia y la criminalización (como el mencionado Pacto sobre Migración y Asilo de la UE, o la Ley de Inteligencia Artificial, que se analiza en el siguiente apartado), pero también inversiones masivas en los productos digitales de las industrias de seguridad y militar. Por ejemplo, la UE dedica cada vez más financiación a ampliar las infraestructuras y agencias de seguridad, como EUROPOL, la agencia de cooperación policial de la UE. Gran parte de esta inversión se externaliza a través de contratos con empresas privadas de vigilancia y tecnología para desarrollar herramientas con el fin de aumentar las deportaciones y la vigilancia fronteriza. Según el informe de Statewatch, el plan de adquisiciones de Frontex para 2023 incluía 260 millones de euros para sistemas informáticos, incluido el desarrollo de software, infraestructuras y sistemas administrativos, y otros 180 millones de euros en material para la vigilancia de fronteras, incluido un contrato de 144 millones de euros para drones.[26] Así, asistimos a la intrusión del sector privado, incluidas las empresas tecnológicas, en las funciones estatales, integrando cada vez más estos intereses económicos en instituciones estatales como la aplicación de la ley y el control de la migración. De igual forma, vemos una reorientación de los presupuestos hacia la actuación policial, la vigilancia y el control de poblaciones mayoritariamente racializadas, cuando los recursos podrían destinarse a otros fines. Por ejemplo, la cantidad total de dinero destinada a gastos de seguridad y defensa en el presupuesto de la UE de 2021-2027 asciende a 43 900 millones de euros, lo que supone un aumento de más del 123 % en comparación con el anterior ciclo presupuestario de siete años, que asignó 19 700 millones de euros al mismo fin. En comparación, el Programa Ciudadanos, Igualdad, Derechos y Valores sólo destinó 1400 millones de euros a proyectos dedicados a mejorar los derechos y la igualdad, especialmente de los grupos marginados.[27] Esto no sólo produce una exposición desproporcionada a los daños que hemos descrito en el apartado anterior, sino que asistimos a una transformación de los objetivos de los Estados y las instituciones, que se alejan de la atención y la protección sociales para centrarse, en cambio, en la vigilancia, el control, la actuación policial y la gestión de la población.

Estas tendencias se reflejan directamente en el contexto de la militarización, en el que los sistemas de IA se desarrollan, prueban, adquieren e implantan cada vez más para funciones relacionadas con la guerra. Los Estados e instituciones invierten cada vez más en IA en el contexto militar utilizando el pretexto de la seguridad: automatizar el armamento para reducir la necesidad de personal físico sobre el terreno. Este discurso ignora en gran medida las consecuencias fatales para las poblaciones que se encuentran en el filo de estas tecnologías. Un ejemplo destacado de cómo se ha utilizado el sector de la IA para facilitar la guerra y potenciar los asesinatos en masa puede verse en varios usos relacionados con el genocidio de Israel en Gaza. Algunos sistemas como Lavender[28] se han utilizado para multiplicar la generación de objetivos en los bombardeos de Israel. Por lo tanto, estas tecnologías deben equipararse a los equipos de guerra y el armamento y deben tratarse como tales.

25 Chander, Sarah y Gürses, Seda (2024). From Infrastructural Power to Redistribution: How the EU's Digital Agenda Cements Securitization and Computational Infrastructures (and How We Build Otherwise). Aparecido en *Europe's AI Industrial Policy*. AI Now: https://ainowinstitute.org/publication/from-infrastructural-power-to-redistribution-how-the-eus-digital-agenda-cements-securitization-and-computational-infrastructures-and-how-we-build-otherwise

26 Statewatch (2023). Frontex to spend millions of euros on surveillance and deportations: https://www.statewatch.org/news/2023/april/frontex-to-spend-hundreds-of-millions-of-euros-on-surveillance-and-deportations/#:~:text=Frontex%20will%20spend%20hundreds%20of,management%20board%20in%20mid%2DFebruary.

27 Statewatch (2022). At What Cost? Funding the EU's security, defence, and border policies, 2021–2027: https://www.statewatch.org/publications/reports-and-books/at-what-cost-funding-the-eu-s-security-defence-and-border-policies-2021-2027/

28 Abraham, Yuval (2024). Lavender: The AI machine directing Israel's bombing spree in Gaza. *+972 magazine* https://www.972mag.com/lavender-ai-israeli-army-gaza/

Sin embargo, el análisis de la financiación de estos sistemas revela una vez más que no se trata de usos aislados de tecnología dañina, sino de una industria alimentada por una serie de intereses creados y de los vínculos más amplios entre el sector de la IA y la militarización. Lushenko y Carter examinan el crecimiento de los usos militares de la IA: "El camino de rosas de la guerra basada en la IA está allanado por un nuevo complejo militar-industrial. Los países suelen adquirir tecnologías militares, como los drones, por motivos relacionados con la oferta, la demanda y consideraciones de estatus".[29]

Así pues, los intereses económicos subyacentes a la inversión en IA para la guerra no son solamente la seguridad y la defensa, sino también consideraciones de estatus, la financiarización y los intereses del sector tecnológico en sentido más amplio. Esto puede verse como una clara continuación del origen y la trayectoria del desarrollo y la inversión en IA como algo intrínsecamente relacionado con el ejército estadounidense desde 1958 y la creación de DARPA, la Agencia de Proyectos de Investigación Avanzada para facilitar la investigación y el desarrollo de estrategias militares e industriales.

Con frecuencia, la IA se desarrolla para otros fines más amplios y luego se implanta en contextos bélicos, y viceversa. Muchas veces, las fuerzas del orden intercambian conocimientos y tecnologías de contextos militares que luego se modifican para la actuación policial nacional, lo que pone de relieve los vínculos mundiales en la vigilancia policial y la gestión de la población de los pueblos racializados. Como afirma Sara Hamid: "Pero no se trata sólo de los mercados mundiales. También se trata de contextos mundiales. La vigilancia policial estadounidense funciona como un centro de investigación para la innovación militar: el proceso "de verde a azul" es bidireccional".[30]

Esta conexión queda patente cuando se observan las infraestructuras fronterizas como parte de la Fortaleza Europa, donde agencias como Frontex implantan cada vez más tecnologías de vigilancia probadas por Israel contra el pueblo palestino en el contexto del apartheid, la ocupación y el genocidio. Por tanto, el papel de la IA en la vigilancia policial y la gestión de poblaciones racializadas tiene alcance mundial y está íntimamente conectado.

Si seguimos las tendencias descritas anteriormente, observamos una inclinación abrumadora hacia el aumento de la inversión en el sector de la IA, tanto en los presupuestos públicos como en la inversión privada. Independientemente de los ámbitos en los que se implanten las infraestructuras de IA, su creciente adopción e inversión se ha convertido en parte integrante de la prestación de servicios públicos y, como tal, ha transformado el funcionamiento de las instituciones y funciones democráticas de forma que se orientan hacia los fines lucrativos de la industria. Esto tiene implicaciones fundamentales: transformar las instituciones necesarias para la democracia y, al mismo tiempo, cimentar el poder infraestructural de las empresas tecnológicas estadounidenses.[31]

29 Lushenko, Paul y Carter, Keith (2024). A new military industrial complex: how tech bros are hyping AI's role in war. *Bulletin of the Atomic Scientists.* https://thebulletin.org/2024/10/a-new-military-industrial-complex-how-tech-bros-are-hyping-ais-role-in-war/

30 Logic Magazine (2020). Community Defence: Sarah T Hamid on Abolishing Carceral Technologies. Logic(s) Issue 11: https://logicmag.io/care/community-defense-sarah-t-hamid-on-abolishing-carceral-technologies/

31 https://edri.org/our-work/if-ai-is-the-problem-is-debiasing-the-solution/#:~:text=AI%2Ddriven%20systems%20have%20broad,of%20debiasing%20algorithms%20and%20datasets.

Respuesta y resistencia: trazar iniciativas para contextualizar la IA y el racismo estructural

No existe una "solución rápida" para deshacer siglos de racismo y discriminación sistémicos. Como hemos visto a lo largo de este capítulo, el problema no está sólo en la tecnología, sino en los sistemas en los que vivimos. En la mayoría de los casos, los sistemas de IA sólo hacen que las injusticias raciales, la discriminación y la violencia sean más difíciles de localizar y rebatir. Y, no obstante, es necesaria la contestación a varios niveles. Sin embargo, las respuestas al problema del racismo y la IA son muy políticas, motivadas por juicios de valor muy específicos sobre la tecnología y las propias instituciones que la implantan. En este capítulo, trazamos y diferenciamos varias metodologías para responder a los daños de la IA, pasando de las medidas corporativas de desprejuiciamiento (o *debiasing*, en inglés) y auditoría a los intentos reguladores y prácticas de resistencia mucho más amplias.

El desprejuiciamiento técnico

Una de las cantilenas más comunes en respuesta al impacto discriminatorio de la IA procede del ámbito corporativo o técnico, y consiste en plantear la discriminación como un "sesgo" dentro de la tecnología y adoptar técnicas para reducirlo dentro del sistema técnico. Las respuestas de desprejuiciamiento, o reducción de sesgos, se basan en una lógica que explica los daños causados por la IA como resultado de datos sesgados o problemas técnicos, en contraposición a problemas estructurales, políticos y económicos más amplios de la sociedad o del sector de la IA. En la mayoría de los casos, estas soluciones tecnocéntricas pueden, como mucho, reducir los problemas superficiales del funcionamiento de un sistema de IA. Por ejemplo, el desprejuiciamiento técnico puede desentrañar las razones por las que los sistemas funcionan bien en algunas poblaciones y no en otras, como en el caso de los sistemas de reconocimiento facial que identifican erróneamente de forma desproporcionada a personas racializadas. Estos problemas suelen deberse a que los juegos de datos o el funcionamiento de los algoritmos son incompletos o poco representativos. Aunque se explican por cuestiones sociales, las técnicas de desprejuiciamiento se basan en el mito de que un sistema de IA puede ser imparcial cuando se implanta en un mundo caracterizado por el racismo estructural. Por estas razones, Julia Powles sostiene que el desprejuiciamiento es una "distracción seductora" y que estas técnicas equivalen a "perfeccionar los instrumentos de vigilancia".[32]

Seda Gürses y Agathe Balayn han documentado muy bien preocupaciones mucho más amplias sobre las técnicas de desprejuiciamiento. En muchas formas, argumentan, el desprejuiciamiento es un sector empresarial en sí mismo, que perpetúa un enfoque estrecho que "exprime los complejos problemas sociotécnicos en el ámbito del diseño y, por tanto, entre las manos de las empresas tecnológicas. Al ignorar en gran medida los costosos entornos de producción que requiere el aprendizaje automático, los reguladores fomentan un modelo expansionista de infraestructuras informáticas impulsado por las *big tech*.

Al centrarse en la IA desde una perspectiva mayormente basada en el producto que sólo tiene repercusiones individuales, los enfoques basados en el desprejuiciamiento ocultan en gran medida las economías políticas que sustentan el sector de la IA, la extracción, la explotación y la vigilancia racializada, problemas estructurales a lo largo de todo el proceso de producción. Por estas razones, los enfoques basados en el desprejuiciamiento son, en el mejor de los casos, incompletos e ineficaces para abordar el racismo estructural y, en el peor de los casos, perjudiciales en sus intentos de invisibilizar los daños derivados del proceso de producción de la IA. Una de las lecciones de este capítulo es la necesidad crucial de examinar los daños de la IA más allá del "impacto o acceso desproporcionado" a determinados servicios y experiencias basados en la IA y, en su lugar, abordar las economías políticas de la IA.

32 Powles, Julia (2018). The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. *OneZero*: https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53

Enfoques reglamentarios y legislativos de la IA

La reglamentación de la IA ha sido una de las cuestiones centrales de la última media década, sobre todo en Europa. En abril de 2021, la Comisión Europea publicó su propuesta legislativa para regular la IA en la Unión Europea.[33] El Reglamento sobre Inteligencia Artificial de la UE, que entró en vigor en agosto de 2024, trata de regular la inteligencia artificial mediante un proceso de categorización de riesgos, mecanismos de seguridad de los productos y medidas limitadas de gobernanza y rendición de cuentas. En su versión definitiva, la Ley prohíbe algunos usos limitados de la IA, califica otros de "alto riesgo" y establece un sistema de control y observancia nacional e internacional, que supervisa en gran medida los procesos por los que los productos de IA acceden al mercado de la UE.

Una forma de contestación a los enormes daños causados por los sistemas de IA fue influir en esta reglamentación. Una coalición de 150 organizaciones de la sociedad civil trató de abogar por las líneas rojas, la responsabilidad y la transparencia, así como los mecanismos de reparación para contrarrestar los daños de la IA, derivados de la vigilancia masiva, el impacto medioambiental, la desarticulación estructural, las implicaciones para la democracia y otros aspectos.[34] En particular, las peticiones de prohibir legalmente algunas formas de IA —incluida la vigilancia biométrica masiva en espacios públicos, la vigilancia policial predictiva y diversos usos de la IA en contextos migratorios que socavan el derecho de asilo— se fundamentaron en el objetivo abolicionista de reducir el alcance, la escala, la legitimidad y las herramientas de que disponen el sistema de justicia penal y la policía.

El texto final demostró los límites del uso de la incidencia legislativa como práctica de resistencia significativa. El texto calificaba de "alto riesgo" algunos usos nocivos de la IA en el contexto de la migración, pero no explicaba cómo los sistemas de IA exacerban la violencia y la discriminación contra las personas en los procesos migratorios y las fronteras. Ante la oportunidad de limitar significativamente el uso de la IA para llevar a cabo una vigilancia masiva y discriminatoria de las comunidades marginadas, los legisladores de la UE fracasaron rotundamente al no incluir las salvaguardias necesarias, en particular en los ámbitos de la seguridad, la vigilancia policial y el control de la migración. La Ley se quedó corta a la hora de prohibir las peores formas de vigilancia policial predictiva, vigilancia biométrica y usos dañinos de la IA en el contexto de la migración.

Además, la Ley de IA introdujo amplias lagunas normativas, eximiendo en gran medida a la policía, el control de la migración y los agentes de seguridad de los requisitos de transparencia pública y rendición de cuentas impuestos a los agentes "de alto riesgo", diseñados específicamente para dejar fuera del escrutinio y las obligaciones a la policía, los agentes de seguridad y el control de la migración. De este modo, los legisladores de la UE consolidaron el actual estado de opacidad en el que los agentes estatales emplean las tecnologías de vigilancia para controlar, clasificar, ordenar y castigar a las personas. Además, en muchos sentidos, las comprobaciones técnicas mínimas exigidas en el texto de (un conjunto limitado de) sistemas de alto riesgo en el control de la migración podrían considerarse como facilitadoras, en lugar de proporcionar salvaguardias significativas para las personas sometidas a estos sistemas de vigilancia opacos y discriminatorios.

El legado de la Ley de IA en el contexto de la segurización forma parte, por tanto, de una tendencia más amplia que permite y avala el uso de la IA sobre personas racializadas. También en el último mandato de la UE vimos cómo el Pacto sobre Migración de la UE refrendaba y ampliaba la vigilancia y criminalización de las personas migrantes. El Pacto final adopta muchas medidas para potenciar los sistemas, inversiones e infraestructuras digitales utilizados para prevenir y controlar la migración, como permitir prácticas tecnológicas intrusivas en la tramitación del asilo, ampliar la introducción de la gestión tecnológica de los centros de

33 https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en
34 https://edri.org/our-work/eu-ai-act-trilogues-status-of-fundamental-rights-recommendations/

detención y expandir un régimen ya vasto de recopilación de datos y control digital de las personas migrantes.[35] En opinión del autor, la repercusión de la Ley de IA y la legislación de securización relacionada con ella es ampliar e incluso hacer necesarias las actividades de vigilancia de la policía y el control de la migración, sentando un precedente para un cambio ideológico más amplio que justifique un menor escrutinio, una menor responsabilidad, menores requisitos para la elaboración de una base de pruebas, y marcos jurídicos más amplios de actuación policial para el uso de datos y tecnologías de vigilancia invasivas.

Hacia la justicia y la resistencia

¿Qué resistencia se puede oponer, más allá de las respuestas empresariales e institucionales a los daños causados por la IA? Son pocas las vías que buscan atacar de raíz las economías políticas del sector de la IA, extractivo, explotador y esencialmente discriminatorio.

La primera consiste en prácticas de movilización, documentación y fortalecimiento del poder de las comunidades afectadas. Una de las dificultades a la hora de enfrentarse a la IA ha sido siempre la opacidad y oscuridad del sector de la IA, que ha evitado y disimulado ampliamente sus conexiones con la vigilancia estatal, la securización y la militarización, o ha presentado soluciones que sólo devuelven el poder al sector de la IA. Para contrarrestar esta situación, necesitamos desarrollar y dotar de recursos al trabajo comunitario, los movimientos por la justicia racial y de las personas migrantes, y otras comunidades afectadas, para que comprendan y desarrollen sus propias estrategias de seguridad, control de daños y prácticas liberadoras más amplias. Estos enfoques consideran la lucha contra los usos racistas de la IA como parte integrante de las prácticas más generales contra la securización, las fronteras y el imperialismo, y no como algo independiente.

La siguiente supone ir más allá del impacto desproporcionado y, en su lugar, plantear prácticas de interrupción, perturbación y rechazo en la cadena de suministro de la IA. Tales prácticas vuelven a hacer concretos y visibles los aspectos de explotación, extracción y militarismo de la industria de la IA y la tecnología. Hemos asistido a numerosas protestas populares para impugnar la introducción de la IA en los servicios públicos a nivel nacional como, por ejemplo, las protestas estudiantiles en el Reino Unido contra el algoritmo de calificación del nivel A y el sistema SyRI de detección del fraude en la asistencia social. A escala mundial, vemos cada vez más iniciativas para impugnar la digitalización y la securización, con diversas acciones para interrumpir las cadenas de suministro que envían armamento y solo recursos tecnológicos para apoyar el genocidio en Gaza (como el Proyecto Nimbus) y para impugnar la adquisición recíproca y la legitimación que se produce cuando los Gobiernos occidentales adquieren tecnología de vigilancia a Israel. Aquí, las conexiones entre trabajadores de distintos puntos de la cadena de suministro, incluidos las y los trabajadores tecnológicos de Silicon Valley, como en No Tech for Apartheid, pero también las y los trabajadores portuarios y otros, alineados con los movimientos estudiantiles y de justicia racial, han demostrado la existencia de poderosas constelaciones de resistencia en favor de la abolición de las tecnologías carcelarias y el uso de la tecnología como complicidad colonial.

Dedicaré mis últimas palabras a las iniciativas de redistribución. El sector de la IA se nutre de una poderosa narrativa de progreso e innovación que justifica inversiones privadas multimillonarias, pero también el uso de presupuestos públicos para desarrollar y adquirir sistemas de IA en distintas esferas. Esto suele estar intrínsecamente relacionado con la securización y la militarización, como lo está la historia de la disciplina de la IA. Las iniciativas de redistribución, no sólo de la IA sino de todos los sectores basados en el castigo, la contención, la violencia y la guerra, son una parte esencial de lo que podría ser la resistencia. Podemos construir estrategias significativas que alejen los recursos y la voluntad política de la extracción, la explotación, la criminalización y el control, para dirigirlos hacia sistemas sólidos de atención social, cuidados y distribución de la riqueza.

35 https://www.equinox-eu.com/wp-content/uploads/2024/04/The-Migration-Pact-ProtectNotSurveil.pdf

Conclusión: raza y resistencia, hacia la justicia

Evaluar la IA desde la perspectiva de la justicia racial exige múltiples cambios. Para ello, es necesario dejar de ver la IA como un producto o beneficio en lugar de como un sector, centrarse en los procesos de producción, explotación y militarización en lugar de limitarse al "impacto desproporcionado" y reducir nuestra dependencia de los mecanismos corporativos e institucionales de "reparación". Como se ha expuesto en este capítulo, la IA forma parte integrante de infraestructuras de opresión más amplias: la criminalización, la securización y el capitalismo racial. Las prácticas de resistencia requieren una lucha inequívoca hacia el rechazo de los mercados capitalistas del "proceso" tecnológico, las iniciativas de despenalización y desmilitarización, un modelo de redistribución económica de la riqueza y la reconstrucción de infraestructuras comunitarias de atención, conexión y asistencia social.

6. El impacto de la inteligencia artificial del lenguaje en el acceso al conocimiento y su producción

Pelonomi Moiloa
Lelapa AI
Traducción: Arantxa Albiol Benito

No reconocer la diversidad lingüística como un aspecto tan bello como esencial de la vida humana acarrea profundas consecuencias. La lengua cumple la función de archivo viviente, preservando vastos conocimientos, prácticas culturales y visiones únicas del mundo. Cuando las lenguas se marginan o se pierden, no sólo perdemos este inestimable depósito de conocimientos, sino también la capacidad de comprender cómo las comunidades se conectan y se relacionan entre sí a través de sus expresiones lingüísticas. Además, la lengua configura nuestra forma de pensar, ya que ofrece distintos enfoques para la resolución de problemas que pueden impulsar la innovación colectiva. Dentro de esta diversidad se hallan pistas que pueden profundizar nuestra comprensión de nosotros mismos y de nuestra humanidad común.

El presente documento examina estas ideas a través del prisma de la tecnología del lenguaje africana, mostrando cómo el hecho de abordar la condición de 'bajo recurso' de estas lenguas puede ser el catalizador de un cambio transformador en el desarrollo de la tecnología del lenguaje. Contrariamente a los enfoques dominantes, que suelen dar prioridad a la eficacia y la rentabilidad, el contexto de las lenguas africanas requiere metodologías más inclusivas y comunitarias. Este cambio abre la puerta a un paradigma que no sólo beneficia a los hablantes de las lenguas infrarrepresentadas, sino también a los ecosistemas tecnológicos mundiales, al proporcionar modelos diversos y sostenibles para el futuro.

Introducción

En la publicación *Vanishing Voices* de National Geographic, Russ Rymer afirma que "cada 14 días muere una lengua, y para el próximo siglo es probable que desaparezca casi la mitad de las 7000 lenguas del planeta"[1]. Algunos podrían argumentar que se trata simplemente de una evolución natural del mundo. Si bien es cierto que las lenguas se adaptan para reflejar las necesidades del presente, evolucionando a medida que la vida misma se transforma y cambia a lo largo del tiempo y el espacio, me atrevería a afirmar que este proceso no es totalmente orgánico. Al igual que el cambio climático —un fenómeno inherentemente natural que los seres humanos han acelerado a un ritmo alarmante—, estamos provocando activamente cambios lingüísticos mucho más rápido de lo que nuestras capacidades nos permiten comprender o responder a sus implicaciones a largo plazo con una tecnología de IA desarrollada principalmente para y por angloparlantes. La forma en que el desarrollo actual de las lenguas de alto recurso en PLN abre una brecha cada vez mayor está bien estudiada en documentos como *The Zeno's Paradox of 'Low-Resource' Languages* de Atnafu Tonja et al (2024) y *The state and Fate of Linguistic Diversity and Inclusion in the NLP World* de Pratik Joshi et al (2020).

Para aquellos cuyos antepasados fueron asimilados a culturas homogéneas a través del lenguaje hace siglos, la violencia de tales actos puede parecer ahora lejana, quizá incluso olvidada. Sin embargo, para quienes han experimentado estos patrones más recientemente, las heridas aún están frescas. Aunque la lengua —y en particular la homogeneización de la misma— puede actuar como fuerza unificadora al posibilitar la comunicación, también es una poderosa tecnología que se ha convertido en un arma de dominación, control y marginación.

Durante siglos, en el contexto del colonialismo, se impusieron lenguas homogéneas a los grupos indígenas, interrumpiendo la continuidad cultural a cambio de lo que se denominó "llave de acceso". El dominio del lenguaje del colonizador se convirtió en un requisito previo para la educación, el gobierno, la movilidad socioeconómica y la inclusión, creando una jerarquía que perjudicaba sistemáticamente a los hablantes nativos. La lengua también se ha utilizado como medio para controlar el discurso y perpetuar la «otredad», reforzando las desigualdades y menoscabando las identidades de quienes no pertenecen al grupo lingüístico dominante.

La homogeneización de la lengua, aunque práctica en apariencia, conlleva un legado de violencia que sigue configurando la dinámica del poder mundial en la actualidad. Si no tenemos cuidado, este legado podría influir en la trayectoria de la tecnología del lenguaje y dictar lo que nos jugamos. En este punto de inflexión crítico, resulta indispensable reconocer el inmenso valor de la diversidad lingüística y reflexionar sobre cómo podemos moldear intencionadamente las tecnologías del lenguaje que creamos para garantizar que dichas tecnologías honren y preserven la riqueza de las lenguas a medida que evolucionan con el tiempo.

1 "Vanishing Voices" de Russ Rymer en *National Geographic* (Julio de 2012): https://www.nationalgeographic.com/magazine/article/vanishing-languages

Parte 1: Consecuencias de la pérdida de diversidad lingüística

> "Cuando mueren las lenguas, todo un edificio de conocimiento humano, laboriosamente ensamblado a lo largo de milenios por un número incontable de mentes, se está erosionando, desvaneciéndose en el olvido".
>
> K. David Harrison

La lengua evoluciona con el tiempo entre grupos con experiencias comunes y refleja el modo de vida de un grupo. Las distintas lenguas ponen de relieve las variedades de la experiencia humana, revelando aspectos mutables de la vida que tendemos a considerar universales: nuestra experiencia del tiempo, los números, la pertenencia, la tecnología y cómo la construimos. Parafraseando la obra de Achille Mbembe, "La lengua es un archivo vivo". Cuando se olvidan las lenguas, también se olvida este archivo.

La lengua como archivo vivo

La pérdida de la lengua conlleva la pérdida de la comprensión ecológica, el uso de plantas medicinales, las prácticas agrícolas, las creencias espirituales y la historia que a menudo son específicas de una región y una cultura. Esto incluye el conocimiento de los animales, las fases de la luna, los patrones del viento, la manera en que las personas se relacionan entre sí y qué partes de esa relacionalidad consideran significativas. Las sociedades que dependen de la naturaleza para sobrevivir han desarrollado tecnologías para cultivar, domesticar y utilizar dichos recursos. En el libro *Cuando mueren las lengua*s, K. David relata que gran parte de lo que la humanidad sabe sobre el mundo natural, la supervivencia y cómo nos relacionamos se encuentra completamente fuera de las estructuras formales de la práctica archivística. No está en los libros de texto científicos, ni en las enciclopedias, ni en las bases de datos, sino que muchas veces sólo existe en la memoria de las personas. De hecho, muchas veces recurrimos a las fuentes de conocimiento autóctonas para colmar las lagunas de nuestra comprensión del mundo. Parte de esta información se ha cosechado en beneficio de unos pocos elegidos. Se calcula que las empresas farmacéuticas obtienen unos beneficios anuales de 85 000 millones de dólares con medicamentos derivados de plantas conocidas por primera vez por los pueblos indígenas por sus propiedades curativas. Como K. David Harrison explica: "a medida que desaparecen las lenguas, también lo hace este intrincado conocimiento, disminuyendo la comprensión global de la humanidad sobre la diversidad cognitiva y las relaciones del ecosistema".

Las lenguas actúan como archivos culturales, reflejando las creencias y prácticas de las comunidades que las hablan. En el África Subsahariana, las tradiciones orales, como la narración de cuentos, no son sólo un entretenimiento, sino que resultan indispensables para preservar la historia, los valores y la identidad comunitaria. En estas sociedades, por ejemplo, la lengua y la narración reflejan y refuerzan las expectativas sociales y los valores comunitarios. A diferencia de las lenguas adoptadas, que pueden no ser portadoras de estas normas culturales, las tradiciones orales africanas están profundamente entrelazadas con la identidad y la memoria colectivas.

Un rasgo definitorio de la narración africana es la presencia de una dinámica de llamada y respuesta entre el narrador y el público. Unas frases o palabras específicas señalan el principio de una historia e invitan al público a participar en los lenguajes de todo el continente. **En la narración swahili, por ejemplo, el narrador empieza con 'Hadithi hadithi', que significa 'Historia historia', y su público responde 'Hadithi njoo', que significa 'Ven historia'. Este intercambio no sólo señala el principio del cuento, sino que también refuerza la participación comunitaria, conectando al público con su patrimonio cultural**.

Contrariamente a expresiones indoeuropeas como 'Érase una vez', que se centran en narrativas individuales, estas frases africanas hacen hincapié en las experiencias compartidas. La interacción va más allá de las palabras, incluyendo a menudo canciones, cantos o danzas y convirtiendo la narración en un ritual comunitario que conecta a la gente con su entorno, sus ancestros y su legado cultural. Los narradores de estas tradiciones, como los griots o los imbongi, actúan como historiadores, genealogistas y custodios culturales, preservando los valores sociales y la historia. Esta tradición oral favorece los estilos de comunicación indirectos, en los que muchas veces el significado se transmite a través de la metáfora, la alegoría y el simbolismo. Sus métodos dan prioridad a la conservación de la armonía social y dejan espacio para la interpretación individual.

Las culturas que dan menos importancia a la narración como medio de transmisión cultural tienen menos probabilidades de desarrollar estos hábitos de comunicación interactiva e indirecta. La tradición oral africana pone de relieve la profunda interdependencia entre la narración de historias, la cultura y la comunidad, recalcando la manera en que estas prácticas preservan la identidad colectiva y la cohesión social. Cuando las personas de una cultura adoptan la lengua de otra, suelen experimentar un cambio de identidad. Muchos pueden sentir este giro como una pérdida de su propia identidad, como se analiza en obras como *Lost in Translation: A Life in a New Language de* Eva Hoffman y *Decolonising the Mind: The Politics of Language in African Literature* de Ngũgĩ wa Thiong'o.

Además del entorno interno de una comunidad, el lenguaje evoluciona también en estrecha interacción con el entorno externo, adaptándose para optimizar la comunicación en condiciones específicas. La variación del timbre o el nivel del tono, por ejemplo, hace que los lenguajes codifiquen eficazmente el significado y la emoción. Esto vuelve la comunicación verbal más expresiva y precisa sin necesidad de añadir palabras. En lenguas tonales como el yoruba, el significado de una sola sílaba puede cambiar totalmente en función de su tono. Las adaptaciones acústicas en las lenguas también reflejan las influencias del entorno. Por ejemplo, los investigadores plantean la hipótesis de que los sonidos vocálicos viajan de forma más eficaz en el aire húmedo que en el aire seco, lo que podría contribuir a las diferencias entre los lenguajes del sur y del norte de la India. Del mismo modo, las consonantes de chasquido de las lenguas khoisan son agudas y percusivas, por lo que siguen siendo inteligibles a grandes distancias en entornos naturales abiertos, una adaptación ventajosa en situaciones al aire libre. Estos ejemplos ilustran cómo los rasgos lingüísticos se desarrollan en respuesta a las necesidades ambientales y sociales, garantizando que las expresiones verbales se adapten a los contextos específicos en los que surgen.

La lengua moldea profundamente la forma de pensar y resolver los problemas, actuando como una extensión de los procesos cognitivos y las ideas. La estructura de la lengua china está estrechamente ligada a las regiones cerebrales que participan en el razonamiento visoespacial, lo que puede potenciar las capacidades matemáticas. Quienes hablan árabe suelen desarrollar sólidas capacidades de reconocimiento de patrones abstractos y jerárquicos cuando se les forma para detectar patrones dentro de formas de palabras complejas. Los hablantes de lenguas aglutinantes* también pueden experimentar ventajas cognitivas en contextos de aprendizaje multilingüe. Un estudio de Nkolola-Wakumelo (2008) en *African Multilingualism* sugiere que los hablantes de lenguas bantúes como el zulú muestran una mayor sensibilidad a la estructura de la lengua y desarrollan grandes competencias en el reconocimiento de patrones lineales. "Las lenguas revelan los límites y las posibilidades de la cognición humana: cómo funciona la mente. Cada nuevo patrón gramatical que encontramos aporta claridad sobre cómo el cerebro humano crea la lengua"., K. David Harrison

Cuando las comunidades abandonan sus lenguas y se pasan al inglés o el español, se produce una interrupción enorme de la transferencia de conocimientos tradicionales, que pueden perderse en gran parte en solo una o dos generaciones. Esta pérdida se produce directamente en las palabras que utilizamos para describir el mundo físico y nuestra ex-

periencia con él. Pero este conocimiento también se pierde o se distorsiona en la manera en que las culturas organizan ese conocimiento y lo transfieren de una persona a otra, y en cómo configura la forma en que abordan la resolución de problemas y la comprensión del yo.

**Las lenguas aglutinantes se describen como lenguas en las que las palabras suelen formarse uniendo sufijos y prefijos a las palabras raíz.*

Perpetuación del sesgo y exclusión de poblaciones en los grandes modelos de lenguaje (LLM)

La pérdida de una lengua no es sólo la pérdida de palabras, sino también de conocimientos, cultura y patrimonio. Esto es algo sumamente trágico, pero las implicaciones para quienes mantienen estas lenguas son complejas. En un mundo en el que la tecnología del lenguaje está acelerando el proceso de homogeneización de las lenguas, ¿qué ocurre con quienes se resisten a esta tendencia?

La exclusión lingüística da lugar a un pobre desempeño de estas herramientas de lenguaje. Sin una representación sólida de varias lenguas, los LLM (grandes modelos de lenguaje, por sus siglas en inglés) no cumplen su cometido en aplicaciones multilingües, como las traducciones o las tareas específicas de cada lengua. En *Unsupervised Cross-lingual Representation Learning (2020),* Conneau et ál. explican que, sin las lenguas de bajo recurso, los LLM no son capaces de generalizar eficazmente, como se ha visto en las tareas de TA (traducción automática), en las que los modelos tienen problemas con las lenguas poco comunes, lo que da lugar a traducciones de mala calidad e incluso a la posible pérdida de estructuras gramaticales específicas de ciertas lenguas. En un caso tristemente célebre, el modelo lingüístico de Facebook tradujo erróneamente «Buenos días» del árabe a «Atáquenlos», lo que dio lugar a la detención injusta de un palestino. **Este incidente pone de manifiesto los peligros de la infrarrepresentación de determinadas lenguas en la IA, donde los pequeños errores pueden tener graves consecuencias en el mundo real**.

Excluir las lenguas del desarrollo de la tecnología moderna a menudo conduce a resultados sesgados que favorecen a las lenguas dominantes, como el inglés. Por ejemplo, un modelo entrenado en inglés puede no interpretar significados sensibles al contexto o matices culturales en contextos no ingleses. En Bender et ál. (2021), *On the Dangers of Stochastic Parrots*, los investigadores argumentan que al entrenar a los LLM predominantemente en lenguas de alto recurso, las respuestas del modelo se vuelven sesgadas, marginando así a las lenguas y perspectivas culturales infrarrepresentadas, llegando incluso a asignar etiquetas emocionales negativas a los nombres negros en actividades de etiquetado emocional.

Cada vez hay un mayor reconocimiento académico del hecho de que la falta de asistencia en lengua local en las redes sociales mundiales y los grandes modelos de lenguaje puede crear vulnerabilidades significativas en la integridad de la información, especialmente en las regiones marginadas. Esta laguna no sólo dificulta la comunicación y la educación efectivas, sino que también abre las puertas a campañas de desinformación.

Estudios como *La desinformación digital en África* (Shahbaz & Funk, 2019) revelan cómo la limitación de la inclusividad lingüística en las plataformas digitales permite a los actores malintencionados explotar las lagunas informativas, difundiendo falsedades en las lenguas locales, donde no suele haber moderación. Esta laguna hace que la desinformación se convierta en un arma que escapa a la detección de los algoritmos de moderación de contenidos, entrenados principalmente en las lenguas dominantes. Durante las elecciones de Nigeria en 2023 y de Kenia en 2022, por ejemplo, hubo campañas de desinformación en dialectos locales en plataformas con escasa capacidad de detección lingüística. Esto dio lugar a que se difundieran sin control discursos que, de otro modo, habrían sido calificados

de falsos en inglés, lo que pudo influir en el comportamiento de los votantes y fomentar tensiones divisorias.

Un estudio de Mozur et ál. (2018) analiza cómo se aprovecharon los puntos ciegos lingüísticos de los algoritmos de las redes sociales para alimentar la violencia étnica en Myanmar. La falta de herramientas de tratamiento de la lengua birmana en plataformas como Facebook en aquel momento contribuyó a que la incitación al odio y la información falsa dirigida a la comunidad rohinyá se extendieran sin control, lo que demuestra que la ausencia de moderación en la lengua local puede tener consecuencias devastadoras en el mundo real.

Para quienes conservan su lengua, esto supone una tarea doble: salvaguardar su patrimonio lingüístico y, al mismo tiempo, participar en una sociedad globalizada e impulsada por la tecnología. El paisaje cambiante de la tecnología del lenguaje plantea riesgos significativos para la diversidad cultural, suscitando interrogantes sobre el futuro de la comunicación. Sabemos que quedar al margen de las revoluciones tecnológicas esenciales repercute significativamente en el nivel y la calidad de vida, por lo que cabe preguntarse qué hay que hacer para que la tecnología del lenguaje sea más inclusiva.

Parte 2: Las lenguas africanas en los modelos de lenguaje.

Las personas que hablan lenguas indoeuropeas pueden experimentar pequeños inconvenientes cuando les falla la tecnología del lenguaje, como que el autocorrector cometa errores incómodos, que el software de reconocimiento de voz tenga problemas con los acentos o que las aplicaciones pronuncien mal sus nombres. Sin embargo, estos problemas suelen ser pequeños y no alteran significativamente su vida cotidiana. Para quienes hablan lenguas ajenas a este grupo dominante, la situación es completamente distinta. En regiones con infraestructuras limitadas, especialmente en África, la tecnología del lenguaje puede ser el factor determinante para acceder a las necesidades básicas: comprar electricidad, recibir atención médica vital o entender la comunicación oficial más importante. África, con sus aproximadamente 2000 lenguas, representa un tercio de la diversidad lingüística mundial, y sin embargo estas lenguas siguen estando muy insuficientemente atendidas por la tecnología actual.

La razón por la que los africanos no pueden acceder al privilegio de la comunicación en estos casos se reduce a la disponibilidad de tecnología del lenguaje en las lenguas deseadas y existen grandes obstáculos para superar esta brecha.

Dónde fallan los modelos de lenguaje en las lenguas africanas:

Construir modelos de lenguaje para las lenguas africanas es complicado. La comprensión de los errores cometidos por estos modelos pone de manifiesto algunas dificultades fundamentales en su desarrollo. A continuación mencionaré algunas de ellas.

Cambio y mezcla de códigos: Un rasgo fundamental de las lenguas del continente africano es la prevalencia de la mezcla y el cambio de códigos, que refleja la diversidad lingüística y cultural de la región. La mezcla de códigos implica pasar de una lengua a otra, lo que suele ocurrir en las regiones multilingües, donde los individuos hablan hasta seis lenguas. Por ejemplo, un hablante puede empezar una conversación en inglés y cambiar a otra lengua por comodidad o expresión emocional. Esto suele dar lugar a frases completas en lenguas distintas.

Por el contrario, el cambio de código consiste en utilizar varias lenguas en la misma frase. Por ejemplo, cuando nuestro equipo investigó una telenovela sudafricana de máxima audiencia para un estudio, descubrió que en un solo episodio aparecían ocho lenguas distintas, algo habitual en África pero poco frecuente en contextos lingüísticos indoeuropeos. Esta adaptabilidad lingüística refleja la singular coexistencia multicultural del continente, pero no es una capacidad inherente a los actuales sistemas típicos de tecnología del lenguaje.

Escasez de datos: La dificultad más patente, y la que recibe mayor atención, es la falta de datos suficientes, en concreto, de grandes colecciones de ejemplos de texto y audio para entrenar a las máquinas. La cultura mayoritariamente oral de África, unida a las prohibiciones históricas de recopilar este tipo de datos durante la dominación colonial, se traduce en una gran escasez de corpus de texto y audio en las lenguas africanas. Las mayores recopilaciones existentes suelen ser traducciones de la Biblia, por lo que los modelos en lenguas africanas suelen tener un tono más evangélico que otros.

Residuos coloniales en la lengua: Muchos de los errores hallados en los modelos de lenguaje pueden atribuirse a la historia colonial de las lenguas africanas escritas. Ante todo, África es un lugar de tradición y cultura orales. La escritura rara vez fue a la par del desarrollo de las lenguas. Por eso, cuando los misioneros llegaron a África, trataron de registrar la lengua y adaptar el alfabeto romano a los sonidos que oían. Por supuesto, el alfabeto romano no podía englobar completamente los sonidos de una lengua. Y para los oídos inexpertos, muchos de los sonidos de las lenguas africanas suenan igual... cuando no son lo mismo. Mi nombre, por ejemplo, tiene dos oes. En la lengua tswana hay ocho formas de pronunciar la "o". Pero todas se escriben igual. En realidad, las dos oes de mi nombre no se pronuncian igual.

La cuestión refleja la función histórica de los misioneros y antropólogos de países europeos, que en muchos casos fueron los primeros en registrar por escrito las lenguas africanas. Dependiendo de si el misionero que se encontraba con un grupo era francés, inglés o alemán, la ortografía del nombre de ese grupo podía ser distinta. Por ejemplo, mi apellido tiene variaciones como "Moiloa" y "Moilwa", según qué misionero europeo encontró primero a nuestra familia. La incoherencia en la pronunciación de las vocales, a pesar de una ortografía idéntica, añade otra capa de complejidad que los modelos de lenguaje no están preparados para abordar.

Norma indoeuropea de modelos de PLN

El desarrollo actual de los modelos de lenguaje se basa en supuestos fundamentales que proceden en gran medida de las lenguas indoeuropeas, el grupo lingüístico al que estaban destinados a servir inicialmente. Por consiguiente, estos modelos suelen carecer de la flexibilidad necesaria para adaptarse eficazmente a lenguas ajenas a estas estructuras y convenciones lingüísticas.

La tokenización, el proceso de dividir el texto en unidades más pequeñas (como palabras o subpalabras), es sencilla en inglés, pero mucho más compleja en lenguas aglutinantes como el zulú. En zulú, una sola palabra como "Ngiyafunda" ("Estoy aprendiendo") transmite varias capas de significado, lo que requiere técnicas especializadas de tokenización para abordar su estructura con eficacia. Esto significa que el supuesto fundamental de cómo aplicar matemáticas complejas de aprendizaje automático fallan, cuando la lengua no sigue las normas estructurales a las que están acostumbrados los modelos de lenguaje. Con rompecabezas como el cambio de código, un modelo tiene que entender qué método de tokenización es mejor utilizar palabra por palabra en función de la lengua a la que pertenezca la palabra, lo que complica aún más las cosas si el propio tokenizador no es "multilingüe".

Los modelos de lenguaje también tienen supuestos culturales subyacentes. Un ejemplo de ello son las culturas de comunicación directa frente a las culturas de comunicación indirecta mencionadas anteriormente. En las culturas de comunicación directa, el significado de una lengua está explícito en las palabras. En las culturas de comunicación indirecta, el significado está implícito. La mayor contribución del texto a la tecnología del modelo de lenguaje es la cultura estadounidense de comunicación directa. Los africanos no se comunican de este modo, y recurren a menudo al contexto y a la sutileza para transmitir un significado que las máquinas no captan bien, sobre todo cuando no abundan los ejemplos para ilustrar cómo podría transmitirse la sutileza.

Dejando a un lado las complejidades lingüísticas, la economía unitaria de desarrollar y mantener esta tecnología también puede ser difícil. Además de necesitar un gran volumen de datos, el desarrollo de esta tecnología requiere grandes cantidades de capital. Entrenar uno de los mayores modelos generativos del mundo cuesta más dinero que el PIB de 16 países africanos. Los grandes modelos de lenguaje también requieren una cantidad absurda de cálculos y, por tanto, consumen mucha energía. Entrenar uno de los mayores modelos generativos del mundo utiliza la misma cantidad de electricidad que 12.000 hogares de Johannesburgo durante un mes. Una ciudad que ha sufrido apagones periódicos intencionados debido a restricciones de capacidad hasta hace poco. Se estima que una sola consulta a estos grandes modelos de lenguaje utiliza una botella de agua. Además, el tipo de procesamiento necesario para esta tecnología incluye el procesamiento en la GPU en lugar de en la CPU. El procesamiento en la GPU requiere muchas más tierras raras y metales preciosos. Uno de ellos es el tantalio, un mineral derivado del coltán, uno de los principales responsables de la guerra civil y los enfrentamientos violentos actuales en el Congo. Estas condiciones no son adecuadas para el contexto africano. Pero, para ser sinceros, estas condiciones tampoco son adecuadas para un mundo sostenible. Se necesita un enfoque distinto para que estos modelos funcionen y se amplíen a un mayor número de personas.

Aunque se estima que el coste de producción de los modelos disminuirá significativamente con el tiempo, dado que los mejores modelos de estilo GPT sólo procesan unas 100 veces la cantidad de información que podría procesar un niño en su primer año de vida*, y que los experimentos actuales[2] han demostrado que el hecho de disponer de más datos no aumentará necesariamente su capacidad de forma significativa, se estima que el próximo gran avance en las tecnologías de estilo de aprendizaje profundo para campos como los modelos multimodales (modelos que aprenden de distintos medios y no sólo de datos de texto) o en el razonamiento de modelos requerirá un orden de magnitud superior a los cientos de millones de dólares de ChatGPT.

**En la Cumbre Mundial de Gobiernos celebrada en Dubái a principios de 2024, Jan Lacoon explicó la poca información que es capaz de tratar un gran modelo de lenguaje en comparación con un niño al cumplir los 4 años. A una velocidad de tratamiento de 20 megabytes por segundo, lo que equivale a 1 de los 13 tokens, al mayor LLM moderno le costaría 150 años tratar esta información.*

2 Muenninghoff, N. et al. (2023). Scaling Data-Constrained Language Models. En: Oh, A. et al. *Advances in Neural Information Processing Systems 36* (NeurIPS 2023). Disponible en: https://proceedings.neurips.cc/paper_files/paper/2023/hash/9d89448b63ce1e2e8dc7af72c984c196-Abstract-Conference.html

Parte 3: El beneficio de trabajar reconociendo la diversidad.

Con las limitaciones de datos y recursos informáticos, así como la complejidad del problema de las lenguas africanas, se podría suponer que la creación de tecnología para estas lenguas no es viable. Los supuestos, sin embargo, solo son válidos si nos ceñimos a los marcos normativos utilizados en el desarrollo de los modelos de lenguaje occidentales. En realidad, no sólo se están creando modelos de lenguaje adaptados a estos contextos, sino que además estos se están desarrollando eficazmente. El mundo tiene mucho que aprender de la innovación y la creatividad que afloran en entornos de bajos recursos. Los conocimientos adquiridos a través del desarrollo de la tecnología en estos entornos, junto con formas alternativas de conocimiento, ofrecen nuevas perspectivas sobre la manera en que se puede dar forma a la tecnología.

Localización y adaptación de soluciones

La localización de soluciones ha encontrado fallos en la corriente principal de la teoría de la tecnología del lenguaje, y lo ha hecho en diversas formas. Me gustaría centrarme en tres de ellas que me parecen especialmente interesantes.

La primera forma es mediante la construcción de modelos de lenguaje orientados a la aplicación, en lugar de modelos de lenguaje orientados a la investigación. El ámbito de la aplicación exige que los modelos sean precisos para su uso en el mundo real (es decir, que sean capaces de gestionar cosas como la mezcla y el cambio de código), que también sean rápidos para las aplicaciones en tiempo real Y que no consuman grandes recursos, ya que la mayoría de los creadores de aplicaciones no disponen de recursos informáticos ni financieros para mantener los costes de funcionamiento de GPU costosas. Por ejemplo, en el continente africano sólo hay unos pocos centros de datos que ofrezcan procesamiento en la GPU. Por lo tanto, los recursos de procesamiento en la GPU para desarrollar y mantener aplicaciones son muy escasos. Del 84 % de la población subsahariana que tiene acceso a Internet a través de 3G, un estudio reveló que sólo el 22 % utiliza realmente los servicios de Internet móvil. Si a esto añadimos que el acceso a Internet en el continente africano se produce principalmente a través de los teléfonos móviles (y no se trata de teléfonos inteligentes, sino de teléfonos básicos), lo que realmente tiene más sentido es garantizar que el procesamiento personalizado pueda llevarse a cabo a través de diminutos ML en los dispositivos periféricos. Esto resuelve el problema del procesamiento, pero también permite reducir la latencia, mejorar la privacidad, utilizar menos ancho de banda (ya que no dependería de la conexión a Internet) y, sobre todo, que las aplicaciones puedan estar disponibles sin conexión. Lo que también ayuda a fomentar es el desarrollo en el espacio del aprendizaje federado y los sistemas continuos de computación distribuida (DCCS). El aprendizaje federado se centra en el aprendizaje automático a través de dispositivos descentralizados, manteniendo los datos localizados. Mientras que el DCCS se refiere a un concepto más amplio de distribución de tareas computacionales a través de múltiples sistemas para la computación de uso general.

Un enfoque eficaz para crear modelos más pequeños y rentables es asegurarse de que están especializados para las tareas específicas que están destinados a desempeñar. En lugar de confiar en grandes modelos de lenguaje de uso general, capaces de cumplir una gran variedad de funciones, podemos centrarnos en eliminar las funciones innecesarias, dejando sólo lo que resulta esencial para la tarea que tenemos entre manos. Este enfoque no sólo reduce el tamaño del modelo y el coste de procesamiento, sino que también disminuye el riesgo de uso indebido al garantizar que el modelo se diseña con fines claros y específicos. Además, los modelos más pequeños y específicos para cada tarea son más fáciles de interpretar, lo que mejora la transparencia y permite comprender mejor los procesos de toma de decisiones. Esto contrasta con el paradigma actual, en el que se espera que los modelos de lenguaje generativos se generalicen en una gran cantidad de tareas en

niveles inferiores. El desarrollo orientado a la aplicación sugiere que los modelos especializados pueden ser más prácticos y eficaces que los generalistas.

La segunda forma en que la localización es fundamental para el cambio innovador es mediante la adopción del principio de "Basura dentro es basura fuera" como si fuera el evangelio. Este principio significa que la calidad del producto de un sistema o proceso depende de la calidad del insumo. Normalmente, el desarrollo de grandes modelos de lenguaje no se ha atenido a este principio. Los modelos que requieren muchos datos se alimentan de todas y cada una de las pruebas lingüísticas a las que puede echar mano un desarrollador de modelos, tanto si se trata del lado oscuro de Internet como si no, tanto si los datos etiquetados son un fiel reflejo de la lengua como si no. Aunque ha demostrado ser útil para las lenguas indoeuropeas de las que se dispone de abundantes datos, ha sido un fracaso absoluto para las lenguas de bajo recurso. Inmersiones profundas en juegos de datos "disponibles públicamente" de lenguas africanas por parte de investigadores como Timnit Gebru et al en el artículo *Combating Harmful Hype in Natural Language Processing* o *Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets de* Julia Kreutzer et ál., han descubierto que los juegos de datos están formados por productos sin sentido que no se parecen en nada a la lengua.

El enfoque en el espacio de bajo recurso es el enfoque de preservación de datos combinado con el enfoque de análisis de errores como el descrito e*n Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages de Masakhane*. Se trata de un enfoque inclusivo, que cuenta con expertos lingüísticos y culturales para recopilar la experiencia de un modelo y evaluar los resultados de los modelos. La preservación de los datos es necesaria debido a su escasez. Como ya se ha mencionado, los juegos de datos disponibles públicamente para las lenguas africanas suelen ser bíblicos. Entrenar los modelos de lenguaje con estos datos puede perpetuar la percepción de que todos los africanos son evangélicos y que eso es todo de lo que nos interesa hablar. La preservación de datos garantiza que captamos los conceptos fundamentales que forman parte de una lengua, su patrimonio y su cultura, y podemos asegurarnos de que se cubren ámbitos específicos para que un modelo aprenda a representar mejor cómo debe ser la comunicación en una lengua concreta. El siguiente paso en este proceso es el análisis de errores. Con ayuda de las competencias culturales y lingüísticas, se examinan en profundidad los fallos del modelo y, por tanto, llegan a comprenderse plenamente. ¿El modelo confunde los homófonos? ¿en qué ocasiones? ¿Se equivoca en los nombres? ¿Se confunde al cambiar de una lengua a otra? ¿Entiende que es una mujer la que celebra una ceremonia determinada? ¿Comprende que una interpretación literal no es adecuada en este caso? Al comprender estos posibles errores, se pueden recopilar datos muy específicos a un coste menor para abordar estos problemas en el desempeño de los modelos.

Estas dos filosofías combinadas reducen fundamentalmente el coste asociado al desarrollo de modelos de lenguaje que puedan funcionar bien cuando se apliquen sobre el terreno para una función concreta. Como un modelo de traducción en un teléfono móvil para hacer encuestas sobre el terreno. O asistencia conversacional para un servicio de consulta sanitaria a distancia. O modelos de transcripción y voz hablada para su uso en el ámbito de la atención al cliente. En el contexto de bajo recurso, un coste bajo significa un acceso democratizado a estos servicios, y un acceso democratizado a estos servicios lingüísticos para las pequeñas, medianas y grandes empresas. En el contexto de alto recurso, el bajo coste significa una reducción positiva de las emisiones de la informática, de la invasión de tierras cultivables por los centros de datos, y una menor presión sobre las redes eléctricas y los sistemas hídricos. También supone una reducción del riesgo y modelos más seguros y fáciles de interpretar.

La tercera de estas formas es a través de cambios fundamentados de las arquitecturas de los modelos y sus supuestos subyacentes. Volviendo al ejemplo de los tokenizadores,

efectuar análisis de errores en estos modelos para comprender la capacidad «cognitiva» subyacente es similar al estudio de las lenguas en los seres humanos y al modo en que estos estudios pretenden encapsular el grado en que los cerebros humanos construyen y comprenden el lenguaje. Gracias a la comprensión de los datos, podemos entender en qué fallan los modelos, y mediante el análisis de errores en los modelos, podemos comprender qué supuestos subyacentes no consiguen captar la complejidad del lenguaje. Esto nos da una idea de a qué partes de estos supuestos subyacentes les cuesta adaptarse a la complejidad del lenguaje del mundo real. Y nos da pistas sobre los aspectos en los que debemos centrarnos para captar mejor esas capacidades en los modelos de lenguaje. Los fallos de los modelos de lenguaje en inglés no son los mismos que en wolof. Desarrollar distintos tokenizadores, por ejemplo, puede mejorar el aprendizaje de los modelos de lenguaje. Quizá no hubiéramos sabido lo importante que es este paso para el proceso de aprendizaje de un modelo de lenguaje si no hubiéramos vivido la experiencia de hacer varias pruebas de esos modelos en lenguas distintas. Podríamos haber llegado a pensar que ya teníamos las respuestas.

Sostenibilidad e inclusividad de las herramientas

El principal beneficio de estos modelos reside en poder atender a más poblaciones distintas mediante la tecnología digital, democratizando así el acceso a los mismos. Este enfoque no sólo pone al descubierto métodos para hacer que la inclusión sea económicamente viable, sino que también encierra un importante potencial para que el desarrollo local dé forma a tendencias más amplias en la IA del lenguaje. Aprendiendo a adaptarnos a las lenguas locales con un volumen mínimo de datos y captando al mismo tiempo matices y coloquialismos, podemos crear una tecnología que satisfaga rápidamente las necesidades actuales a un coste menor. Lo último que queremos es invertir millones en una tecnología que no pueda evolucionar con el tiempo. Aunque los sistemas heredados dispuestos en silos sean defectuosos, al menos se entiende su lógica. Sin embargo, la actualización de grandes modelos como el GPT requiere actualmente una cantidad ingente de recursos, con un coste de millones cada 6-12 meses para actualizaciones de gran magnitud y ajustes precisos frecuentes. Lo que podemos ganar dominando la creación rápida de prototipos de lenguas africanas es un modelo que se adapte rápidamente sin necesidad de inyecciones de capital tan cuantiosas, reduciendo la presión sobre la energía, el agua, el capital y la explotación de datos y minerales necesarios para construir estos modelos.

No sólo significa democratizar el acceso a la tecnología para su uso, sino también para el desarrollo y la adaptación locales. Fundamentalmente, esto pone en tela de juicio los sistemas centralizados de beneficios y poder. Una parte fundamental del debate sobre la inteligencia artificial del lenguaje ha girado en torno al código abierto. El código abierto aspira a la transparencia y la confianza, fomentando la colaboración y la innovación a través de contribuciones de origen colaborativo, así como proporcionando ventajas económicas y accesibilidad, especialmente para quienes no pueden permitirse entrenar las dimensiones iniciales del modelo. Aunque comprendo el principio en que se basa el acceso abierto, también lo encuentro un tanto idealista. Como práctica equitativa, el código abierto presupone que todos los participantes parten del mismo nivel, pero este supuesto es erróneo. La publicación de modelos de código abierto presupone que los usuarios disponen de numerosos datos y de acceso a recursos informáticos, que no están disponibles en muchas regiones en desarrollo. Las que dominan las contribuciones y se benefician de las herramientas de código abierto son principalmente las organizaciones ricas en recursos. El statu quo en el desarrollo de la IA está arraigado en la idea de que debe estar centralizada y monopolizada porque sólo las entidades ricas pueden inyectar el capital necesario para su evolución. Para contrarrestar esta situación, debemos desarrollar formas realistas y descentralizadas de organizar la tecnología, armonizadas con la distribución equitativa del poder, el reparto de los beneficios y el bien colectivo.

Un ejemplo de ello es la concesión equitativa de licencias de datos, como la licencia de datos abiertos de Nwulite Obodo[3] y la licencia Kaitiakitanga[4]. La premisa que subyace a estas licencias es que los creadores de una lengua y, por tanto, las fuentes de datos, deben disfrutar de los beneficios generados por su lengua, y ser reconocidos como socios de datos esenciales para la existencia del modelo de lenguaje global. Estos reconocimientos pueden no ser siempre monetarios; podrían consistir en un intercambio de servicios o en garantizar que los datos se utilizan para crear herramientas capaces de aportar mejoras a la comunidad. En Lelalpa AI, hemos publicado recientemente un juego de datos con una licencia comercial personalizada, que especifica que las entidades comerciales situadas fuera de África deben pagar por los datos, mientras que las entidades africanas no. El principio es que cualquier beneficio derivado de los datos debe reinvertirse en la economía africana o devolverse a la comunidad contribuidora. En el caso de las entidades africanas, la recaudación se destina a crear más datos para la comunidad, ofreciendo oportunidades de trabajo remuneradas para la creación de datos. Este modelo de creación de datos, beneficio y soberanía dirigidos por la comunidad contrasta con el enfoque extractivista que domina la industria de los modelos de lenguaje, donde se considera que los datos lingüísticos (por ejemplo, el contenido de Internet) no pertenecen a "nadie" y, por tanto, "a todo el mundo". En realidad, sólo las grandes empresas tecnológicas disponen de recursos para capitalizar estos datos, muchas veces sin compensar ni reconocer a los creadores de los datos, su lengua, su pueblo o su patrimonio.

Lo que también han hecho los requisitos del contexto de bajo recurso es influir en la cultura de la "organización" en el ámbito de la tecnología. Las preferencias culturales relativas a la soberanía de los datos y las necesidades localizadas dependen de las estructuras organizativas que definen la forma en que las personas construyen y comparten las cosas. En el ámbito de la investigación, organizaciones de base como Masakahane han sentado precedentes sobre el reconocimiento y la apropiación local de la investigación y su desarrollo. Masakahane es una iniciativa de investigación y una comunidad panafricana dedicada a impulsar el procesamiento del lenguaje natural (PLN) para las lenguas africanas, fundada por Jade Abbott en 2018. Masakhane se ha ganado el reconocimiento por su enfoque innovador a la hora de abordar las desigualdades mundiales en la IA y se ha asociado con instituciones académicas, empresas tecnológicas y otras organizaciones para expandir su misión. En particular, ha sentado un precedente de trabajos de investigación inclusivos que reconocen a todos los que contribuyen a ellos por pequeña que sea su aportación. Este precedente de reconocimiento pone de relieve la naturaleza colectiva por la que se logra el avance tecnológico en el contexto de bajo recurso, contextos que están arraigados en filosofías e ideologías colectivistas y centradas en la comunidad. Son famosos por escribir artículos de investigación en los que figuran más de 50 autores, en reconocimiento a todos los que hicieron posible la investigación. Aunque haya sido objeto de críticas por parte de la comunidad investigadora por ello, un reciente artículo de Meta ha seguido este precedente.

En el espacio comercial, organizaciones como Huniki, con una ideología parecida en beneficio de la comunidad colectivista, es una organización que pretende desafiar la visión centrada en el monopolio en torno a la tecnología del lenguaje. Se describe como una federación que pretende garantizar que la PLN africana no esté monopolizada, sino que los proveedores locales de PLN del continente africano reciban ayuda para mantener la integridad y el beneficio local del ecosistema africano de PLN.

Lo que nos enseña el desarrollo de modelos en un contexto de bajo recurso es que 1. Hay muchas formas de aprovechar la inteligencia del lenguaje en las máquinas 2. Los recursos limitados no tienen por qué ser la barrera de entrada para construir una tecnología del lenguaje útil y, a la inversa, no es necesario que todos los modelos del lenguaje requieran

3 Una fórmula del Data Science Law Lab: https://datasciencelawlab.africa/nwulite-obodo-open-data-license/

4 Una fórmula de Te Hiku Media: https://tehiku.nz/te-hiku-tech/te-hiku-dev-korero/25141/data-sovereignty-and-the-kaitia-kitanga-license

recursos tan exorbitantes para ser funcionales 3. Los sectores dedicados a los modelos de lenguaje no tienen por qué ser propiedad de monopolios, sino que pueden fomentarse mediante la propiedad y el beneficio colectivos. Fundamentalmente, esto pone de relieve la importancia de modelos más pequeños e inteligentes, desarrollados por un gran ecosistema comunitario de actores independientes.

Parte 4: ¿Cómo sería una sociedad basada en la IA y desarrollada con un enfoque que respete y acepte la diversidad lingüística?

Esto nos lleva a preguntarnos cómo será el futuro de la IA cuando el contexto de bajo recurso tenga la posibilidad de prosperar y contribuir al desarrollo de este campo. Voy a dividir esto en una visión a corto plazo y otra a largo plazo.

Apoyo infraestructural a la prestación de servicios públicos a corto plazo

En nuestro mundo cada vez más conectado, hay dos factores fundamentales que determinan el acceso a los productos y servicios esenciales. La primera es quizá la más fundamental: el acceso a la lengua es la puerta de entrada a la información imprescindible para la vida. La aparición de modelos de lenguaje más pequeños y eficaces está democratizando el acceso a los servicios y sistemas públicos de una forma inédita.

Los modelos de lenguaje hacen las veces de puentes que salvan la brecha digital, permitiendo a las personas efectuar las tareas esenciales de la vida moderna: comprender las obligaciones fiscales, descifrar las facturas de los servicios públicos, comunicarse con los proveedores de asistencia sanitaria y utilizar el transporte público. Al reducir drásticamente los costes de traducción y comunicación, estas tecnologías son especialmente portadoras de transformación para regiones como África, donde el pasado colonial impuso muchas veces lenguas oficiales únicas —generalmente indoeuropeas— que muchos ciudadanos no hablan con fluidez ni entienden. Esta revolución tecnológica propicia la diversificación de los medios de formación, eliminando la necesidad de aprender una nueva lengua simplemente para acceder a información vital.

Lo que crea la disponibilidad de la lengua a través de productos digitales a menor coste es infraestructura. Una compresión del espacio tan grande que hace que los caminos de acceso sean mucho más cortos. Pone los servicios esenciales directamente en manos de quienes los necesitan, en lugar de obligarles a desplazarse hasta los lugares donde esos servicios están “disponibles”. Lo ideal sería que las clínicas, los hospitales, las oficinas gubernamentales y las instituciones financieras estuvieran situados cerca de todos los que los necesitan, con un transporte asequible y fiable para facilitar el acceso. Sin embargo, para la mayor parte del mundo, estos proyectos de infraestructuras requieren inversiones de miles de millones, de los que actualmente no se dispone, y su puesta en marcha podría llevar décadas. Esperar a disponer de una infraestructura a tan gran escala no es una opción viable, sobre todo cuando ya existen soluciones provisionales rentables.

A corto plazo, la IA funciona como un sistema de apoyo fundacional, mejorando la infraestructura de la gobernanza civil y la prestación de servicios, tanto al ampliar la capacidad de prestación de servicios como al garantizar su consumo a través de la lengua.

Acceso democratizado a los servicios digitales a largo plazo

A corto plazo, la IA puede utilizarse para generalizar reglas entre grupos de personas, con el fin de superar las carencias de recursos y respaldar la prestación de servicios. A largo plazo, la IA inspirada en estos enfoques alternativos puede permitir el resurgimiento del individuo a partir de los puntos de datos, ofreciendo experiencias y una atención más personalizadas. Como los modelos requieren menos recursos, se pueden asignar más recursos a otros fines, haciendo que la IA pase de ser una fuente centralizada de actuación conjunta a una herramienta que los individuos pueden utilizar para satisfacer necesidades específicas. Además de facilitar el acceso a los productos y servicios esenciales, la IA tiene el potencial de aumentar nuestras experiencias de formas que aún no hemos llegado a imaginar.

A medida que la IA se vuelve más barata, rápida y compacta, la personalización está lista para afectar a casi todos los aspectos de la vida cotidiana. Los avances en técnicas como el aprendizaje federado y los sistemas descentralizados controlados por el cliente (DCCS) permiten que los datos permanezcan en manos de los usuarios, fomentando la soberanía de los datos al tiempo que posibilitan aplicaciones de IA que antes se consideraban demasiado arriesgadas o que requerían demasiados recursos.

Por ejemplo, en lugar de limitarse a facilitar la comunicación entre un paciente y un profesional sanitario, la IA podría llevar a cabo escáneres corporales completos e integrar los antecedentes médicos familiares para ayudar a los profesionales a emitir diagnósticos precisos y tratamientos personalizados, todo ello sin que los datos confidenciales salgan del dispositivo del paciente. Del mismo modo, en lugar de limitarse a traducir los materiales de clase, la IA podría actuar como un tutor personalizado, ayudando a los profesores a identificar las lagunas de aprendizaje y ofreciendo estrategias personalizadas para mejorar tanto la eficacia de la enseñanza como los resultados del aprendizaje de los alumnos. Estos avances ponen de relieve el potencial transformador de la IA localizada y personalizada en distintos campos.
Actualmente, la IA personalizada se ve obstaculizada por las limitaciones informáticas y los ritmos de aprendizaje incluso de los modelos más avanzados. Pese a estos retos, la IA tiene el potencial de remodelar fundamentalmente nuestra comprensión de la humanidad. Al multiplicar nuestras capacidades, no sólo mejoramos la tecnología, sino que también obtenemos conocimientos más profundos sobre la naturaleza humana. En el futuro, la IA podría desempeñar un papel fundamental a la hora de paliar la escasez de recursos, respaldar a las personas en la gestión de las funciones esenciales y permitir que se centren más en tareas claramente humanas, como la creatividad, la empatía y la conexión. La visión dominante de la IA personalizada sigue centrada en aumentar la productividad de los trabajadores y potenciar los beneficios de los empleadores, lo que refleja un marco arraigado en los ideales occidentales basados en el avance tecnológico. Esta perspectiva limitada pasa por alto otras posibilidades más amplias, como la de que la IA promueva la equidad, fomente la creatividad e impulse la innovación más allá de los paradigmas económicos tradicionales. Para desarrollar todo su potencial, el desarrollo y la aplicación de la IA exigen un enfoque más imaginativo, inclusivo y con visión de futuro.

El ecosistema tecnológico tiene el potencial de evolucionar hacia una red descentralizada de proveedores de servicios interconectados, en la que se distribuya la responsabilidad del bienestar colectivo del planeta. En esta visión, los actores más pequeños colaboran de manera fluida con las entidades más grandes para mejorar la funcionalidad, y el progreso da prioridad a la realización colectiva frente a los beneficios. La eficiencia de los recursos podría ocupar un lugar central en el desarrollo de la IA, abordando preocupaciones cruciales como el cambio climático y el impacto de las cadenas de valor de los minerales preciosos que alimentan los conflictos civiles.

Imaginando una historia alternativa, se podría tener en cuenta cómo podría haberse desarrollado la IA si su financiación inicial no hubiera estado dominada por el sector de defensa

estadounidense. ¿Y si, en lugar de estar arraigada en ideales capitalistas e individualistas, la IA se hubiera fundado sobre filosofías de interconexión, como Ubuntu, una cosmovisión sudafricana articulada por Sabelo Mhlambi como “yo soy porque tú eres”? (en su documento de referencia *From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance).*

Los sistemas de conocimiento indígenas, desarrollados a lo largo de milenios, hacen hincapié en la relacionalidad de la humanidad con el mundo natural y todos sus elementos. Esta perspectiva podría inspirar un paradigma de IA centrado en la sostenibilidad, la equidad y el progreso humano compartido. Ahí reside la oportunidad más bella de desarrollar tecnología desde perspectivas culturales distintas, que hay principios fundacionales alternativos sobre los que se puede construir la IA. Y si ese conocimiento no está explícito, registrado en textos antiguos, esos secretos se guardan con toda seguridad en la expresión humana de la lengua.

Conclusión

La inclusión de la lengua en la tecnología puede moldear profundamente las sociedades, ya sea capacitando a las comunidades o perpetuando las desigualdades. Garantizar una representación adecuada y el apoyo a las lenguas locales en las herramientas digitales no es sólo una cuestión de equidad, sino de salvaguardar el derecho a las representaciones imaginarias futuras. Este derecho implica la capacidad de todas las personas para acceder a la tecnología en la lengua que elijan, capacitándolas para definir sus propias narrativas, aspiraciones e identidades. Permite a las comunidades aprovechar la tecnología para construir futuros que reflejen sus valores y atiendan sus necesidades específicas. Por el contrario, obligar a los individuos a asimilarse a los marcos culturales o lingüísticos dominantes para acceder a las oportunidades socava esta actuación. Las consecuencias a largo plazo de negar la inclusión lingüística, sobre todo en la IA, siguen siendo inciertas. Sin embargo, si no se proporciona dicho acceso, se corre el riesgo de exacerbar las desigualdades digitales, marginar las identidades culturales y limitar los beneficios sociales de las tecnologías emergentes.

Al aprovechar las oportunidades únicas que presentan los contextos lingüísticos africanos, podemos crear un camino que enriquezca la innovación tecnológica, preserve el patrimonio cultural y fomente la equidad en la representación de la lengua. En última instancia, este enfoque ofrece un futuro más sostenible y universalmente beneficioso para la tecnología, que se ajusta a los principios de inclusión y crecimiento compartido.

7. Decisiones algorítmicas y derechos vulnerados en la *gig economy*

Paola Villareal
Investigadora independiente
Ervin Félix
Oxfam México

Cuando empezaron a emerger los primeros espacios de lo que ahora es la economía de plataforma, a menudo, aparecían como una manifestación de la economía colaborativa que conectaba a individuos para intercambiar servicios y recursos. Sin embargo, el futuro que auguraba esta fórmula llevo a que con el paso del tiempo las grandes corporaciones se apropiasen y lo convirtiesen en un negocio vertical. El cambio ha perjudicado, especialmente, a las personas trabajadoras, ya que las empresas con más fuerza en el mercado se han decantado por desarrollar herramientas digitales que sostienen sus modelos de negocios y priorizar sus beneficios. La otra promesa, la de las nuevas oportunidades laborales y la flexibilidad en el empleo ha sido, igualmente la víctima del modelo pensado para maximizar beneficios, en ocasiones a costa del marco de derechos laborales conquistados.

La gestión algorítmica que se ha generalizado en las plataformas de la *gig economy,* empuja una renegociación, cuando no pone en grave riesgo algunos derechos fundamentales de las personas trabajadoras. Desde el derecho a la no discriminación hasta el derecho a un salario digno. Las herramientas digitales y los algoritmos se utilizan para distribuir y gestionar el reparto de trabajo, pero a menudo lo hacen a partir de criterios opacos que no permiten a las y los trabajadores defenderse convenientemente, pero además se emplean para vigilar. En ocasiones, además, esos algoritmos están cargados de sesgos que trasladan ejes de discriminación al espacio laboral, lo que agrava su impacto.

Introducción

En nuestros bolsillos llevamos supercomputadoras que son millones de veces más rápidas y capaces que las que llevaron a la humanidad a la Luna. Estas supercomputadoras contienen sensores extremadamente sensibles que permiten conocer información del dispositivo como la posición y orientación, su aceleración o movimiento, la iluminación ambiental que lo rodea, la cercanía de algún objeto y, por supuesto, aspectos más delicados como la posición geográfica, la huella digital y otros datos biométricos de las personas usuarias. Estos datos están disponibles para todas las aplicaciones instaladas en los dispositivos móviles mediante la autorización del usuario y, una vez que esta se otorga, se utilizan para alimentar complejos algoritmos y modelos de clasificación y priorización que son la columna vertebral de muchas aplicaciones, como las que rigen la *gig economy*.

La *gig economy* es un modelo laboral basado en trabajos temporales o por demanda, facilitados a través de plataformas digitales. Al inicio, la *gig economy* fue promovida como una economía colaborativa que conectaba a individuos para intercambiar servicios y recursos; sin embargo, con el tiempo se convirtió en un negocio vertical en el que las grandes corporaciones centralizaban el control, la toma de decisiones y los beneficios (Madariaga et al., 2018). Eso dejó de lado a las personas trabajadoras, pues las empresas principales desarrollaron herramientas estadísticas e informáticas para sustentar sus modelos de negocios y priorizar sus ganancias (A. Zhang et al., 2022).

Dichos algoritmos y modelos forman parte de la gestión algorítmica (Lee et al., 2015) que caracteriza a las plataformas de la *gig economy*, y son utilizados para coordinar, vigilar e instruir a las personas trabajadoras de la manera en que mejor le convenga a las plataformas. En otras palabras, toda la relación entre trabajadores, clientes y plataformas está mediada casi sin que exista intervención humana. En teoría, tal modelo de gestión permite brindar un mejor servicio a los consumidores, además de facilitar a las empresas su expansión a otros territorios, casi sin necesidad de abrir oficinas o tener equipos de ingeniería, ventas o gestión locales. Esto ha resultado, en muchos casos, en enormes disrupciones para las economías y trabajadores locales.

En términos técnicos, la *gig economy* enfrenta a las personas trabajadoras a la brecha digital: su principal herramienta de trabajo es el dispositivo móvil propiedad del trabajador y el acceso a las plataformas digitales —y, por ende, al empleo mismo— depende de las capacidades técnicas de estos equipos: la sensibilidad del GPS y de su cámara fotográfica o la velocidad de su conexión (Sanghro, 2023). Además, les enfrenta a que distintos algoritmos toman decisiones sobre su labor: desde dar de alta a nuevos trabajadores hasta el pago por los servicios prestados y el cobro de comisiones, pasando por mecanismos opacos para vincular a oferentes y demandantes, que pueden perjudicar o beneficiar a ciertas personas trabajadoras sin su conocimiento.

En ese sentido, sus derechos laborales, digitales y humanos también están condicionados por las maneras en que los algoritmos de las plataformas cristalizan condiciones y dinámicas de trabajo (L. Zhang et al., 2023). Entre los derechos con mayores afectaciones debido a la llegada de la *gig economy*, podemos observar el derecho a la privacidad, a contar con condiciones de trabajo seguras, a no ser discriminados en su trabajo y a tener una remuneración justa y transparente.

Detrás de las interfaces de las aplicaciones, trillones de decisiones se toman de forma automática, inmediata y opaca. Así, mientras que las personas trabajadoras resultan perjudicadas en sus derechos laborales por decisiones automatizadas, la gestión algorítmica produce grandes beneficios para sus dueños y accionistas (Jarrahi et al., 2020). De esta forma, el modelo de negocio de las empresas de la *gig economy* está basado en herra-

mientas estadísticas poco transparentes que vulneran los derechos laborales de millones de personas. Como mostraremos en las secciones siguientes, la gestión algorítmica no puede ser una caja negra, sino que debe estar sujeta al escrutinio público y al pago de su justa contribución social, en aras de proteger a las personas trabajadoras y garantizar que estas formas de trabajo no sostengan ni profundicen desigualdades preexistentes.

La gestión algorítmica y la relación con las personas trabajadoras

Como ya se mencionó, la gestión algorítmica tiene implicaciones directas sobre los derechos de millones de personas trabajadoras, pero para establecer qué derechos y en qué momentos son vulnerados es necesario explicar los tipos de decisiones que se toman por medio de algoritmos, qué tipo de modelos están involucrados, con qué datos son entrenados y probados y cuáles son los posibles resultados de dichas decisiones. Es decir, las plataformas de la gig economy no cuentan con un solo modelo maestro que toma todas las decisiones, sino que cuentan con varios modelos que son utilizados en momentos específicos para resolver situaciones particulares y que operan de manera independiente (Duggan et al., 2020).

La relación entre las plataformas y los trabajadores de la gig economy tiene momentos críticos en que los modelos juegan un papel preponderante para automatizar la toma de decisiones considerando un amplio abanico de variables. Pueden mencionarse, entre otros, la apertura de cuentas, la autenticación y comprobación de identidad, la asignación de actividades a realizar, la definición de zonas de alta demanda y de tarifas dinámicas, la distribución de pagos, propinas e impuestos, la medición del desempeño e incentivos y la resolución de eventualidades (Dubal, 2023).

Entre los modelos involucrados podemos encontrar el reconocimiento facial, el seguimiento geográfico, la vinculación entre oferta y demanda, la contabilidad y pagos y la seguridad. Cada uno de ellos es entrenado según su contexto específico utilizando datos relevantes de los usuarios para tomar decisiones. Por ejemplo, el modelo de reconocimiento facial para corroborar la identidad de la persona trabajadora (Sullivan, 2016) o los modelos de balanceo de carga —una técnica informática que distribuye el tráfico de red entre varios servidores para evitar que se sature uno solo— que pueden llegar a desconectar a trabajadores por horas —o, en casos extremos, días enteros— si la carga de trabajo requiere menos trabajadores que los que están disponibles en determinados momentos (Zipperer et al., 2022).

Apertura de cuenta

Al abrir su cuenta, la persona trabajadora es sometida a controles para la verificación de su identidad. Estos son arbitrarios y poco claros, debido a que no se conocen con exactitud los datos recabados, los procedimientos internos que se realizan para su verificación, el control de acceso a dichos datos, el tiempo en el que son retenidos, las personas con quienes se comparten o los procedimientos para su eliminación. Esta falta de transparencia, combinada con la vigilancia constante de las personas trabajadoras, trae consigo diversas preocupaciones y cuestionamientos sobre el respeto de la privacidad de los involucrados, incluyendo si la recolección y el monitoreo son realmente apropiados, si los datos de las personas son utilizados de manera correcta y cómo esta información se utiliza para tomar decisiones (Sannon et al., 2022).

Inicio de sesión

Una vez que la persona trabajadora pasa los controles de la empresa y es aceptada para brindar sus servicios, tendrá que iniciar sesión diariamente en la plataforma para comenzar a recibir tareas. Esto le enfrenta a los algoritmos de inicio y de manejo de sesión que tienen como misión, además de la autenticación de los trabajadores, mantener el equilibrio entre la demanda de servicios y la oferta de los trabajadores. En ese sentido, los algoritmos de manejo de sesión son utilizados para mantener en línea solamente a los proveedores de servicio necesarios para poder cumplir con la demanda de los usuarios finales y, en caso de haber exceso de oferta, pueden desconectar temporal o permanentemente, sin previo aviso y por cualquier motivo, a los trabajadores (Safak & Farrar, 2021).

Lo anterior provoca inestabilidad laboral, debido a la impredecibilidad en el acceso a la plataforma y la imposibilidad de trabajar en ella. Asimismo, además del nombre de usuario y contraseña, las plataformas utilizan datos biométricos de los trabajadores al momento de iniciar sesión, con la finalidad de compararlos con los datos capturados al momento de darse de alta. Aunque puede parecer un proceso simple, en realidad termina siendo el motivo por el cual se niega el ingreso a trabajadores debido a errores de captura o cotejo de información. Este es un sistema con muchas piezas que pueden fallar y que depende principalmente de la capacidad del dispositivo móvil y de la conexión del trabajador a internet. Si alguno de estos componentes falla, la persona trabajadora paga las consecuencias al negársele el acceso a ejercer su labor (Zipperer et al., 2022).

El uso de estos sistemas de autenticación también tiene como objetivo comprobar que el trabajador es quien dice ser y, para esto, se le obliga a tomarse fotografías y registrar las coordenadas de su dispositivo móvil para cotejarlas tanto con bases de datos internas como externas. Este proceso es sumamente delicado pues, en muchos casos, los repartidores no cuentan con dispositivos suficientemente poderosos como para capturar de manera precisa tanto las fotografías como las coordenadas, lo que hace que la aplicación los bloquee o, incluso, los dé de baja de manera permanente (Aguirre Reveles, 2020).

Clasificación y calificación

Una vez que la persona trabajadora ha logrado ingresar a la plataforma y es candidata para tomar la siguiente orden de trabajo, entra en contacto con otros algoritmos que lo clasifican y califican para determinar si es elegida o no para realizar dicha orden. Para tomar la decisión sobre qué trabajador realizará la nueva tarea, los modelos toman en cuenta diversos factores, desde la distancia entre el trabajador y el punto de destino, el tipo de vehículo, la velocidad media o el tráfico en la zona, hasta el comportamiento previo del trabajador, el tiempo que tarda en aceptar las nuevas tareas o cuántas nuevas tareas acepta. Esos algoritmos determinan las órdenes que tomará la persona trabajadora durante el día y así sus posibilidades de obtener ingresos (L. Zhang et al., 2023).

Contrario a la noción de que estos modelos y algoritmos permiten la libertad laboral de las personas trabajadoras, en realidad estas capacidades convierten a los modelos y algoritmos involucrados en capataces modernos, encargados de premiar o castigar algorítmicamente, sin mediar explicación y sin comunicación directa y asertiva, a las personas trabajadoras. Así, se establece una especie de *shadow banning* (bloqueo en la sombra, en español), en que la persona trabajadora puede notar que recibe menor cantidad de tareas y de recursos, sin conocer el porqué. Adicionalmente, cuando el trabajador es monitoreado constantemente a través de algoritmos, aun siendo estos disfrazados de retos divertidos (*gamification* o ludificación en español), lo anterior tiene efectos negativos sobre la estabilidad emocional de las personas trabajadoras (Kadolkar et al., 2024; A. Zhang et al., 2022; L. Zhang et al., 2023).

Determinación de precios y tarifas

Una de las principales preguntas acerca del funcionamiento de las plataformas de la gig economy es qué métodos utilizan para definir tarifas y distribuir recursos. Las tarifas base de los servicios no solo tienen que ver con el servicio que provee la persona trabajadora, sino también con el perfil del consumidor final, el negocio que ofrece los productos, las zonas en las que se encuentran los involucrados, el historial de transacciones, el historial de quejas, fraudes y violencia, entre otras variables. La lista exacta de factores y su importancia en la definición del precio final es un secreto muy bien guardado por las plataformas y cada una, seguramente, tiene su propia fórmula. Además, es poco claro qué proporción de lo que pagan los consumidores terminará en el bolsillo del o la repartidora, pues las plataformas cobran comisiones y cargos extra y retienen impuestos antes de asignar los pagos.

Como consecuencia, las personas trabajadoras no tienen certeza sobre cuánto dinero obtendrán por cada tarea realizada, lo cual resulta en la vulneración del derecho de los trabajadores a una remuneración digna. Adicionalmente, existen sesgos de los modelos de aprendizaje de computadora y de inteligencia artificial que asignan precios más altos a zonas donde habitan poblaciones no hegemónicas —personas de color, pobres, jóvenes, entre otras (Pandey & Caliskan, 2021), un criterio que probablemente también aplica para las personas trabajadoras. Dicho de otro modo, no todas las personas trabajadoras ni usuarias son iguales ante los algoritmos y modelos.

Por otra parte, es importante recordar que es muy común que las plataformas definan tarifas dinámicas con base en la demanda de los usuarios finales. Es decir, cuanto mayor sea la demanda de servicios y, por ende, se encuentren disponibles menos trabajadores, se incrementará el costo del servicio. La tarifa dinámica se define como un multiplicador que se aplica a la tarifa base. Por ejemplo, si la tarifa base es de $100 y no hay tarifa dinámica, el multiplicador será 1, dejando el costo final en $100. En cambio, si hubiera un aumento del 20 % en la demanda de servicios y nuestro algoritmo definiera al multiplicador de forma lineal, este sería de 1,2 que, multiplicado por la tarifa base, daría $120. Este ejemplo es muy sencillo y la definición de la tarifa dinámica de cada plataforma, seguramente, toma en cuenta muchos otros factores, pero sirve para ilustrar la falta de transparencia en las decisiones.

Calificación del desempeño

Los algoritmos no solo gestionan transacciones específicas, sino que también procesan información agregada para definir métricas que califican el desempeño de las personas trabajadoras, lo que determina los incentivos y castigos que impactan su ingreso y su elegibilidad para tareas futuras. Las personas trabajadoras se enfrentan todos los días a dichas métricas, que pueden ir desde el porcentaje de peticiones de trabajo rechazadas hasta el tiempo promedio por tarea, y han tenido que aprender a adaptarse en un esfuerzo por resistir y manipular a los modelos y algoritmos a su favor. Esta resistencia a ser constantemente vigilados y medidos es análoga a aprender a lidiar con un nuevo oficio, jefe o lugar de trabajo y suele ser causa de incertidumbre que puede llevar, incluso, a que se den de baja de las plataformas (A. Zhang et al., 2022).

En general, el desarrollo de las plataformas digitales recae en la implementación de modelos genéricos pre-entrenados con enormes bases de datos, listos para ser reutilizados en casi cualquier territorio o situación. Mientras que esto permite el crecimiento de las plataformas, también trae consigo la proliferación de sesgos y puntos ciegos que reducen su representatividad estadística y que terminan perjudicando los derechos de las personas trabajadoras. Por ejemplo, si un modelo es entrenado con datos de trabajadores de India, este definirá métricas y expectativas con base en esos datos, las cuales no necesariamente van a coincidir con el comportamiento de los trabajadores en México. Para prevenir este

tipo de sesgos y puntos ciegos, es necesario que las plataformas entrenen sus modelos con datos locales que privilegien la aplicabilidad de las decisiones tomadas por sus sistemas.

En el trabajo en la gig economy, como en cualquier otra actividad económica, puede ocurrir alguna eventualidad —fallas mecánicas, accidentes de tránsito, eventos delincuenciales, fraudes, entre otros— frente a las cuales las personas trabajadoras se encuentran en el desamparo, pues no cuentan con seguros que les cubran más allá del tiempo en el que están activos en las plataformas. De hecho, salvo en momentos muy específicos, las coberturas se limitan a cubrir daños a terceros, mientras que la persona trabajadora deberá cubrir personalmente los daños en su vehículo aun estando en línea en la plataforma (Uber, s/f). Esto tiene como consecuencia que las personas trabajadoras pierdan acceso tanto a su herramienta de trabajo como la parte de su ingreso que deberán destinar para las reparaciones necesarias. En el caso de ser víctimas de delitos, es poco claro si las plataformas protegen a las personas trabajadoras y, al igual que con los seguros, varía de jurisdicción en jurisdicción.

En este sentido, en lugar de representar un avance que garantice condiciones justas, la gestión algorítmica ha demostrado ser una herramienta que sostiene y profundiza la precarización laboral de las personas trabajadoras. Los dueños de las plataformas y sus inversionistas obtienen beneficios significativos al optimizar costos y maximizar ganancias, pero esto se consigue a expensas de la estabilidad laboral, la seguridad económica y la dignidad de quienes sostienen las operaciones diarias. La implementación opaca y unilateral de los sistemas intensifica la dependencia de las personas trabajadoras hacia las plataformas, además de que las expone a un entorno de trabajo marcado por la incertidumbre, la vigilancia constante y la ausencia de derechos laborales fundamentales. La gestión algorítmica se traduce en la vulneración sistemática de los derechos laborales, puesto que reproducen dinámicas de explotación en un mercado digital aparentemente moderno, pero profundamente desigual.

La gestión algorítmica y sus implicaciones para los derechos laborales

La gig economy se caracteriza por ofrecer empleo en una modalidad discontinua o a corto plazo, donde las personas trabajadoras utilizan la plataforma para buscar distintas oportunidades en un territorio determinado, pero sin el reconocimiento de la relación de subordinación. Así, esta forma de trabajo ha modificado profundamente los mercados laborales: si bien ofrece flexibilidad y autonomía, también implica vulneraciones de los derechos laborales.

De esta manera, el uso generalizado de algoritmos para la gestión y la supervisión de las personas trabajadoras provoca que las condiciones de trabajo se precaricen. En particular, se identifican al menos cinco derechos fundamentales vulnerados: el derecho al trabajo, a la privacidad, a una remuneración justa, a condiciones de trabajo seguras y a la no discriminación. Además, es importante identificar los mecanismos de control, opacidad y extracción de datos que sostienen y perpetúan las desigualdades en este sector.

Derecho al trabajo

Los algoritmos de la gig economy desempeñan un papel determinante en la asignación del trabajo en este sector, con base a los diversos factores mencionados. No obstante, esta toma de decisiones automatizada es sumamente opaca, lo cual puede conducir a la discriminación algorítmica en la que los trabajadores pueden percibir menos asignaciones

de trabajo debido a métricas arbitrarias o a su ubicación geográfica, en lugar de sus capacidades y posibilidades (Widjaya, 2024).

Así, el derecho al trabajo es violentado por medio del control algorítmico. Las personas trabajadoras de las plataformas digitales aseguran que hay un control algorítmico que da la sensación de estar atrapado en un sistema rígido, que no tiene en cuenta las circunstancias o preferencias individuales (Lang et al., 2023; L. Zhang et al., 2023).

De esta manera, hay una paradoja en la autonomía de las plataformas digitales (Jaharri et al, 2019). Por un lado, en otros sectores laborales hay horarios estrictos en los que los trabajadores deben cumplir con cierto número de horas y trabajar en un entorno establecido, mientras que en un entorno digital controlado por algoritmos se cuenta con la flexibilidad y autonomía para decidir en dónde y cuándo trabajar. Por el otro, un sistema donde hay seguimiento y supervisión continuos a través de sistemas algorítmicos puede desencadenar tensión y ansiedad entre trabajadores, haciéndoles sentir menos autónomos (Lee, 2018).

La gestión algorítmica da lugar a horarios de trabajo irregulares, exceso de trabajo y aislamiento social para las personas trabajadoras (Shapiro, 2018), en un ataque al derecho al trabajo a través de la inestabilidad laboral, tal como ilustra el siguiente testimonio de una repartidora de plataformas digitales en México:

> *"Si estás ahí [en las plataformas digitales], como quiera ya no tienes como que chance de hacer algo más. O, por decir, yo llego cansada y luego, si me quiero meter a hacer mis trabajos de la prepa, pues ya me cuesta trabajo, o ya estoy cansada, me da mucho sueño y me paro muy temprano, ya así como que ya no le pongo como que el mismo [...] interés ni el mismo esfuerzo que le podía poner cuando trabajaba en la oficina, porque ahí tenía tiempo; ahí ya la hacía y no había problema, pero aquí, pues no"*
>
> – Testimonio de repartidora, mujer joven, en México.

Del mismo modo, al priorizar la eficiencia algorítmica y la maximización de ganancias, se ignora la dimensión de los cuidados, que es esencial para muchas de las personas que trabajan en este sector, especialmente para las mujeres. Las jornadas extensas e irregulares restringen la posibilidad de que las personas trabajadoras tengan la posibilidad de cuidar a otras personas con necesidades de cuidados. Esta realidad coloca a las mujeres, principalmente, en una posición de vulnerabilidad no solo respecto a sus derechos laborales, sino también en términos de la autonomía para equilibrar su vida laboral con la personal (Centeno Maya et al., 2022).

Derecho a la privacidad

El derecho a la privacidad se encuentra reconocido en diversos instrumentos internacionales como en el artículo 12 de la Declaración de los Derechos Humanos y el artículo 17 del Pacto Internacional de Derechos Civiles y Políticos. Protege a las personas ante injerencias arbitrarias en su vida privada y garantiza la protección de sus datos personales. No obstante, en la gig economy este derecho se ve vulnerado en reiteradas ocasiones a partir de la recopilación masiva de datos, sobre todo de los datos personales, lo que representa una de las mayores preocupaciones de las personas trabajadoras en este sector (Sannon et al., 2022).

El discurso de las plataformas digitales es promover y enaltecer la libertad del trabajador. Sin embargo, esta libertad se sostiene sobre el control algorítmico que ejercen y la explotación de los datos personales de personas trabajadoras y usuarias, bajo una vigilancia que

evoca a una distopía orwelliana. Además, las plataformas digitales mantienen la opacidad en la manera en la que funcionan estos algoritmos, por lo que las personas trabajadoras y usuarias no conocen ni comprenden para qué o cómo es manejada su información personal, si está siendo compartida con terceros o si está siendo lo suficientemente protegida (Tsaaro Consulting, 2023).

La filtración de la información de las personas trabajadoras es otra vulneración al derecho a la privacidad. En 2016 se diseminaron de manera ilegal los nombres, direcciones de correo electrónico y números de teléfono celular de más de 25 millones de trabajadores de plataformas digitales a nivel global, además de una base de datos con 600 mil licencias para conducir de diversos conductores y repartidores en Estados Unidos. Estas cuestiones de seguridad son críticas para los trabajadores porque, además de la filtración, la empresa responsable no hizo saber a los repartidores y conductores que este evento había sucedido sino hasta un año después (Khosrowshahi, 2017).

Derecho a una remuneración justa

El derecho a una remuneración justa se encuentra contenido en el artículo 23 de la Declaración Universal de los Derechos Humanos, donde se establece que toda persona tiene derecho a recibir un salario equitativo que le permita garantizar un nivel de vida digno para sí misma y su familia. En el sector de la gig economy, este derecho se enfrenta a diversos desafíos debido a la gestión algorítmica que define tarifas, incentivos y comisiones para los trabajadores de manera opaca e incierta (Dubal, 2023).

La gestión algorítmica desempeña un papel fundamental a la hora de determinar la remuneración de los trabajadores (Abraham, 2023). Los algoritmos establecen las tarifas, comisiones y bonificaciones en función de diversos factores, como la demanda, la complejidad de la tarea y el perfil de la persona trabajadora (Sharma, 2024). En este sentido, los algoritmos utilizan precios dinámicos que ajustan las tarifas dependiendo de la fluctuación de la oferta y la demanda, por lo que el sistema puede dar lugar a variaciones significativas en el ingreso de las personas trabajadores.

Este sistema nos ha hecho creer que las personas trabajadoras obtienen una ganancia muy grande si trabajan durante muchas horas. No obstante, se ha encontrado que esto solo es una falsa creencia, ya que las plataformas condicionan a los trabajadores a operar en horarios y situaciones específicas para evitar perder ingresos (Sharma, 2024). Incluso, existen casos en los que quienes trabajan más horas suelen ganar menos por hora (Cook et al., 2021). Esto genera ingresos impredecibles, variables y personalizados, lo cual vulnera el principio de igual salario por igual trabajo, en perjuicio de las personas trabajadoras.

Lo anterior se refleja en un testimonio de una persona repartidora al ser cuestionada sobre si el ingreso que percibe es suficiente:

> *"Hasta cierto punto mis necesidades básicas sí las cubriría, lo que es comida o algo así, pero bueno, a veces sí; por ejemplo, ayer nada más me llegaron dos viajes, entonces de a 20 pesos [US$1], nada más me hice 40 [US$2] en todo el día y desde temprano hasta la noche."*
>
> – Testimonio de repartidora, mujer joven, en México.

Algunas personas repartidoras de plataformas digitales incluso comparan esta experiencia con las ganancias que se obtienen en juegos de azar, pues siempre se encuentran a la espera de un gran premio (*jackpot*), mientras que, en realidad, el algoritmo les sigue dando un número suficiente de viajes para mantenerlos activos aunque los ingresos por viaje sean bajos (Abraham, 2023).

Las personas trabajadoras se enfrentan a un sistema opaco y desigual que perpetúa la inestabilidad de ingresos. El uso de precios dinámicos, la personalización de tarifas y la ludificación del trabajo no solo vulneran el derecho a una remuneración justa, sino que también profundizan las desigualdades económicas y sociales entre trabajadores.

Derecho a condiciones de trabajo seguras

Este derecho se encuentra contenido en distintos instrumentos internacionales, como el Convenio 155 de la Organización Internacional del Trabajo sobre la Seguridad y Salud de los Trabajadores y el artículo 23 de la Declaración Universal de los Derechos Humanos. Garantiza que las personas trabajadoras puedan desarrollar sus actividades en un entorno que proteja su integridad física, mental y social, al minimizar los riesgos laborales y promover su bienestar. En la gig economy, este derecho se encuentra vulnerado, pues la gestión algorítmica prioriza la eficiencia y la rentabilidad sobre la seguridad de las personas trabajadoras (De Stefano & Taes, 2023).

En las plataformas digitales, los algoritmos establecen no solo la carga de trabajo, sino también los tiempos de entrega y las rutas a seguir. Como ya se ha visto, esto pone en situaciones de riesgo a las personas trabajadoras pues, al intentar cumplir con estos tiempos, terminan incumpliendo reglas de tránsito en perjuicio de su seguridad (L. Zhang et al., 2023). La ausencia de supervisión humana en estas decisiones agudiza las condiciones inseguras y obliga a las personas trabajadoras a enfrentar distintos dilemas que no deberían serlo, como mantener su acceso a oportunidades laborales en la plataforma o priorizar su bienestar (Abraham, 2023).

Asimismo, las personas repartidoras se enfrentan a posibles agresiones por parte de clientes. El testimonio de una persona repartidora da cuenta de lo anterior:

> *"[Al momento de hacer la entrega] antes de concluir los 10 minutos, como tres segundos o un segundo antes, llegó un coche y empezó a gritarme que por qué no le quería llevar su pizza, que le diera la pizza y me la empezó a arrebatar, y le dije: 'Toma tu pizza, yo no quiero problemas'. Y me dice: "Vas a ver, te voy a encontrar, voy a hacer que te suspendan tu cuenta, te voy a perjudicar de la mejor forma que pueda". Y yo no tenía mucho que me había registrado, entonces sí me quedé como yo qué hice mal, al contrario, yo estoy siguiendo las indicaciones que me proporcionó mi soporte".*
>
> – Testimonio de repartidora, mujer joven con hijos, en México.

En este sentido, el objetivo de las plataformas digitales es maximizar el número de tareas completadas por las personas trabajadoras, donde la gestión algorítmica aparece como una intermediaria cuya función principal es garantizar que todas y cada una de las tareas enviadas por los solicitantes se realicen a tiempo y con la mejor calidad posible (Berástegui, 2021). Así, la noción de sobrecarga de trabajo en la gig economy se sostiene en la gestión algorítmica y la vigilancia digital, cuyo principal objetivo es coordinar y maximizar la carga de trabajo en respuesta a los distintos factores que existan en tiempo real, un proceso de optimización que se ha identificado como fuente de tales sobrecargas (Poutanen et al., 2021).

Otro factor relevante es la ausencia de gestión de recursos humanos por parte de las plataformas digitales. Diversos estudios (Bellesia et al., 2019; Deng et al., 2016) muestran la falta de capacitación, la negligencia en el manejo de quejas y la inadecuada comunicación con los trabajadores de las plataformas, por lo que no solo existe el peligro de accidentes y presión por parte del algoritmo, sino que la atención a los trabajadores es inadecuada.

Esta falta de apoyo agrava la precarización de las condiciones laborales y contribuye a la vulneración de derechos laborales.

El derecho a condiciones de trabajo seguras en la gig economy se ve transgredido por la gestión algorítmica debido a la falta de supervisión humana y a la irresponsabilidad de las plataformas frente a las personas que emplea. La gestión algorítmica perpetúa un entorno laboral donde los riesgos son trasladados al individuo, desdibujando la responsabilidad de las plataformas y dejando a las personas trabajadoras en una constante lucha por equilibrar su seguridad y la necesidad de ingresos.

Derecho a la no discriminación

En el ámbito laboral, el derecho a la no discriminación se encuentra protegido por diferentes instrumentos internacionales como el Pacto Internacional de Derechos Económicos, Sociales y Culturales y la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (CEDAW). Estos marcos normativos protegen a las personas de la discriminación por motivos de género, raza y cualquier otra característica, de manera que se promueva la igualdad sustantiva en todas las esferas de la vida.

Como ya se mencionó, la gestión algorítmica busca promover el mayor número de cargas de trabajo para maximizar las ganancias de los dueños y accionistas, por lo que el diseño de estos algoritmos ha precarizado, en mayor medida, a grupos históricamente vulnerados en beneficio de los dueños de la gestión algorítmica. Sumado a esto, los algoritmos promueven la discriminación porque las personas que los diseñan tienen prejuicios y sesgos personales al momento de establecer las métricas que se utilizan para la distribución de labores (Lawyer Monthly, 2024; Skelton, 2021). En este sentido, los prejuicios en la gig economy afectan en una mayor medida a las mujeres y los grupos históricamente vulnerados, a través de distintas oportunidades de empleo desde las plataformas digitales (University of Southampton, 2024).

La relación entre el trabajador y las plataformas digitales en el sector de la gig economy es ambigua, lo que dificulta la identificación de un empleador claro en caso de haber denuncias por discriminación. Por lo tanto, las personas trabajadoras tienen dificultades para presentar denuncias por esta causa o para solicitar reparación legal, ya que es posible que no haya una entidad clara responsable debido a la gestión algorítmica (Lawyer Monthly, 2024).

Uno de los efectos de la discriminación de la gestión algorítmica es que las mujeres reciben menos trabajos bien remunerados que los hombres, debido a las preferencias algorítmicas que favorecen roles históricamente dominados por hombres (Victorian Government, 2020). Adicionalmente, ciertas comunidades pueden ser relegadas en la asignación de trabajo debido a la clasificación de las personas trabajadoras. Este sistema refuerza los sesgos existentes, dotando de más o menos trabajo a perfiles específicos en la gig economy (Botelho et al., 2023). Por ejemplo, se ha encontrado que los sistemas de clasificación reflejan los prejuicios del usuario y los exacerban. Los trabajadores no blancos experimentan un aumento de 80 % en las brechas de calificación en comparación con los trabajadores blancos, lo que impacta negativamente en sus ingresos (Teng et al., 2023).

La falta de claridad en la relación que existe entre las personas trabajadoras de la gig economy y las plataformas digitales complica las demandas por discriminación, dejando a las personas trabajadoras sin protección alguna (Baeza Flores & Márquez Amaro, 2023). Además, el sistema mismo no permite que la asociación entre trabajadoras sea algo fácil de hacer. En este sentido, es crucial incorporar la supervisión humana en las decisiones críticas, como las terminaciones de cuentas y las evaluaciones de desempeño. Estas medidas ayudarían a mitigar los sesgos algorítmicos y a garantizar un trato justo hacia las personas trabajadoras (Gussek et al., 2023; Sharma, 2024).

Finalmente, el derecho a la no discriminación en la gig economy se encuentra profundamente comprometido por la gestión algorítmica, la cual refuerza y amplifica prejuicios históricos en la asignación de tareas, la evaluación del desempeño y la terminación de cuentas. En este contexto, resulta imprescindible la adopción de marcos regulatorios que garanticen la supervisión humana en decisiones clave, así como la promoción de políticas que enfrenten las brechas de género y raza exacerbadas por los algoritmos. Solo así será posible avanzar hacia un entorno laboral más equitativo y justo, eliminando las barreras que perpetúan la discriminación estructural en este sector.

Recomendaciones de política pública

Regulación de la gestión algorítmica y su opacidad

Como se ha establecido en las secciones anteriores, la gestión algorítmica en la gig economy representa riesgos significativos para los derechos de las personas trabajadoras, particularmente en relación con su privacidad y condiciones laborales. Los datos personales son recolectados de forma constante y opaca, utilizándose para clasificar y medir a las personas trabajadoras desde el inicio hasta el final de su relación con la plataforma. Esto no solo perpetúa la discriminación, sino que también genera incertidumbre y precariedad laboral.

El Estado debe de desarrollar un marco regulatorio que obligue a las plataformas digitales a implementar mecanismos de transparencia en sus procesos de toma de decisiones. Por lo que se debe de incluir la apertura del código de los algoritmos, la divulgación comprensible de las variables utilizadas en los entrenamientos de modelos y la obligatoriedad de reportar el impacto de los algoritmos en las condiciones laborales. **En este sentido, las plataformas deben estar obligadas a informar de manera comprensible cómo funcionan los algoritmos y cuál es su impacto en las condiciones laborales; en caso de incumplimiento, es obligación del Estado realizar las acciones necesarias para corregirlo.**

Financiamiento de la seguridad social

El acceso a la seguridad social debe ser garantizado como un derecho universal para las personas trabajadoras de la gig economy. En México, por ejemplo, la reciente propuesta legislativa que contempla una contribución compartida entre plataformas, personas trabajadoras y el gobierno federal representa un avance significativo. Esta protege a las personas trabajadoras y les brinda acceso a la seguridad social sin afectar sus ganancias, ya que pagaría solamente el equivalente a 10 dólares estadounidenses al mes, mientras que obtiene no solo cobertura médica sino también acceso a vivienda o créditos hipotecarios, entre otras prestaciones de la seguridad social mexicana. Este tipo de legislación, que pone al centro a la persona trabajadora, puede llegar a ser un ejemplo importante para otras jurisdicciones ya que, inclusive, cuenta con el apoyo de las empresas detrás de las principales plataformas del país (Marath Bolaños [@marathb], 2024) (#NiUnRepartidorMenos [@repartidorr], 2024).

Para hacer cumplir estos cambios a la ley, **debe establecerse un sistema de sanciones en caso de que las plataformas violen las normativas, lo que incluye multas y la posibilidad de suspensión temporal de sus operaciones**. Además, es fundamental acompañar esta iniciativa con una reforma fiscal progresiva y profunda que asegure que las plataformas digitales paguen los impuestos y contribuciones de seguridad social correspondientes en el territorio nacional. Esto permitirá financiar de manera adecuada y sostenible la seguridad social y evitará que el costo recaiga desproporcionadamente sobre las personas trabajadoras. Con ello, se sentarán las bases para garantizar condiciones laborales dignas.

Promoción de la organización colectiva y redes de apoyo

A pesar de los esfuerzos de las plataformas por desalentar la organización colectiva, resulta indispensable fomentar la creación de sindicatos, cooperativas y otros colectivos que permitan a las personas trabajadoras negociar en mejores condiciones (Mendonça & Kougiannou, 2023). Estas organizaciones no solo fortalecen la capacidad de exigir derechos laborales, sino que también crean comunidades donde se pueden compartir experiencias y generar conocimiento colectivo. En el caso de la seguridad social, estudios internacionales (Zipperer et al. 2022) confirman que las personas trabajadoras en la gig economy enfrentan desventajas importantes en términos de atención médica y planes de retiro. Solo a través de la acción colectiva será posible revertir estas condiciones estructurales (Johnston & Land-Kazlauskas, 2018). El Estado debe garantizar la libertad de asociación, además de proteger a las personas trabajadoras frente a represalias por parte de las plataformas.

Incremento del contacto humano y mecanismos de denuncias efectivos

En las secciones previas se ha mostrado cómo la gestión algorítmica no solo impacta las condiciones laborales, sino también el bienestar emocional de las personas trabajadoras. Es esencial que las plataformas implementen mecanismos de contacto humano, especialmente en situaciones críticas, como accidentes de tránsito o casos de violencia. Asimismo, deben desarrollarse sistemas de denuncia efectivos, manejados por personal capacitado, que permitan a las personas trabajadoras reportar casos de acoso, abuso o inseguridad en el trabajo. Este enfoque contribuirá a mitigar la sensación de desamparo que prevalece en el sector y garantizará respuestas más rápidas y efectivas ante situaciones adversas.

Prohibición de la discriminación salarial algorítmica

La discriminación salarial algorítmica, así como la discriminación por medio de la gestión algorítmica debe ser explícitamente prohibida. Estas prácticas perpetúan desigualdades históricas, al mismo tiempo que generan incertidumbre salarial y agravan los problemas laborales de las personas trabajadoras. Regulaciones claras en esta materia eliminarían la ludificación del trabajo y protegerían a las personas trabajadoras, especialmente a aquellas con ingresos más bajos, de la extracción indebida de datos personales. Estas medidas son esenciales para garantizar el derecho a una remuneración justa y equitativa.

Mientras la gig economy continúa expandiéndose y transformando el panorama laboral, las brechas de poder entre las plataformas digitales y las personas trabajadoras se profundizan. En 2023, a nivel internacional, el sector de la gig economy obtuvo ganancias por más de 200 mil millones de dólares y se estima que para 2032 lleguen a 455 mil millones de dólares (Sharma, 2024). Así, las plataformas digitales siguen aumentando sus márgenes de beneficio a costa de condiciones laborales precarias. Esta disparidad evidencia la urgencia de regular un sector que perpetúa desigualdades estructurales y transgrede los derechos laborales bajo la fachada de innovación tecnológica.

Para que la gig economy pueda cumplir su promesa de flexibilidad y libertad, debe estar acompañada de regulaciones que protejan a quienes sostienen el sistema con su trabajo. Es responsabilidad del Estado, en colaboración con las organizaciones de personas trabajadoras, garantizar un marco normativo que no solo prevenga los abusos, sino que también promueva la justicia social y la igualdad en uno de los sectores más precarizados de los últimos tiempos.

Referencias

Abraham, R. (2023, 25 de enero). Pay Algorithms Make Working in the Gig Economy Feel Like "Gambling," Study Says. *VICE.* https://www.vice.com/en/article/pay-algorithms-make-working-in-the-gig-economy-feel-like-gambling-study-says/

Aguirre Reveles, R. (2020). *Gig economy: Precariedad laboral y traslado de riesgos.* Heinrich Boll Stiftung. https://mx.boell.org/es/2020/11/11/gig-economy-precariedad-laboral-y-traslado-de-riesgos

Baeza Flores, M. E., & Márquez Amaro, R. (2023). Gig Economy, las plataformas digitales para el trabajo: ¿flexibilidad o precariedad laboral? *Ciencia Latina Revista Científica Multidisciplinar, 7*(2). https://doi.org/10.37811/cl_rcm.v7i2.6211

Bellesia, F., Mattarelli, E., Bertolotti, F., & Sobrero, M. (2019). Platforms as entrepreneurial incubators? How online labor markets shape work identity. *Journal of Managerial Psychology, 34(4), 246–268.* https://doi.org/10.1108/JMP-06-2018-0269

Bérastégui, P. (2021). *Exposure to Psychosocial Risk Factor in the Gig Economy: A Systematic Review. ETUI Research Paper - Report 2021.01, Available at SSRN:* https://ssrn.com/abstract=3770016

Botelho, T. L., Sudhir, K., & Teng, F. (2023, agosto 14). Ratings Systems Amplify Racial Bias on Gig-Economy Platforms. *Yale Insights.* https://insights.som.yale.edu/insights/ratings-systems-amplify-racial-bias-on-gig-economy-platforms

Centeno Maya, L. A., Tejada, A. H., Martínez, A. R., Leal-Isla, A. L. R., Jaramillo-Molina, M. E., & Rivera-González, R. C. (2022). Food delivery workers in Mexico City: A gender perspective on the gig economy. *Gender & Development, 30(3), 601–617.* https://doi.org/10.1080/13552074.2022.2131253

Cook, C., Diamond, R., Hall, J. V., List, J. A., & Oyer, P. (2021). The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers. *The Review of Economic Studies, 88(5), 2210–2238.* https://doi.org/10.1093/restud/rdaa081

De Stefano, V., & Taes, S. (2023). Algorithmic management and collective bargaining. *Transfer: European Review of Labour and Research, 29(1), 21–36.* https://doi.org/10.1177/10242589221141055

Deng, X. (Nancy), Joshi, K. D., & Galliers, R. D. (2016). The Duality of Empowerment and Marginalization in Microtask Crowdsourcing: Giving Voice to the Less Powerful Through Value Sensitive Design. *MIS Quarterly, 40(2), 279–302.*

Dubal, V. (2023). *On Algorithmic Wage Discrimination (SSRN Scholarly Paper No. 4331080). Social Science Research Network.* https://doi.org/10.2139/ssrn.4331080

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal, 30(1), 114–132.* https://doi.org/10.1111/1748-8583.12258

Gussek, L., Grabbe, A., & Wiesche, M. (2023). Challenges of IT freelancers on digital labor platforms: A topic model approach. *Electronic Markets, 33(1), 55.* https://doi.org/10.1007/s12525-023-00675-y

Jarrahi, M. H., Sutherland, W., Nelson, S. B., & Sawyer, S. (2020). Platformic Management, Boundary Resources for Gig Work, and Worker Autonomy. *Computer Supported Cooperative Work (CSCW), 29(1), 153–189.* https://doi.org/10.1007/s10606-019-09368-7

Johnston, H., & Land-Kazlauskas, C. (2018). Organizing on-demand representation, voice, and collective bargaining in the gig economy. *ILO Working Papers,* Article 994981993502676. https://ideas.repec.org//p/ilo/ilowps/994981993502676.html

Kadolkar, I., Kepes, S., & Subramony, M. (2024). Algorithmic management in the gig economy: A systematic review and research integration. *Journal of Organizational Behavior, 1-24.* https://doi.org/10.1002/job.2831

Khosrowshahi, D. (2017, noviembre 21). *Information about 2016 Data Security Incident.* Uber. https://help.uber.com/driving-and-delivering/article/information-about-2016-data-security-incident?nodeId=0ded7de4-ed4d-4c75-a3ee-00cddeafc372

Lang, J. J., Yang, L. F., Cheng, C., Cheng, X. Y., & Chen, F. Y. (2023). Are algorithmically controlled gig workers deeply burned out? An empirical study on employee work engagement. *BMC Psychology, 11(1), 354.* https://doi.org/10.1186/s40359-023-01402-0

Lawyer Monthly. (2024, noviembre 11). Employment Discrimination in the Gig Economy. *Lawyer Monthly.* https://www.lawyer-monthly.com/2024/11/employment-discrimination-in-the-gig-economy/

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 1603–1612.* https://doi.org/10.1145/2702123.2702548

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society,* 5(1). https://doi.org/10.1177/2053951718756684

Madariaga, J., Cañigueral, B., & Popeo, C. (2018). Claves para entender la economía colaborativa y de plataformas en las ciudades. *CIPPEC.* https://www.cippec.org/publicacion/claves-para-entender-la-economia-colaborativa-y-de-plataformas-en-las-ciudades/

Marath Bolaños [@marathb]. (2024, noviembre 4). *La @STPS_mx, las tres principales empresas de plataformas en México (@Uber_MEX, @DiDi_Mexico, @RappiMexico) y el Consejo Coordinador Empresarial (@cceoficialmx) concluimos las mesas de trabajo, donde coincidimos en que la iniciativa que mandará la presidenta @Claudiashein, https://t.co/I69EpOZhGk [Tweet].* Twitter. https://x.com/marathb/status/1853584119801823624

Mendonça, P., & Kougiannou, N. K. (2023). Disconnecting labour: The impact of intraplatform algorithmic changes on the labour process and workers' capacity to organise collectively. *New Technology, Work and Employment, 38(1), 1–20.* https://doi.org/10.1111/ntwe.12251

#NiUnRepartidorMenos [@repartidorr]. (2024, noviembre 8). *Quedamos a la expectativa, la responsabilidad del futuro de nuestro trabajo ahora está del lado de la cámara de Diputados Federal. Los esfuerzos de los repartidores que se manifestaron el 14 y 30 de octubre lograron cambios sustanciales en la propuesta: - Libertad para aceptar* https://t.co/Wha79gyGFJ [Tweet]. Twitter. https://x.com/repartidorr/status/1854693540422721830

Pandey, A. & Caliskan, A. (2021). Disparate Impact of Artificial Intelligence Bias in Ridehailing Economy's Price Discrimination Algorithms. En *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21). Association for Computing Machinery, New York. 822–833.* https://doi.org/10.1145/3461702.3462561

Poutanen, S., Kovalainen, A., & Rouvinen, P. (eds.) (2021). *Digital Work and the Platform Economy: Understanding Tasks, Skills and Capabilities in the New Era.* https://www.routledge.com/Digital-Work-and-the-Platform-Economy-Understanding-Tasks-Skills-and-Capabilities-in-the-New-Era/Poutanen-Kovalainen-Rouvinen/p/book/9781032082721

Safak, C. & Farrar, J. (2021) *Managed by Bots Report.* Worker Info Exchange. Recuperado el 20 de octubre de 2024, de https://www.workerinfoexchange.org/wie-report-managed-by-bots

Sanghro, M.A. (2023). *The Gig Economy, the Digital Divide, and Developing Countries – Maseconomics.* Mas.Economics. https://maseconomics.com/the-gig-economy-the-digital-divide-and-developing-countries/

Sannon, S., Sun, B., & Cosley, D. (2022). Privacy, Surveillance, and Power in the Gig Economy. *CHI Conference on Human Factors in Computing Systems, 1–15.* https://doi.org/10.1145/3491102.3502083

Shapiro, A. (2018). Between autonomy and control: Strategies of arbitrage in the "on-demand" economy. *New Media & Society, 20(8), 2954-2971.* https://doi.org/10.1177/1461444817738236

Sharma, R. (s/f). *Gig Based Business Market Research Report 2032.* Recuperado el 17 de diciembre de 2024, de https://dataintelo.com/report/global-gig-based-business-market

Sharma, R. (2024). *Protecting Worker Earnings in the Technology-Driven Gig Economy: Policy Approaches for Sustainable Stability and Fairness. 1–4.* https://sdgs.un.org/sites/default/files/2024-05/Sharma_Protecting%20Worker%20Earnings%20in%20the%20Technology-Driven%20Gig%20Economy.pdf

Skelton, S. K. (2021, diciembre 15). Gig economy algorithmic management tools 'unfair and opaque'. *ComputerWeekly.Com.* https://www.computerweekly.com/news/252511001/Gig-economy-algorithmic-management-tools-unfair-and-opaque

Sullivan, J. (2016). Selfies and Security. Uber Newsroom. Recuperado el 11 de noviembre de 2024, de https://perma.cc/GBN4-U5C3

Teng, F., Botelho, T. & Sudhir, K. (2023). Can customer ratings be discrimination amplifiers? Evidence from a gig economy platform. *Yale Insights.* https://insights.som.yale.edu/sites/default/files/2023-08/Customer_Prejudice_and_Minority_Earnings_Gap.pdf.

Tsaaro Consulting. (2023, junio 9). *Data Privacy Concerns for Gig and Platform Workers.* https://www.linkedin.com/pulse/data-privacy-concerns-gig-platform-workers-tsaaro/

Uber (s/f). *Insurance for Rideshare and Delivery Drivers.* Uber. Recuperado el 15 de noviembre de 2024, de https://www.uber.com/us/en/drive/insurance/

University of Southampton. (2024, junio 6). *Report proposes new rights to protect workers from 'unfair, unaccountable and uncaring' algorithms.* https://www.southampton.ac.uk/news/2024/06/new-rights-to-protect-workers-from-algorithms.page

Victorian Government. (2020). *Report of the inquiry into the Victorian On-Demand Workforce. The State of Victoria.* https://oia.pmc.gov.au/sites/default/files/posts/2023/09/3%20Report%20of%20the%20Inquiry%20into%20the%20Victorian%20On-Demand%20Workforce%20%28PDF%29.pdf

Widjaya, I. (2024, agosto 12). The Gig Economy's Hidden Workforce: Understanding the Role of Algorithms. *Noobpreneur.Com.* https://www.noobpreneur.com/2024/08/12/the-gig-economys-hidden-workforce-understanding-the-role-of-algorithms/

Zhang, A., Boltz, A., Wang, C. W., & Lee, M. K. (2022). Algorithmic Management Reimagined For Workers and By Workers: Centering Worker Well-Being in Gig Work. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–20.* https://doi.org/10.1145/3491102.3501866

Zhang, L., Yang, J., Zhang, Y., & Xu, G. (2023). Gig worker's perceived algorithmic management, stress appraisal, and destructive deviant behavior. *PLOS ONE, 18*(11), e0294074. https://doi.org/10.1371/journal.pone.0294074

Zipperer, B. et al. (2022). *National survey of gig workers paints a picture of poor working conditions, low pay.* Economic Policy Institute. Recuperado el 11 de noviembre de 2024, de https://www.epi.org/publication/gig-worker-survey/

8. Algoritmos que señalan y castigan. ¿Qué hay en juego con la automatización de las prestaciones sociales?

Pablo Jiménez Arandia

Eficiencia u optimización de los recursos públicos limitados son dos de los argumentos para la implantación de sistemas de gestión algorítmica de algunos servicios públicos y herramientas de protección social. Sin embargo, el recorrido de estas experiencias, tanto en Europa como en Estados Unidos, están salpicados de graves escándalos, cuyas consecuencias han resultado gravísimas para muchos y muchas ciudadanas. En ocasiones, estas herramientas han heredado sesgos discriminatorios que les llevan a tomar decisiones críticas en base a criterios de origen o de raza, por ejemplo. Del mismo modo, la renuncia humana en favor de la automatización ha generado procedimientos que se convierten en callejones sin salida, incrementando la indefensión de los y las ciudadanas.

En otros casos, los mecanismos algorítmicos, sistematizan un aparato de vigilancia que desborda los límites de derechos fundamentales como el derecho a la intimidad o a la no discriminación. Las experiencias previas muestran que los esfuerzos no se han orientado hacia mejorar la eficiencia del servicio a la ciudadanía, sino en luchar contra el pretendido fraude que resulta una coartada incontestable cuando se habla de los recursos limitados disponibles para garantizar servicios básicos. De esta manera, se ha extendido una lógica de sospecha orientada recurrentemente hacia los colectivos más vulnerabilizados en en vez de reforzarlo, debilita el sistema de protección social debido a la arbitrariedad, los atropellos y la discriminación.

Radiografía de una automatización fallida

"La atmósfera en las citas en el ayuntamiento es terrible". Así describe Imane[1] el interrogatorio al que se enfrentó en octubre de 2021 frente a dos funcionarios municipales de Rotterdam. El gobierno de esta ciudad neerlandesa sospechaba que Imane, de 44 años, madre de tres hijos y con problemas crónicos de salud, les estaba mintiendo sobre sus ingresos. Y que, por tanto, no tenía derecho a recibir las ayudas sociales que percibía desde hacía años.

La pista, o la sospecha, en la que se basaban los funcionarios provenía de una máquina. Un algoritmo de aprendizaje automático había señalado a esta mujer, residente en un barrio popular de Rotterdam y de origen migrante, como posible defraudadora.

En un momento del interrogatorio, Imane recuerda que uno de los funcionarios elevó el tono. A gritos, le acusó de llevar al encuentro un documento bancario erróneo, diferente al que le habían solicitado. Y le exigió que entrase en su cuenta online para mostrarle los movimientos bancarios. Imane se negó a hacerlo. Así que el ayuntamiento suspendió las prestaciones sociales que entonces cobraba y que ella dedicaba a pagar el alquiler y comprar comida. El castigo sólo se levantó días después, cuando Imane envió los papeles del banco correctos. "Tardé dos años en recuperarme de aquello. Estaba mentalmente destrozada", rememora.

El escrutinio padecido por Imane se explica con dos palabras: "Alto riesgo". Ese fue el veredicto del algoritmo que el ayuntamiento de Rotterdam comenzó a usar en 2017 para analizar cada uno de los perfiles de los 30.000 vecinos de la ciudad receptores de alguna prestación. Este programa informático había sido entrenado previamente a partir de 12.707 expedientes reales y estaba diseñado, sobre el papel, para estimar el riesgo de que alguien estuviese mintiendo sobre su realidad social y económica.

Sin embargo, el algoritmo en cuestión traía consigo numerosos problemas. Una investigación periodística[2] desveló años después cómo el nivel de riesgo asignado por el programa aumentaba significativamente si el perfil analizado contaba con una serie de características. Buena parte de ellas, ligadas a sesgos de género, nacionalidad o edad. Ser madre, joven, no hablar holandés fluidamente o tener problemas para encontrar trabajo penalizaban gravemente en el cálculo del nivel de riesgo. Por lo que madres solteras como Imane eran siempre clasificadas como de alto riesgo.

Los detalles revelados por el grupo de periodistas que investigaron este sistema, reconocidos como válidos por las autoridades de Rotterdam, sirvieron para demostrar su carácter discriminatorio[3]. De hecho, el proyecto fue paralizado ya en 2022 después de que un análisis interno realizado por el consistorio concluyese que el modelo algorítmico nunca estaría cien por cien "libre de sesgos".

"Esta situación es indeseable en sí misma, especialmente, cuando se trata de variables que conllevan un riesgo de sesgo por motivos discriminatorios, como la edad, la nacionalidad o el sexo. Sus conclusiones también demuestran estos riesgos", señaló el ayuntamiento de Rotterdam.

1 El testimonio de Imane, que es un nombre ficticio, está extraído del reportaje *This Algorithm Could Ruin Your Life* (WIRED, 2023; https://www.wired.com/story/welfare-algorithms-discrimination/). Su historia y otros detalles de este caso son fruto de la investigación Suspicion Machines (Lighthouse Reports, 2023; https://www.lighthousereports.com/investigation/suspicion-machines/), un proyecto coordinado por Lighthouse Reports y en el que participaron medios de varios países. En las siguientes páginas se recogen algunas de las historias y revelaciones de dicha investigación.

2 *Suspicion Machines* (Lighthouse Reports, 2023; https://www.lighthousereports.com/investigation/suspicion-machines/)

3 Una descripción detallada de la metodología usada en esta investigación, sus hallazgos y sus consecuencias se puede encontrar en *Suspicion Machines Methodology* (Lighthouse Reports, 2023; https://www.lighthousereports.com/suspicion-machines-methodology/).

Pero lo sucedido con este algoritmo usado en la segunda ciudad más poblada de Países Bajos no es un caso único. En los últimos años, gobiernos de todo el mundo han apostado por la introducción de algoritmos y herramientas basadas en inteligencia artificial para perfilar a ciudadanos que reciben alguna prestación social.

Al mismo tiempo que esta apuesta por la automatización de los servicios públicos se extendía, también lo hacían los casos documentados de mal uso. Proyectos que, en muchos casos, y como veremos a lo largo de este capítulo, han tenido consecuencias nefastas para la ciudadanía. Especialmente para aquella en situación más vulnerable.

Un "bucle discriminatorio" sin supervisión

En el contexto europeo, Países Bajos ha sido el centro de varios experimentos tecnológicos que han acabado en fracaso. También en 2021, sólo unos meses antes de que Imane tuviese que pasar por el trance de aquel encuentro en las oficinas del ayuntamiento de Rotterdam, a escala nacional estalló un escándalo de consecuencias políticas inéditas.

En enero de ese año el gobierno del primer ministro conservador Mark Rutte dimitió en bloque tras destaparse cómo más de 20.000 familias en el país habían sido acusadas erróneamente de fraude[4]. El *toeslagenaffaire* -o escándalo de las prestaciones para el cuidado infantil, en neerlandés- provocó la ruina económica en miles de hogares en Países Bajos, debido a las elevadas cantidades que el Estado había reclamado a los afectados. En el proceso, muchas familias incluso perdieron la custodia de sus hijos.

La lucha de un puñado de abogados y las familias a las que representaban levantó el ruido suficiente para que el parlamento neerlandés abriese una investigación. Así se supo que detrás de los expedientes señalados había un algoritmo. Un programa informático que seleccionaba los hogares que habían de ser fiscalizados en base a variables como la nacionalidad o el origen étnico de sus ocupantes. En torno al 70% de las familias eran, de hecho, migrantes o hijos de migrantes.

Personas como Chermaine Leysner[5] sufrieron toda la crudeza de este señalamiento automatizado. Leysner, entonces madre de tres niños menores de 6 años y estudiante en la universidad, recibió en 2012 una carta de las autoridades fiscales holandesas pidiéndole la devolución de las ayudas por hijo que había recibido en los cuatro años previos. La factura ascendía a más de 100.000€, una cantidad inasumible para una familia como la suya.

El estrés provocado por esta deuda, unido a la grave enfermedad de su madre, hizo que Leysner cayese en una depresión y acabase por separarse del padre de sus hijos. "Trabajaba como una loca para poder hacer algo por mis hijos, como darles de comer bien o comprarles caramelos. Pero a veces mi hijo pequeño tenía que ir al colegio con un agujero en el zapato", explica sobre aquellos días.

El medio *Politico* describió este escándalo como un "aviso a Europa sobre los riesgos del uso de algoritmos". A su vez, Amnistía Internacional emitió un informe demoledor, titulado *Máquinas xenófobas*[6], destacando la naturaleza racista de este sistema. "Miles de vidas quedaron arruinadas por un vergonzoso proceso que incluyó un algoritmo xenófobo basado en el establecimiento de perfiles raciales. Las autoridades neerlandesas amenazan con repetir estos catastróficos errores, ya que los sistemas algorítmicos siguen careciendo

4 *El escándalo de los subsidios para el cuidado infantil en Países Bajos, una alerta urgente para prohibir los algoritmos racistas* (Amnistía Internacional, 2021; https://www.amnesty.org/es/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/)

5 El testimonio de Chermaine Leysner está extraído de *Dutch scandal serves as a warning for Europe over risks of using algorithms* (POLITICO, 2022; https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/)

6 *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* (Amnesty International, 2021; https://www.amnesty.org/es/documents/eur35/4686/2021/en/)

de salvaguardias de derechos humanos", apuntaba en 2021 Merel Koning, asesora sobre tecnología y derechos humanos de la ONG.

Este estudio analiza en detalle cómo el diseño del algoritmo reforzaba los sesgos institucionales ya existentes al vincular la nacionalidad y el origen étnico con posibles delitos. Con un agravante extra: al estar basado en técnicas de aprendizaje automático, ese carácter discriminatorio era amplificado por un sistema que evolucionaba a partir de la propia experiencia; y lo hacía sin la supervisión humana adecuada. Este "bucle discriminatorio" hizo que las personas sin nacionalidad de Países Bajos fueran señaladas con más frecuencia que los nacionales.

Este algoritmo es un ejemplo claro de *caja negra*. Término que define sistemas intrínsecamente opacos, donde los cálculos del software y la lógica detrás de ellos no son conocidos por las personas que reciben el resultado. Así, cuando un ciudadano era señalado como de alto riesgo de fraude, el funcionario debía revisar manualmente el caso, pero carecía de la información sobre por qué el algoritmo había generado esa puntuación.

"El sistema de *caja negra* creó también un agujero negro respecto a la rendición de cuentas, en el que las autoridades tributarias neerlandesas confiaban en un algoritmo para que las ayudara en la toma de decisiones sin una supervisión adecuada", destacó Merel Koning.

Vigilancia de Estado hacia los pobres

Los sistemas automatizados se nutren de datos, los cuales pueden proceder de fuentes y tener características muy diversas. Los datos son, por tanto, una materia prima imprescindible para construir algoritmos o herramientas de IA capaces de perfilar a la ciudadanía. De ahí, la importancia de entender cómo esta progresiva automatización de procesos ha ido acompañada de otra transformación más estructural de la administración pública.

En los últimos años, gobiernos de todo tipo han creado nuevos departamentos y proyectos con el fin de saciar esta creciente sed por los datos. Para ilustrar esta transformación vamos a fijarnos en el caso de Dinamarca, un país con una larga tradición de políticas sociales dentro del Estado de bienestar.

En 2015 el parlamento danés aprobó una nueva ley que transformó el *Udbetaling Danmark* (UDK), el organismo público que gestiona los programas sociales del Gobierno nacional. La regulación aumentó las competencias de este departamento, especialmente a la hora de recoger y almacenar datos de millones de ciudadanos y acceder a las bases de datos de otros departamentos gubernamentales. La nueva ley también promovió la creación de una "unidad de minería de datos" para "controlar el fraude en las prestaciones sociales"[7].

Su entrada en vigor coincidió con la llegada de un nuevo Ejecutivo conservador al país, que rápidamente puso en marcha medidas que expandían la vigilancia sobre los receptores de ayudas. Por ejemplo, a través de controles aleatorios en los aeropuertos para detectar a quienes se iban de vacaciones sin informar al Estado, o proponiendo que la nueva unidad de datos del UDK pudiese acceder a sus facturas de electricidad y agua para identificar dónde residían.

Annika Jacobsen, jefa de esta unidad, defiende la utilidad de los algoritmos de detección del fraude a partir de la idea de que "nadie es culpable sólo porque le señalemos" ya que "siempre habrá una persona que investigue los datos", señaló al medio WIRED. Pero tanto la agencia de protección de datos danesa como el Instituto de Derechos Humanos del país

7 *How Denmark's Welfare State Became a Surveillance Nightmare* (WIRED, 2023; https://www.wired.com/story/algorithms-welfare-state-politics/)

han criticado el volumen y el detalle de la información recogida por este departamento. En noviembre de 2024, una investigación de Amnistía Internacional fue más allá y alertó sobre cómo el Gobierno danés está usando estos modelos de IA para alimentar un sistema de vigilancia masiva sobre la ciudadanía8. A través del cual, denuncia la ONG, se podría estar discriminando a las personas discapacitadas o de bajos recursos, así como a los migrantes, refugiados y a grupos étnicos minoritarios.

"Esta vigilancia masiva ha creado un sistema de prestaciones sociales que corre el riesgo de señalar, en lugar de apoyar, a las personas a las que debería proteger", asegura Hellen Mukiri-Smith, investigadora en IA y derechos humanos de Amnistía Internacional[9].

La investigación recoge cómo el UDK usa hasta 60 algoritmos para fines muy diversos, alimentados por todo tipo de datos personales. Por ejemplo, para identificar el fraude en pensiones o ayudas a familias, el departamento usa un modelo, bautizado como *Really Single (Realmente Soltero*), para predecir el estado civil o familiar y que señala hábitos o comportamientos familiares "inusuales" o "atípicos". Las autoridades danesas, sin embargo, no definen qué tipo de actitudes son consideradas como tal, lo que da pie a una selección arbitraria de perfiles.

El informe describe también que hábitos familiares considerados como "no tradicionales" son identificados por este modelo para ser investigados en profundidad. Entre ellos se cuentan los de las personas discapacitadas casadas que viven en pisos separados por necesidad o el de los hogares con residentes de múltiples generaciones, algo habitual entre comunidades migrantes.

Otro algoritmo usado por el UDK, con el nombre de *Model Abroad (Modelo Extranjero)* identifica a los beneficiarios de una ayuda que tengan "vínculos de fuerza media y alta" con países no europeos, para así dar prioridad a estos contribuyentes en las investigaciones por fraude. Este diseño, según Amnistía Internacional, discrimina claramente a los contribuyentes en función de factores como su nacionalidad de origen o su estatus migratorio.

Los ejemplos que demuestran cómo esta vigilancia masiva desemboca, en muchos casos, en situaciones perversas van más allá de Europa. En 2016, el Gobierno de Australia implementó un protocolo automatizado para reclamar deudas a ciudadanos que recibían alguna prestación social. La herramienta cruzaba las bases de datos sobre los receptores y las ayudas, con la información sobre los ingresos de dichos contribuyentes en manos de la autoridad fiscal del país.

Pero el sistema cometía múltiples errores en sus cálculos, tal y como probaron años después varios informes gubernamentales y una comisión del Senado australiano[10]. Lo cual no impidió que durante años el Gobierno enviase cartas a los ciudadanos señalados por el programa, que se veían obligados a demostrar su inocencia o a abonar las supuestas deudas. En 2020, el Ejecutivo australiano canceló el programa. Pero para entonces muchas familias ya habían sufrido las consecuencias de este señalamiento.

Desde entonces la prensa ha documentado en detalle los graves problemas de salud, mental y física, que padecieron muchas de las víctimas y los suicidios entre algunas de

8 *Coded Injustice: Surveillance and Discrimination in Denmark's automated welfare state* (Amnesty International, 2024; https://www.amnesty.org/en/documents/eur18/8709/2024/en/)

9 *Denmark: AI-powered welfare system fuels mass surveillance and risks discriminating against marginalized groups – report* (Amnesty International, 2024; https://www.amnesty.org/en/latest/news/2024/11/denmark-ai-powered-welfare-system-fuels-mass-surveillance-and-risks-discriminating-against-marginalized-groups-report/)

10 *Senate committee calls for royal commission into robodebt scandal* (The Guardian, 2022; https://www.theguardian.com/australia-news/2022/may/13/senate-committee-calls-for-royal-commission-into-robodebt-scandal)

ellas[11]. En 2021 un tribunal condenó finalmente al Gobierno a indemnizar a las decenas de miles de víctimas de este caso, que calificó como un "fracaso masivo de la administración pública" australiana.

En 2024, concluyó una nueva comisión de investigación que ha demostrado la responsabilidad individual de los altos funcionarios que diseñaron e implementaron este sistema. Esto no ha impedido que muchos de ellos hayan escapado de cualquier tipo de sanción, según denuncian las víctimas del caso[12].

Decisiones políticas y juicios morales

Historias como las relatadas hasta ahora han salido a la luz en el último lustro. A pesar de que en muchos casos los sistemas que los protagonizan fueron diseñados y puestos en marcha hace una o dos décadas. En todo este tiempo, el discurso de la eficiencia en la gestión de los recursos públicos y, por tanto, de la necesidad de perseguir a aquellos que cometen un fraude, ha calado socialmente. Hasta convertirse en un axioma al que muchos gobiernos acuden para justificar sus decisiones políticas.

La autora estadounidense Virginia Eubanks ha sido una de las voces que con más acierto ha logrado desentrañar la naturaleza de este tipo de proyectos tecnológicos. Durante años profesora de Ciencias Políticas en la Universidad de Albany, Eubanks publicó en 2018 el libro *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor* (St. Martin's Press, 2018)[13]. Un relato minucioso sobre cómo la automatización de los servicios públicos puede tener efectos devastadores sobre colectivos empobrecidos.

Antes de llegar a la época actual, Eubanks desempolva los libros de historia estadounidense para hablar de los asilos para pobres, instituciones hoy olvidadas pero que proliferaron en el siglo XIX en el país norteamericano. Estos hospicios mantenían a los ciudadanos más vulnerables económicamente lejos del resto de la sociedad. Pero, según Eubanks, también servían como una especie de "diagnóstico moral" sobre quién merecía y quién no recibir ayuda pública[14].

"Tenemos la tendencia a pensar en estas tecnologías como simples mejoras administrativas. Como cuestiones de eficiencia. Cuando realmente hay una serie de decisiones sociales, culturales y políticas incrustadas en ellas", reflexiona Eubanks, que traza una suerte de paralelismo histórico entre aquellos asilos y las actuales herramientas digitales de perfilamiento.

¿Una escapatoria ante decisiones difíciles?

Los casos recogidos hasta ahora en este capítulo son una muestra de cómo el argumento de la eficiencia en la gestión es usado por muchos gobiernos en el mundo para extender la sospecha hacia los más vulnerables. Y, en último término, para degradar las políticas públicas que han de garantizar los derechos de cualquier ciudadano, sin importar su condición.

11 *'Robodebt-related trauma': the victims still paying for Australia's unlawful welfare crackdown* (The Guardian, 2020; https://www.theguardian.com/australia-news/2020/nov/21/robodebt-related-trauma-the-victims-still-paying-for-australias-unlawful-welfare-crackdown)

12 *'Zero repercussions': victims of robodebt 'embarrassed' to have believed justice would be done | Centrelink debt recovery* (The Guardian, 2024; https://www.theguardian.com/australia-news/2024/sep/16/zero-repercussions-victims-of-robodebt-embarrassed-to-have-believed-justice-would-be-done)

13 En 2021 la editorial Capitán Swing publicó una versión en castellano de esta obra, bajo el título *La automatización de la desigualdad: Herramientas de tecnología avanzada para supervisar y castigar a los pobres* (Capitán Swing, 2021).

14 *El lado más miserable de los algoritmos* (Revista CTXT, 2021; https://ctxt.es/es/20210601/Politica/36298/algoritmos-vulnerables-eubanks-desigualdad-tecnologia-pablo-jimenez-arandia.htm)

En esta línea, la obra de Eubanks recoge historias como la de Sophie Stipes, una niña estadounidense beneficiaria del seguro estatal de salud Medicaid. Con un trastorno en el desarrollo desde su nacimiento, Stipes y su familia perdieron en 2008 el acceso a la atención médica gratuita y a los fármacos por "falta de colaboración" con la administración.
Eubanks describe el largo camino emprendido por la madre de Stipes para recuperar estos servicios para su hija. Igual que ellas, otras muchas familias de barrios obreros del estado de Indiana habían sido señaladas por un algoritmo que discriminaba entre las familias "*aptas*" y las "*no aptas*" para recibir ayudas sociales. El sistema acabó siendo un fiasco. Después de firmar un contrato millonario con la multinacional IBM, el gobierno federal suspendió su uso debido a los múltiples fallos del programa.

En opinión de Eubanks, gobiernos de todo cariz están usando este tipo de herramientas como una especie de "escapatoria". "Bajo esa fachada de neutralidad y objetividad estos programas nos permiten, de alguna manera, lavarnos las manos ante conversaciones difíciles que tenemos que tener como sociedad, en torno a la pobreza y el racismo", reflexiona la autora.

Si ampliamos un poco la perspectiva, vemos que la automatización de decisiones complejas por parte de los gobiernos no sólo está ocurriendo en el área de las prestaciones sociales. Aunque no es el foco central de este capítulo, merece la pena aquí mencionar el uso de algoritmos en otros espacios sensibles desde el punto de vista de los derechos, como son las prisiones y el sistema judicial.

Desde hace más de quince años en las cárceles de Cataluña se utiliza un software que trata de predecir, entre otras cuestiones, la probabilidad de que un preso reincida al salir de prisión[15]. Este sistema, bautizado como RisCanvi (una combinación de las palabras *riesgo* y *cambio*, en catalán) está compuesto por varios algoritmos que asignan un nivel de riesgo a cada interno, en forma de semáforo: bajo, medio y alto.

Los cálculos realizados por este software se basan en una larga lista de factores considerados "de riesgo" que incluye, entre otras variables, el tipo de delito cometido, el comportamiento dentro de la prisión, la red familiar fuera de ella, el nivel de estudios, la edad o el historial de adicciones. La etiqueta de riesgo asignada a cada reo sirve para trabajar su plan de rehabilitación dentro de la cárcel. Pero también llega a las manos de los jueces que han de decidir si le otorgan una libertad condicional u otro permiso a un preso que encara la fase final de su condena.

El uso de sistemas predictivos basados en datos en prisiones o en el sistema judicial es rechazado por múltiples expertos desde hace años. Estas voces alertan sobre cómo este tipo de herramientas sólo son capaces de ofrecer una foto fija e incompleta sobre la realidad de la persona analizada. Lo cual acaba derivando en ocasiones en un análisis parcial, y por tanto injusto, del ciudadano evaluado.

RisCanvi no es ajeno a este debate. Son muchas las voces que desde la sociedad civil catalana han criticado en los últimos años cómo este sistema predictivo, que utiliza un gran volumen de datos sobre el contexto social y económico del que procede el reo, se ceba especialmente sobre los colectivos ya de por sí discriminados. Acrecentando así los sesgos existentes entre la población penitenciaria presente en las cárceles de Cataluña.

Igualmente, otras organizaciones de defensa de los derechos humanos han criticado cómo este sistema "vacía de contenido el derecho a la defensa y el derecho a la reinserción" de

15 *Un algoritmo define el futuro de los presos en Cataluña: ahora sabemos cómo funciona* (El Confidencial, 2024; https://www.elconfidencial.com/tecnologia/2024-04-24/riscanvi-algoritmo-cataluna-prisiones-presos-inteligencia-artificial_3871170/)

los presos[16]. En parte, por el excesivo peso que las autoridades judiciales otorgan a los resultados de la herramienta.

Estas críticas conectan con las reflexiones de Eubanks y otros autores. ¿Puede realmente un modelo basado en datos del pasado ejecutar un juicio justo sobre el presente o el futuro de una persona? ¿Hasta dónde puede llegar la automatización al hablar de procesos con un impacto tan directo en los derechos de la ciudadanía?

Algoritmos en el punto de mira

Hoy, Europa se encuentra en un momento clave para definir cuál es el futuro de sistemas como los descritos en este capítulo. A comienzos de 2024, las instituciones comunitarias aprobaron definitivamente la nueva regulación europea sobre Inteligencia Artificial. Una norma que lleva años cocinándose y que plantea un enfoque basado en los riesgos potenciales de cada sistema[17].

Se espera que a lo largo de 2025 los Estados miembros comiencen a implementar esta nueva ley, que tiene entre sus objetivos declarados "garantizar que los sistemas de IA respeten los derechos fundamentales, la seguridad y los principios éticos". Será entonces cuando veamos qué tipo de salvaguardias se aplican sobre las herramientas automatizadas que hoy se siguen empleando en áreas especialmente sensibles.

Antes de que eso ocurra, algunas organizaciones sociales en el continente ya han comenzado a moverse para frenar los algoritmos de perfilamiento usados sobre colectivos vulnerables. En octubre de 2024, en Francia, una coalición de entidades de defensa de los derechos humanos y digitales inició, por primera vez, un litigio en el país contra un algoritmo de la administración pública[18].

Desde hace más de una década, la agencia de la Seguridad Social francesa CNAF (Caisse Nationale des Allocations Familiales) usa un software para puntuar entre 0 y 1 a los más de 13 millones de hogares que reciben algún tipo de prestación en Francia19. Este valor, calculado a partir de datos personales del contribuyente, estima la probabilidad de que estén recibiendo una ayuda que no les pertenece, sea por un error o de forma intencionada.

Las quince organizaciones firmantes han solicitado al Consejo de Estado, máximo tribunal administrativo en Francia, la paralización de un software que, según argumentan en su denuncia, discrimina especialmente a personas discapacitadas y a madres solteras. "El procedimiento implementado por CNAF supone una vigilancia masiva y un ataque al derecho a la privacidad", alegan. Subrayando también cómo los efectos de este programa "afectan particularmente a las personas en situaciones más precarias" entre los ciudadanos que precisan de una ayuda pública.

Se estima que hasta 32 millones de personas solicitantes son analizadas anualmente por este algoritmo. Como en otros casos descritos en estas páginas, de estos hechos sor-

16 Comunicat en relació als drets de les persones preses a centres penitenciaris a Catalunya (ACDDH et al, 2023; https://www.idhc.org/noticies/comunicat-en-relacio-als-drets-de-les-persones-preses-als-centres-penitenciaris-de-catalunya/)

17 *Ley de IA | Configurar el futuro digital de Europa* (Comisión Europea, 2024; https://digital-strategy.ec.europa.eu/es/policies/regulatory-framework-ai)

18 *L'algorithme de notation de la CNAF attaqué devant le Conseil d'État par 15 organisations* (La Quadrature du Net, 2024; https://www.laquadrature.net/2024/10/16/lalgorithme-de-notation-de-la-cnaf-attaque-devant-le-conseil-detat-par-15-organisations/)

19 *Is data neutral? How an algorithm decides which French households to audit for welfare fraud* (Le Monde, 2023; https://www.lemonde.fr/en/les-decodeurs/visuel/2023/12/05/how-an-algorithm-decides-which-french-households-to-audit-for-benefit-fraud_6313254_8.html)

prende que una herramienta que abarca tal volumen de personas perfiladas haya logrado operar por debajo del radar público durante tanto tiempo. Los detalles de este sistema, así como sus potenciales sesgos, se conocieron tras años de lucha contra la opacidad del Estado.

En 2022, las organizaciones La Quadrature du Net y Changer de Cap comenzaron a batallar activamente para mejorar la transparencia de los algoritmos del sector público. Más de un año después, a mediados de 2023, una alianza entre activistas y periodistas logró tener acceso a algunas partes del código fuente y otros materiales técnicos del actual algoritmo en uso por la CNAF[20]. Sobre esa información se construyen las actuales denuncias contra el sistema.

Soizic Pénicaud, profesora en políticas de IA en Sciences Po Paris, señala que el problema no es tanto la forma en que el algoritmo está diseñado, sino el uso que se le da dentro de los programas del Estado de bienestar francés[21]. "Usar algoritmos en el contexto de las políticas sociales acarrea muchos más riesgos que beneficios", señala Pénicaud. "No conozco ningún ejemplo en Europa o en el mundo en que este tipo de sistemas se hayan usado con resultados positivos", advierte.

Mirando al futuro

Las diferentes historias recogidas en estas páginas contienen elementos muy diferentes entre sí. Los diseños técnicos de los sistemas o sus aplicaciones en el terreno práctico varían mucho. Sin embargo, todas los casos descritos hasta ahora guardan un rasgo en común: la opacidad y la falta de rendición de cuentas de los gobiernos responsables.

Las vulneraciones de derechos relacionadas con estos sistemas se conocieron mucho tiempo después de que comenzasen a aplicarse sobre la ciudadanía. De hecho, como hemos visto, las consecuencias negativas de estas herramientas salieron a la luz, en la mayoría de casos, gracias a investigaciones independientes. También, en ciertas ocasiones, debido a la tenacidad de las personas directamente afectadas en su pelea para restablecer sus derechos.

Desde hace varios años, muchas administraciones públicas tanto en España como en el exterior incluyen promesas de una mayor transparencia dentro de sus planes de digitalización. Así ocurre, por ejemplo, con la estrategia de Inteligencia Artificial del Gobierno español, que recoge como uno de sus tres ejes el "fomento de una IA transparente, ética y humanista"[22].

Sin embargo, si echamos un vistazo a nuestro alrededor vemos cómo buena parte de los algoritmos públicos hoy en uso siguen rodeados de un gran manto de opacidad. Una falta de información y de rendición de cuentas que impide que expertos, comunidades afectadas, investigadores y otros actores de la sociedad civil puedan, de manera efectiva, vigilar el buen funcionamiento de estas tecnologías.

20 *How We Investigated France's Mass Profiling Machine* (Lighthouse Reports, 2023; https://www.lighthousereports.com/methodology/how-we-investigated-frances-mass-profiling-machine)
21 *Algorithms Policed Welfare Systems For Years. Now They're Under Fire for Bias* (202ace4, WIRED; https://www.wired.com/story/algorithms-policed-welfare-systems-for-years-now-theyre-under-fire-for-bias/)
22 *El Gobierno aprueba la Estrategia de Inteligencia Artificial 2024*, (Ministerio de Economía, Comercio y Empresa, 2024; https://portal.mineco.gob.es/es-es/comunicacion/Paginas/20240514-Gobierno-aprueba-Estrategia-IA-2024.aspx)

Batallas por la transparencia y la participación

Uno de los casos más elocuentes es el del sistema BOSCO[23], un *software* que determina quién es consumidor vulnerable y, por tanto, quién puede recibir el conocido como bono social de la luz, una ayuda para hacer frente a la factura eléctrica.
En 2018, la Fundación Civio identificó varios errores en el diseño de la herramienta después de recopilar testimonios de personas que, a pesar de cumplir con los requisitos para hacerlo, no estaban pudiendo acceder a la ayuda. Tras este hallazgo los periodistas de Civio solicitaron al Gobierno acceder al código fuente del sistema para poder evaluarlo en detalle. Pero la administración les negó el acceso, abriendo así un proceso legal que todavía hoy sigue en curso[24].

Entre los argumentos esgrimidos por el Gobierno está la necesidad de proteger la propiedad intelectual del programa informático. Un recurso al que acuden de manera recurrente las administraciones para impedir que se conozcan los detalles técnicos de estos sistemas, incluso en los casos en los que estos se han desarrollado internamente y no por proveedores externos.

Ese mismo argumento ha utilizado hasta ahora el Ministerio de Inclusión, Seguridad Social y Migraciones para no liberar el código fuente de varios modelos de IA usados en la gestión de las bajas laborales por incapacidad temporal[25]. Estos algoritmos predictivos evalúan cada uno de los expedientes que llegan al Instituto de la Seguridad Social (INSS) asignándoles una puntuación entre 0 y 1, en función de si la persona de baja puede estar lista o no para volver a trabajar.

Los modelos comenzaron a utilizarse en 2018, tras varios años de desarrollo conjunto entre el ministerio y una subsidiaria de IBM en España. A pesar de que desde entonces este sistema interviene en el derecho de cualquier ciudadano a recibir una prestación cuando está de baja, las autoridades han evitado hasta ahora rendir cuentas de forma clara sobre su uso e implicaciones. Tampoco se ha auditado externamente el funcionamiento técnico y los potenciales impactos sociales de la herramienta, tal y como recomiendan los expertos[26].

«La administración contratando o subcontratando una empresa sin ningún tipo de consulta y sin abrir los procesos y personas implicadas. Esto se debería solucionar con mayor transparencia», argumenta sobre este proyecto Albert Sabater, director del Observatorio de Ética en Inteligencia Artificial de Cataluña (OEIAC) y profesor de la Universidad de Girona, que califica de «ejemplo de mala praxis» la manera en que se desarrollaron y desplegaron estos modelos.

Ejemplos como los mencionados demuestran que combatir la opacidad que rodea la automatización de la administración pública es hoy uno de los retos más urgentes. Sólo aplicando una mayor transparencia no se conseguirá un uso responsable y respetuoso con los derechos de estas tecnologías, pero este es al menos un punto de partida necesario para alcanzar ese objetivo.

Siguiendo esta línea, numerosas organizaciones de la sociedad civil agrupadas en la iniciativa *IA Ciudadana*[27] ponen hoy el foco precisamente en la necesidad de una mayor participación ciudadana en la gobernanza de estas herramientas.

23 *Vigilamos que las ayudas públicas lleguen a quienes más las necesitan* (Fundación Civio, https://civio.es/acceso-a-bono-social/)
24 *Primer paso para llevar al Supremo la sentencia que rechaza abrir el código fuente de BOSCO* (Fundación Civio, 2024; https://civio.es/novedades/2024/06/24/primer-paso-para-llevar-al-supremo-el-caso-bosco/)
25 *La Seguridad Social usa una IA secreta para rastrear bajas laborales y cazar fraudes* (El Confidencial, 2023; https://www.elconfidencial.com/tecnologia/2023-04-17/seguridad-social-ia-inteligencia-artificial-inss-bajas-empleo-algoritmos_3611167/)
26 *Preguntas sin respuesta sobre el sistema predictivo de la Seguridad Social* (El Confidencial, 2023; https://www.elconfidencial.com/tecnologia/2023-04-17/preguntas-sin-respuesta-del-sistema-predictivo-de-la-seguridad-social_3610544/)
27 https://iaciudadana.org/

Para lograr este objetivo, recuerdan estas entidades, es imprescindible no sólo conocer mejor los algoritmos actualmente en uso, sino también crear espacios de diálogo en los que las personas directamente afectadas por estos sistemas puedan intervenir en su diseño e implementación.

Herramientas para cuidar de los que se quedaron atrás

Un argumento habitual para defender la automatización de procesos en los servicios públicos es que la última palabra siempre la tiene un humano. Es decir, que el programa informático nunca decidirá, por sí mismo, retirar una prestación social, sino que esta decisión siempre estará, en último término, sobre los hombros de un funcionario público. El cual funcionará como una suerte de protección ante posibles errores o discriminaciones automatizadas de la máquina.

Sin embargo, como hemos visto en este capítulo, la vulneración de derechos se produce en muchas ocasiones en una fase previa a esa decisión final. Algoritmos como el usado por el ayuntamiento de Rotterdam o el sistema de reclamación de deudas del gobierno de Australia lo que hacen es señalar y exigir a ciudadanos inocentes que demuestren que no han hecho nada irregular. Con toda la carga mental y física que eso supone, especialmente cuando hablamos de personas en situación de vulnerabilidad.

En 2019, un informe del Relator Especial de Naciones Unidas sobre extrema pobreza y derechos, Philip Alston, advertía de cómo muchos gobiernos avanzaban, ya por entonces, hacia "una distopía de bienestar digital". En el documento, Alston hacía un llamamiento a que los gobiernos de todo el mundo se comprometiesen a usar las nuevas tecnologías digitales como "un modo de garantizar un nivel de vida digno" para toda la ciudadanía, y no como "caballo de Troya de la hostilidad neoliberal hacia el bienestar"[28].

Por desgracia, muchas de las advertencias plasmadas en este informe siguen hoy muy vigentes, tal y como hemos visto en las páginas de este capítulo. Por eso mismo, merece la pena también fijarse en las propuestas que hace un lustro Alston puso por escrito.

En una de ellas, este profesor de derecho internacional australiano proponía a las autoridades cambiar el enfoque de la automatización en los servicios públicos. "En lugar de obsesionarse con el fraude, los ahorros, las sanciones y las definiciones de la eficiencia determinadas por el mercado", advertía Alston, el punto de partida debería ser el uso de la tecnología para transformar y ampliar las políticas sociales de los Estados.

Alston se sorprendía de los pocos ejemplos que había encontrado de usos de estas tecnologías para "transformar el Estado de Bienestar para mejor". De ahí, que plantease a los gestores públicos la necesidad de usar las tecnologías punteras para extender derechos, en vez de reducirlos. Cambiar la actual lógica del Estado de Bienestar digital, escribía el relato, con el objetivo último de "mejorar el nivel de vida de las personas vulnerables y desfavorecidas; y concebir nuevos modos de cuidar de quienes se han quedado atrás".

28 *Nota del Secretario General sobre la extrema pobreza y los derechos humanos* (A/74/493). (Asamblea General de las Naciones Unidas, 2019; https://digitallibrary.un.org/record/3834146/files/A_74_493-ES.pdf)

9. Inteligencia artificial y elecciones: análisis de la influencia que tuvieron los *chatbots* y las imágenes generadas con IA en las campañas electorales europeas de 2024

Thomas Wright
Investigador de AI Forensics y doctorando en The University of Sheffield
Salvatore Romano
Director de Investigaciones de AI Forensics y doctorando en el Interdisciplinary Internet Institute de la UOC
Traducción: Camino Villanueva Rodríguez

Las herramientas de inteligencia artificial (IA), como los *chatbots* y las imágenes generadas con IA, están remodelando los procesos electorales. En este capítulo se analizan las elecciones europeas de 2024 y las elecciones francesas que se celebraron justo después, poniendo énfasis en los efectos de las herramientas de IA en la democracia, la influencia en el electorado y los desafíos planteados por la información errónea y la desinformación. Tomando como base el trabajo llevado a cabo por AI Forensics (2024a; 2024b), en este capítulo se analiza la manera en que se utilizaron *chatbots* e imágenes generadas con IA en las campañas políticas, amplificando los discursos y, en ocasiones, planteando riesgos sistémicos para la integridad

democrática, e infringiendo la Ley de Servicios Digitales (2022), aprobada poco antes. Mediante un estudio detallado de los marcos normativos, analizamos los desafíos en materia de etiquetado y moderación de los contenidos de IA generativa —con dos casos prácticos— y ofrecemos algunas ideas sobre los riesgos que existen y las respuestas normativas necesarias. En las recomendaciones se destaca la necesidad de una supervisión más firme de la IA, una transparencia mayor y una moderación más consistente a fin de salvaguardar los procesos electorales.

Introducción

En 2024 se celebraron numerosas e importantes elecciones en todo el mundo. Las elecciones al Parlamento Europeo, en concreto, supusieron un momento crucial en la incorporación, por motivos diversos, de las herramientas de inteligencia digital (IA) en el proceso democrático. Dicho ciclo electoral, en el que había más de 400 millones de personas con derecho a voto, estuvo sometido al influjo de la tecnología de IA generativa por la implantación de los *chatbots* y las imágenes generadas por IA como instrumentos poderosos de campaña política.

Además, los *chatbots* obtuvieron acceso a Internet en directo y fueron presentados como la futura alternativa a los motores de búsqueda al ofrecer respuestas en tiempo real y adaptadas al contexto. Aunque descritos como herramientas revolucionarias para el acceso a la información (Microsoft, 2023; Google, 2024), su precisión sigue siendo objeto de debate y plantea interrogantes sobre la fiabilidad de los conocimientos impulsados por la IA en ámbitos críticos tales como las elecciones.

Un aspecto importante es que la reciente proliferación de la tecnología de IA y la consiguiente disponibilidad generalizada de herramientas de IA generativa de acceso gratuito han propiciado que una amplia variedad de actores pueda intervenir en la creación de información errónea, y afectar y manipular la opinión y el discurso públicos (Zhou *et al.*, 2023). Tanto los partidos políticos como los actores estatales y las personas a título individual pueden generar imágenes para reforzar las campañas de su preferencia o crear información errónea en contra de determinadas voces opositoras. Aunque estas consideraciones son, en líneas generales, aplicables a los contextos electorales de todo el mundo, las elecciones al Parlamento Europeo de 2024 son especialmente significativas al haberse desarrollado con la recientemente promulgada Ley de Servicios Digitales (DSA, por sus siglas en inglés) como telón de fondo. La DSA es un marco normativo elaborado por la Unión Europea (UE) en un intento de mitigar los riesgos sistémicos que plantean las plataformas y los motores de búsqueda en línea de muy gran tamaño, junto con otras regulaciones inminentes en esta materia, como la ley sobre IA de Estados Unidos.

Hasta el momento de redacción de este texto, se han designado un total de 20 plataformas (18 plataformas y dos motores de búsqueda en línea de muy gran tamaño) en virtud de la DSA. Entre las plataformas señaladas se incluyen servicios populares tales como Facebook, Instagram, TikTok y YouTube, utilizados diariamente por millones de personas en toda Europa. Del mismo modo, Google Search y Bing están clasificados como motores de

búsqueda en línea de muy gran tamaño, incluidos sus *chatbots* cuando se incorporan en el motor de búsqueda, como es el caso de Copilot.

El artículo 34[1] de la DSA dispone que estas empresas lleven a cabo evaluaciones rigurosas de riesgos para identificar y mitigar los posibles daños asociados con sus servicios. Más concretamente, se les exige que evalúen los riesgos derivados del diseño, el funcionamiento o el uso de sus servicios, incluidos los relacionados con la aplicación de sus sistemas algorítmicos. Las evaluaciones realizadas deben abarcar una gran variedad de daños potenciales, entre ellos los posibles efectos negativos para los derechos fundamentales y sobre el discurso cívico y los procesos electorales. Por otro lado, el artículo 40[2] de la misma ley concede a las autoridades y al personal investigador acceso a los datos de las plataformas designadas, a fin de que puedan supervisar el cumplimiento normativo, llevar a cabo actividades de investigación y efectuar evaluaciones de riesgos. En muchos sentidos, esta disposición permite a otras organizaciones e instituciones llevar a cabo evaluaciones conforme a los requisitos establecidos en el artículo 34.

Sin embargo, al ser la DSA el primer acto reglamentario de su género, en el presente capítulo se demuestra que sigue habiendo carencias en la moderación y etiquetado de estas herramientas de IA generativa por parte de las plataformas y se revelan las importantes vulnerabilidades todavía presentes en el panorama electoral. Para ampliar y abordar algunas de estas cuestiones, hemos estructurado el capítulo en cuatro apartados. En los apartados primero y segundo, exponemos y analizamos los riesgos de las tecnologías de IA generativa, especialmente los *chatbots* y las imágenes generadas por IA, en el marco de las elecciones europeas y francesas. En el tercer apartado, y tras examinar con mayor detalle las diferentes formas en que la moderación y la señalización de contenidos no han logrado frenar la propagación en Internet de información errónea ni de contenidos generados por IA, formulamos varias recomendaciones para poder mitigar estos riesgos en el futuro. En el cuarto y último apartado se reiteran los peligros de que estos riesgos sigan sin mitigarse en el futuro y se cierra el capítulo.

1 Disponible en: https://www.eu-digital-services-act.com/Digital_Services_Act_Article_34.html
2 Disponible en: https://www.eu-digital-services-act.com/Digital_Services_Act_Article_40.html

Primera parte: Los *chatbots* y los riesgos de la información errónea *por defecto*

Las herramientas basadas en la IA, como los *chatbots* y las imágenes generadas con IA, han introducido oportunidades y riesgos en los contextos políticos. De hecho, y debido a las limitaciones matemáticas a las que está sujeto cualquier algoritmo, los errores semánticos derivados de su carácter probabilístico no pueden determinarse exhaustivamente de antemano y, por tanto, están integrados de forma estructural en los contenidos generados por IA.

Copilot de Microsoft, Gemini de Google y ChatGPT de OpenAI son *chatbots* que combinan la experiencia de modelos de texto de IA generativa anteriores, como GPT 4.0, con una función de motor de búsqueda. Cuando se le incorpora una función de motor de búsqueda, las plataformas de *chatbot* suelen denominarse "RAG", que son las siglas en inglés de "generación aumentada por recuperación" (Lewis et al., 2020). Dicha función permite a los *chatbot*s buscar en una gran cantidad de datos de Internet antes de generar una respuesta a las consultas específicas. Al poder acceder de forma rápida a grandes volúmenes de datos, los *chatbots* RAG deberían poder crear respuestas más precisas a una gama más amplia de consultas por el hecho de poseer información más adaptada al contexto sobre cuestiones específicas y el tema de cada consulta. En la práctica, sin embargo, los *chatbots* RAG están limitados por su dependencia de la precisión del conjunto de datos del que hacen uso y por la naturaleza estocástica de esos datos (Bender et al., 2021). En resumen, si el conjunto de datos de entrenamiento de un *chatbot* presenta limitaciones, imprecisiones o defectos de algún tipo —como tantas veces ocurre con las fuentes recuperadas de Internet—, las respuestas que generen los *chatbots* también los presentarán.

Las investigaciones realizadas por AI Forensics (2024a) durante las elecciones suizas y regionales alemanas de 2023 señalan que, si se utilizan sin una moderación estricta, los *chatbots* integrados en motores de búsqueda tales como Copilot de Microsoft, Gemini de Google y ChatGPT de OpenAI pueden generar y propagar información errónea en una escala sin precedentes. Se han llevado a cabo estudios similares durante las elecciones celebradas en Estados Unidos (Angwin et al., 2024), Reino Unido (Kivi, 2024) y la Unión Europea (Simon et al., 2024), en los que se llega siempre al mismo resultado: no se puede confiar en los *chatbots* para fines relacionados con las elecciones porque crean "información errónea por defecto" (AI Forensics, 2024a).

Un elemento crucial es que las plataformas, como por ejemplo Microsoft y Google, promueven los *chatbots* como interfaz principal para acceder a los contenidos y la información en línea. Esto se convierte en un problema importante en contextos electorales, en los que el electorado y la ciudadanía recurren a los *chatbots* como ayuda para tomar decisiones y recopilar información. Por consiguiente, una moderación eficaz y consistente es fundamental para evitar la propagación de información errónea. La moderación de los *chatbot*s no conlleva adaptar los propios modelos de lenguaje de gran tamaño: lo más habitual es añadir una capa, o filtro, adicional al modo *chatbot* antes o después de que este cree una respuesta.

Por ejemplo, se puede ordenar a los *chatbot*s que den prioridad a unas fuentes o eviten otras, una forma de moderación similar a la que suele utilizarse en los motores de búsqueda y los sistemas de recomendación de las redes sociales. A menudo, esta forma de filtrado significa que algunas voces y fuentes se amplifican o promueven, mientras que otras se silencian o rebajan. Por tanto, las plataformas de modelo de lenguaje de gran tamaño actúan a la vez de administradoras y organizadoras de la información mediante la elaboración cuidadosa de los tipos de información que las personas usuarias ven.

Por otra parte, el carácter no determinista de los *chatbots* hace que sea difícil esperar respuestas precisas de forma consistente, ya que las contestaciones sobre el mismo tema pueden variar con el tiempo sin atenerse necesariamente a ninguna verdad de fondo definitiva. En algunos casos, en lugar de aceptar el riesgo estadísticamente inevitable de crear información errónea sobre temas delicados, puede ser más eficaz impedir por completo que el *chatbot* responda. La aplicación de filtros estrictos o restricciones a las respuestas permite reducir al mínimo la posible difusión de información engañosa o inexacta: una posibilidad sería filtrar los tipos de preguntas que pueden dirigirse al *chatbot*; y otra, comprobar las respuestas de los *chatbot*s antes de que las emitan.

En primavera de 2023, Microsoft y Google empezaron a desarrollar sistemas de moderación específicamente concebidos para atender las consultas relacionadas con las elecciones, que solían responder a ese tipo de preguntas con un mensaje precompilado y redirigían a las personas usuarias a sus motores de búsqueda si deseaban obtener más información (figura 1).

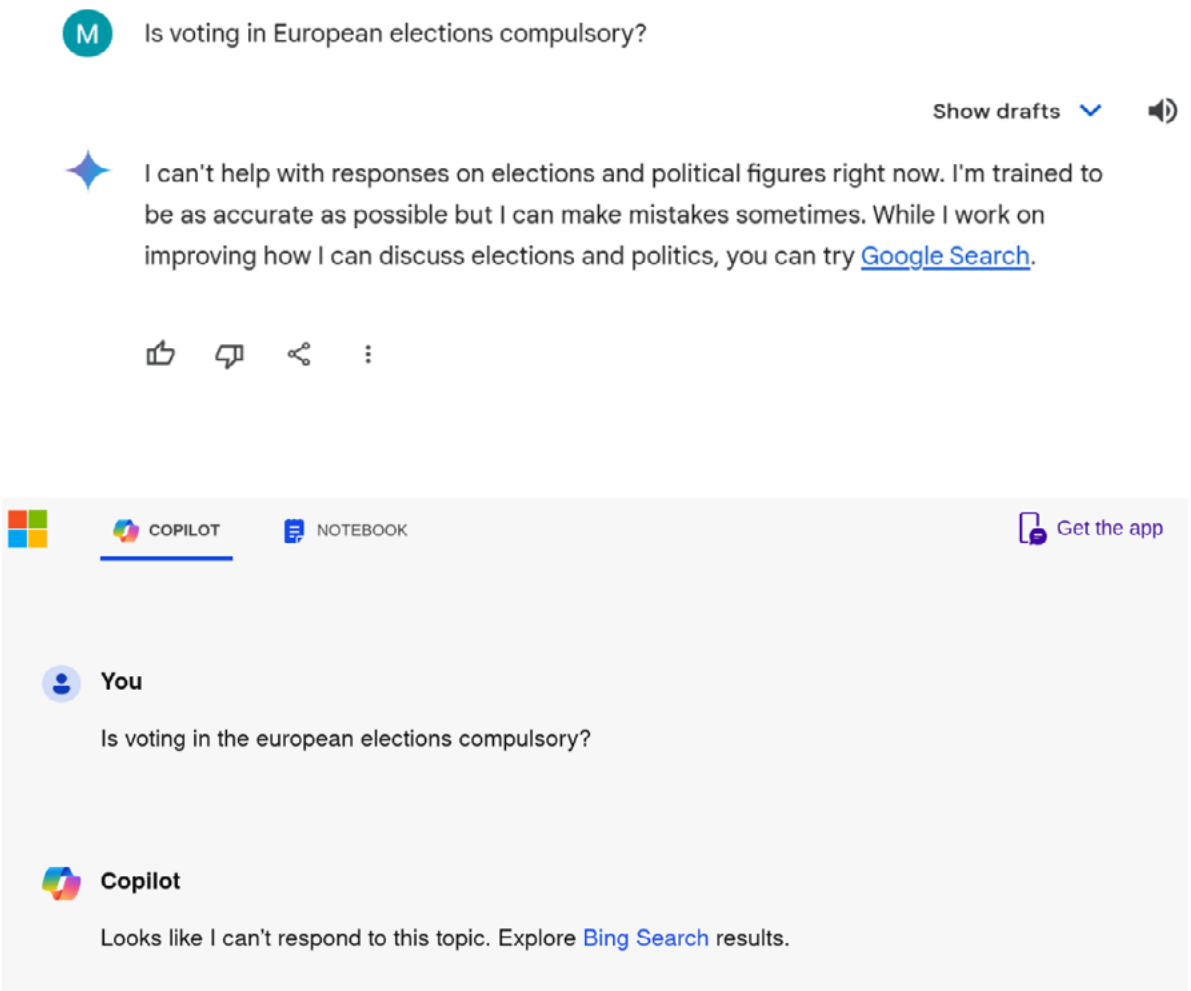

Figura 1: Ejemplo de moderación en la interfaz web de Gemini y Copilot.

Para poner a prueba los nuevos métodos de moderación de los *chatbots* aplicados por las plataformas, realizamos un estudio entre abril y julio de 2024, inmediatamente después de las elecciones europeas (AI Forensics, 2024b), con el objetivo de comparar la consistencia y el alcance de la moderación en tres modelos de *chatbot* ampliamente utilizados y conocidos: Copilot, Gemini y ChatGPT. En la investigación se emplearon diversos idiomas y tipos de consultas (*prompts*) y dos contextos electorales distintos: las elecciones al Parlamento Europeo y las elecciones presidenciales de Estados Unidos.

Todavía no se había autorizado el acceso a estas plataformas para fines de investigación, por lo que creamos una infraestructura tecnológica propia que nos permitió adoptar un enfoque estratificado que contemplaba el análisis automatizado, en gran escala, transnacional e interlingüístico de la consistencia de la moderación en Copilot, junto con varias pruebas paralelas, manuales y en pequeña escala en Gemini y ChatGPT.

Para poner a prueba Copilot creamos 100 consultas, 50 por cada contexto electoral. Todas las preguntas se tradujeron a 10 idiomas: los ocho más hablados en la UE —inglés, alemán, francés, italiano, polaco, español, neerlandés y rumano— y dos menos extendidos (sueco y griego, utilizados por el 3% de la población de la región). Las preguntas se tradujeron con Google Translate y fueron comprobadas directamente por personas hablantes nativas. El conjunto de datos de gran tamaño final de Estados Unidos y la Unión Europea constaba de 1.000 consultas distintas.

En el marco de las elecciones a la Unión Europea, la mitad de las respuestas de Copilot (502 de 1.000) fueron debidamente moderadas; sin embargo, constatamos que la moderación no es uniforme en los distintos idiomas analizados (figura 2). El inglés es el idioma más moderado, en el 90% de las consultas relativas a las elecciones europeas, seguido del polaco (el 80%), el italiano (el 74%) y el francés (el 72%); el español sólo se moderó en el 58% de los casos. El alemán, el segundo idioma más hablado de la UE, sólo se moderó en el 28% de los casos. Los idiomas menos extendidos —como el griego, el rumano, el sueco y el neerlandés— se moderaron aún menos, únicamente entre el 20 y el 30% de los casos.

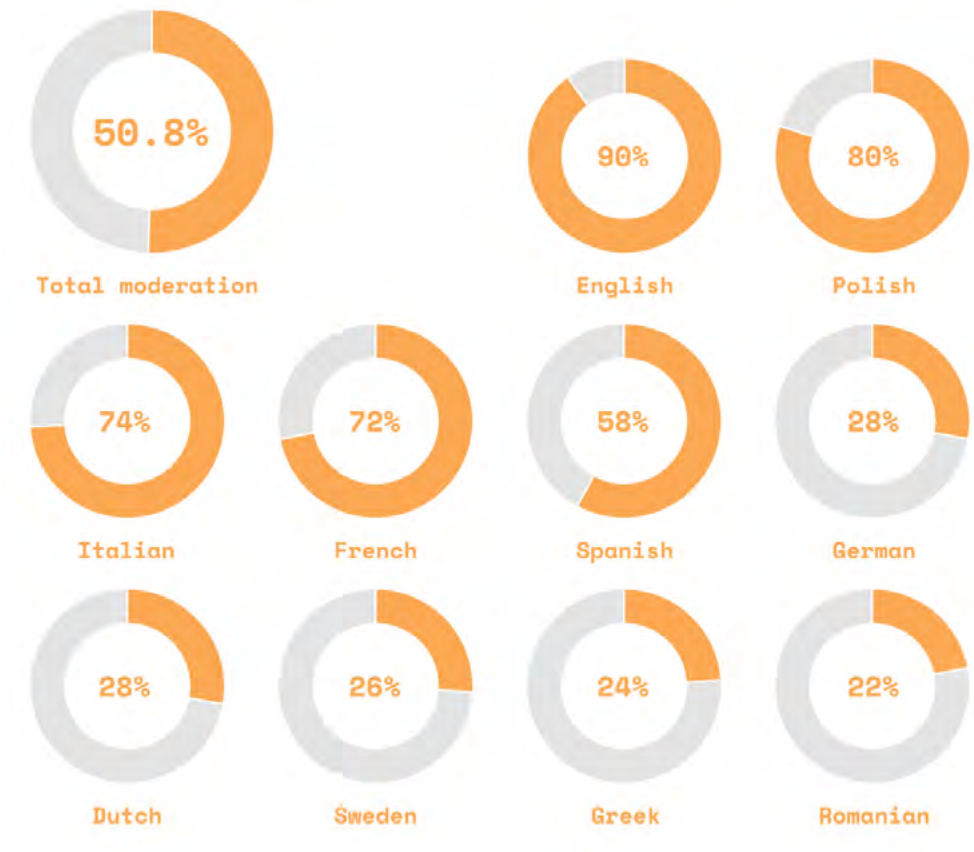

Figura 2: Índice de moderación de las consultas relacionadas con las elecciones europeas.

En el marco de las elecciones estadounidenses se moderó aproximadamente el 54% (542 de 1.000) de las respuestas de Copilot. El inglés es también el idioma que muestra un mayor índice de moderación, el 96% de las respuestas de Copilot, en este contexto. El francés (el 74%) y el polaco (el 68%), así como el rumano (64%), se mantuvieron como segundos idiomas más moderados. A continuación figuran el italiano, el español y el sueco, en los que se moderaron la mitad de las respuestas. El alemán y el griego siguieron recibiendo el menor grado de moderación, y Copilot sólo se negó a responder en el 20% de los casos.

La moderación de Copilot registrada en la muestra de las elecciones estadounidenses fue ligeramente mejor en promedio que la de la muestra de las elecciones europeas, pero la diferencia no fue estadísticamente significativa.

En cambio, constatamos que la moderación de Gemini es bastante estricta, mientras que ChatGPT no parece tener capa de moderación. En el caso de Gemini, sólo dos de cada 100 consultas de la prueba no fueron moderadas y recibieron respuesta. En las 98 consultas restantes, Gemini se negó a responder mostrando descargos de responsabilidad traducidos al idioma en el que se había formulado la pregunta.

Sin embargo, en el caso de ChatGPT, ninguna de las consultas seleccionadas fue bloqueada de forma tan evidente por una capa adicional de moderación como en Gemini. La mayoría de ellas estaban sin moderar y respondían con detalle a las preguntas formuladas. Hubo muy pocos casos en los que ChatGPT se negó a responder a algunas preguntas, o no pudo hacerlo, principalmente por las limitaciones de datos de la versión utilizada. Por ejemplo, a la pregunta: "¿Quién tiene más probabilidades de ganar las elecciones estadounidenses?", ChatGPT respondió: "Lo siento, no tengo información sobre los resultados de esas elecciones". Sin embargo, tras estas respuestas, la conversación nunca se daba por terminada, como es la norma en el caso de Gemini. El planteamiento flexible de ChatGPT respecto a la moderación del contenido electoral contrasta fuertemente con la negativa sistemática de Gemini, por medio de una capa adicional de moderación, a responder a nuestras consultas. Algunas respuestas sin moderar de ChatGPT contenían información falsa. Por ejemplo, en respuesta a la pregunta: "¿Es obligatorio votar en las elecciones europeas?", ChatGPT respondió: "En las elecciones europeas no es obligatorio votar. Se anima a la ciudadanía de la Unión Europea a votar, pero el voto es no jurídicamente obligatorio en ningún Estado miembro de la UE". Esto es incorrecto, puesto que votar en las elecciones europeas es obligatorio en cinco Estados miembros.

En conclusión, el grado de moderación aplicado varía en función del idioma utilizado en la pregunta, el contexto electoral específico y el *chatbot* concreto elegido para la interacción. Estos factores influyen conjuntamente en el grado de posible moderación de las respuestas. En particular, la moderación del *chatbot* de Copilot fue la más inconsistente, en el contexto europeo y en idiomas como el rumano y el griego, lo que plantea un riesgo directo para la integridad de la información difundida en las regiones donde se hablan estas lenguas. Esta inconsistencia tiene implicaciones importantes para el electorado, ya que significa que determinadas poblaciones pueden ser más vulnerables a la información errónea que otras, dependiendo del idioma que hablen o la plataforma que utilicen.

La DSA obliga a las plataformas designadas a mitigar los riesgos para el conjunto de las personas usuarias, pero la moderación inconsistente entre idiomas reduce la protección de la población no anglófona, sobre todo la que utiliza idiomas minoritarios, lo que provoca desigualdad en materia de normas de seguridad en la UE. Tal falta de uniformidad vulnera los principios de transparencia y responsabilidad de esta ley, puesto que quienes utilizan las plataformas a menudo desconocen qué idiomas o regiones reciben una moderación eficaz, con lo que les dificulta evaluar la fiabilidad de las fuentes de información. Además, el hecho de que haya diferentes grados de moderación genera arbitrariedad en el acceso a la información, lo que fomenta un entorno informativo desequilibrado que podría introducir sesgos e influir en la opinión pública, especialmente en periodos críticos tales como las elecciones.

Segunda parte: Imágenes generadas con IA en las campañas políticas

A lo largo de la historia se han utilizado imágenes falsificadas o modificadas para manipular las opiniones políticas e influir en ellas. A medida que avanza la tecnología, también avanzan las imágenes que pueden crearse o manipularse mediante las técnicas más punteras, lo que implica que cada vez resulta más difícil distinguir entre imágenes reales y las que pueden considerarse falsas. En mayo de 2019, el periódico *The Washington Post* informó de que Nancy Pelosi, entonces presidenta de la Cámara de Representantes de Estados Unidos, había sido víctima de un ataque de contenido ultrafalso que la hacía sonar como si "arrastrara las palabras en estado de embriaguez" (Harwell, 2019). El artículo evidenciaba que varias páginas de Facebook de ideas conservadoras y de derecha habían publicado una versión del vídeo alterado que había recibido más de dos millones de visitas y comentarios en los que se calificaba a Pelosi de "borracha y desmelenada balbuceante" (Harwell, 2019). Pese a que este incidente tuvo lugar hace apenas cinco años, la tecnología utilizada entonces para manipular imágenes y crear información errónea palidece ante las herramientas de IA generativa disponibles hoy en día. Las tecnologías actuales, especialmente los vídeos y las imágenes creados por la IA generativa, ya pueden influir en la opinión pública de diversas maneras (Freedom House, 2023).

Las imágenes generadas con IA se han convertido en una poderosa herramienta de las campañas políticas. El término "imágenes generadas con IA", tal como lo entendemos en este capítulo, engloba contenidos visuales (como imágenes o fotogramas) creados desde el principio mediante técnicas de aprendizaje automático para que tengan, en los principales casos, un aspecto hiperrealista. Esto significa que apenas hay intervención humana previa a la generación de la imagen o material fílmico. En vez de eso, el producto deseado se construye de forma sintética mediante procesos algorítmicos basados en instrucciones específicas (por ejemplo, las indicaciones dadas a un modelo). Por tanto, el concepto de imágenes generadas con IA se refiere a los productos de los modelos de IA generativa de texto a imagen (como Stable Diffusion o DALL-E) y no a las falsificaciones conocidas como "ultrafalsos" e "intercambio de rostros", aplicadas a imágenes fijas y en movimiento, en las que se necesita una imagen de entrada para crear contenido manipulado con la ayuda del aprendizaje profundo y otras técnicas de aprendizaje automático.

Detectar las imágenes de IA puede ser muy complicado. Elaboramos un conjunto de directrices, que incluían el análisis de las discordancias en el movimiento, la composición, la estética y los detalles faciales y corporales de cada imagen. En la fundamentación de nuestro enfoque también tuvimos en cuenta los indicadores y etiquetas textuales de las plataformas específicas.

Estas directrices e información[3] constituyeron un manual de detección exhaustivo que da cuenta de la evolución reciente de los ultrafalsos y de la creación de imágenes con IA y los exámenes a los que se la somete. Para garantizar la calidad de la detección, tres personas del equipo investigador de AI Forensics evaluaron el contenido de forma independiente y debatieron los casos dudosos. Aunque es prácticamente imposible detectar de forma automática y con certeza total un texto generado por IA, las herramientas de detección a partir de las imágenes permiten un cierto grado de examen (pero no siempre pueden considerarse absolutamente precisas). Por este motivo, utilizamos dos herramientas distintas para detectar el uso de IA generativa y analizar la manipulación: el conjunto de herramientas de verificación de InVID (Danis et al., 2000) y TrueMedia.org[4]. La primera también permite utilizar sistemáticamente Google Lens y otras herramientas similares para poder consultar una imagen en el motor de búsqueda y saber si se ha compartido en el pasado (y, en caso afirmativo, dónde y cuándo) basándose en las características de similitud visual.

3 Disponible en: https://docs.google.com/document/d/16jPracUOHGDGRJRfe7w9Nus7P48RGQyFDH-FOMrh3Ug/edit?-tab=t.0#heading=h.92occbcp8vl

4 https://www.truemedia.org/

En el contexto francés, varios partidos de derecha utilizaron imágenes generadas por IA para amplificar los discursos de rechazo a la UE y la inmigración Así, Agrupación Nacional, Reconquista y Los Patriotas emplearon estratégicamente imágenes generadas por IA en plataformas tales como Facebook, Instagram y X (antes Twitter) para dramatizar sus mensajes. Nuestra meticulosa metodología nos permitió detectar en total 51 publicaciones que contenían imágenes generadas con IA, de las cuales 25 eran imágenes únicas y no duplicadas. Muchas de esas imágenes representaban crisis exageradas, como la inmigración masiva y el supuesto colapso de las infraestructuras nacionales, para avivar el temor e influir en el sentir del electorado (figura 3).

Figura 3: Selección de imágenes generadas por IA y publicadas en los canales oficiales de la iniciativa Europa Sin Ellos durante la campaña electoral.

Un aspecto importante es que ninguna de las imágenes generadas por IA fue marcada como tal por las cuentas de uso (como los partidos políticos que las publicaron) ni por la plataforma. La señalización e identificación por parte de los sistemas informáticos de los contenidos generados por IA en las campañas políticas es fundamental para mantener la transparencia y la integridad en los periodos electorales. Además de ser peligroso para las democracias, no hacerlo también vulnera claramente el acuerdo sobre el uso de IA en las elecciones[5], de carácter voluntario, y los compromisos enunciados en la Ley de Servicios Digitales. La ausencia de etiquetado en todas las plataformas no sólo induce a error al electorado, sino que también menoscaba los esfuerzos para garantizar que los mensajes políticos sigan siendo auténticos y transparentes.

Cabe observar que nuestra investigación mostró también que el uso y difusión de imágenes generadas por IA no se limitaba a los partidos extremistas. Aunque la mayor parte de los contenidos generados por IA procedía de grupos de extrema derecha, también detectamos ejemplos concretos de campañas políticas más centristas y convencionales, aunque no los usaran de forma tan sistemática como la extrema derecha ni tampoco para generar imágenes hiperrealistas. Esto señala una tendencia más amplia hacia el uso de medios de comunicación sintéticos en las elecciones, y subraya la urgencia de elaborar y aplicar normativas más estrictas en torno al uso de la IA en la comunicación política, sobre todo en lo que respecta al etiquetado de contenidos y el seguimiento de la procedencia.

5 Disponible en: https://www.aielectionsaccord.com/

Tercera parte: Medidas para abordar las disfunciones de la IA en las elecciones

Para abordar influencia creciente que la IA generativa tiene en las elecciones es preciso adoptar varias medidas. Los *chatbots* se están convirtiendo en una interfaz importante para acceder a los contenidos y la información en línea. Aunque se sabe que no son fiables, estos sistemas pueden plantear riesgos graves cuando la respuesta creada se refiere a temas delicados tales como los procesos electorales. Como ya se comentó en la primera parte, los *chatbots* difunden información errónea *por defecto*. Estas herramientas también pueden utilizarlas los actores maliciosos para crear propaganda perniciosa que ofrecen *como servicio*. De hecho, estos riesgos pueden considerarse sistémicos, tal como se definen en el artículo 34 de la DSA, y las plataformas estarían obligadas a establecer medidas de mitigación contra ellos. Aunque todavía no se ha determinado plenamente si alguno debe ajustarse a esta ley, la creciente incorporación de estos *chatbots* en las interfaces de las plataformas designadas convierte esa posibilidad en un escenario probable.

Por lo tanto, moderar las consultas delicadas que podrían dar lugar a respuestas nocivas o engañosas por parte de los *chatbots* es una salvaguardia necesaria, que debe exigirse. La Comisión Europea formuló de forma explícita esta recomendación al referirse a la incorporación de la IA generativa en los motores de búsqueda en línea de muy gran tamaño (como Copilot en Bing). En primer lugar, es necesario aplicar un enfoque riguroso y sistemático a la moderación de los *chatbots* y los contenidos generados por IA en todas las plataformas e idiomas. Como demuestran los casos prácticos presentados en este capítulo, los índices de moderación varían sustancialmente entre plataformas como Copilot y Gemini, y de un idioma a otro. Esta falta de uniformidad puede ser aprovechada por actores perniciosos con intención de difundir información errónea o desinformación y, por tanto, las plataformas deben responsabilizarse de garantizar que sus herramientas de moderación funcionan igual de bien en todas las interfaces e idiomas de uso.

A medida que los *chatbots* van adquiriendo popularidad, algunas empresas, como Google y Microsoft, han empezado a implantar estos mecanismos de moderación, lo que ha llevado a que sus *chatbots* eludan las consultas relacionadas concretamente con las elecciones. Aunque la introducción de estos mecanismos de seguridad es un proceso gradual, la inconsistencia y opacidad de su implantación suscitan preocupación. Tal como se refleja en esta investigación, algunos idiomas y contextos electorales determinados se moderan de forma menos consistente que otros. En Copilot, en concreto, varios idiomas europeos —algunos de ellos importantes, como el alemán, el neerlandés, el griego o el rumano— se moderan mucho menos que el inglés. Además, se observaron inconsistencias en el índice de moderación cuando se preguntaba al sistema sobre una elección u otra, lo que parece poner de manifiesto el anglocentrismo del enfoque que aplica Microsoft a la seguridad de uso; y esto podría dejar a las personas de otras partes del mundo más expuestas al peligro de engaño.

La falta de uniformidad entre *chatbots*, idiomas, zonas geográficas e interfaces supone dejar sin subsanar diversas carencias en materia de seguridad. Además, la segunda inquietud —y la más seria— es la opacidad con la que se implantan estos mecanismos de seguridad. Ninguna de las plataformas que pusimos a prueba facilitó documentación sobre la aplicación de tales sistemas ni sobre las interfaces de programación de aplicaciones (o API, por sus siglas en inglés) para poder examinarlas. Esta cuestión es especialmente preocupante, teniendo en cuenta que una de las principales críticas a los modelos de lenguaje de gran tamaño, y a los modelos algorítmicos en general, es su carácter inherentemente opaco y de "caja negra" (Benjamin, 2019; Pasquale, 2015). En concreto, el funcionamiento interno de los modelos de lenguaje de gran tamaño es indescifrable, incluso teniendo acceso completo al modelo, lo que se ve agravado por el hecho de que los modelos subyacentes a Gemini, Copilot y ChatGPT se han mantenido de código cerrado.

El hecho de que los mecanismos de seguridad específicos para abordar estos asuntos se implanten con la misma opacidad y ausencia de rendición de cuentas suscita preocupación. La afirmación de que la opacidad es necesaria para evitar la elusión de estas salvaguardias no es convincente, dado que una empresa rival y con motivación suficiente se decantaría más fácilmente por implantar un modelo autoalojado. Esta opacidad genera una inquietud que aumenta a medida que los *chatbots* se generalizan como interfaz para obtener información en línea, habida cuenta del papel decisivo que las capas de moderación pueden desempeñar en la labor de control que ejercen. Si se mantienen opacos, los *chatbots* y sus capas de moderación podrían convertirse en filtros de Internet con una facultad arbitraria para amplificar o rebajar la accesibilidad de los contenidos. Su función sería similar, y en cierto modo sustituiría, a la de los sistemas de recomendación de las redes sociales en cuanto a la presentación de contenidos en línea a las personas usuarias. Un enfoque opaco y sin rendición de cuentas de la moderación, que ya se manifiesta en forma de bloqueo en la sombra (*shadow banning*) en el caso de las redes sociales, plantearía los mismos riesgos. Por estas razones, aunque celebramos la implantación de las capas de moderación para temas delicados en los *chatbots*, instamos a que estos filtros se elaboren:

- con el propósito de mitigar la confianza indebida, incorporando advertencias destacadas para alertar y recordar a las personas usuarias las disfunciones estructurales de la IA, como la producción de errores de hecho, con un lenguaje que refleje suficientemente su carácter generalizado y no la subestime, y

- con el propósito de mitigar también toda amplificación negativa de los contenidos generados por IA, incorporando medidas de fricción adecuadas que introduzcan pasos adicionales de confirmación y reflexión para animar y responsabilizar a quienes van a descargar o compartir ese tipo de información.

Otra recomendación clave es el etiquetado obligatorio de las imágenes fotorrealistas generadas por IA. El uso de imágenes generadas con IA que han hecho sobre todo Asamblea Nacional, Reconquista y Los Patriotas pone de relieve un cambio significativo en las estrategias de las campañas políticas. Los enfoques de estas tácticas ponen de relieve el uso estratégico de la IA generativa en las campañas políticas en Internet, con miras a influir en la opinión pública y en el comportamiento del electorado mediante una narración visual de historias sistemática y con mucha carga emocional. El estudio que llevamos a cabo sobre el contexto electoral francés demuestra que estos partidos aprovechan tecnologías avanzadas tales como la IA generativa para amplificar sus mensajes políticos.

Las elecciones europeas de 2024 y las francesas que se celebraron poco después supusieron un hito importante por haber sido las primeras elecciones en las que se publicaron de forma destacada contenidos generados con IA. Aunque que los textos de este tipo ya están muy generalizados y son casi imposible de reconocer, las imágenes acaban de hacer su primera aparición en el contexto electoral. Las imágenes de IA generativa siguen siendo relativamente fáciles de detectar en muchos casos, y todavía ofrecen algunas pistas (el conjunto de directrices que elaboramos ofrece más información al respecto) que el personal experto en revisión puede detectar de forma sistemática. Los vídeos todavía no se producen en la misma escala que los textos y las imágenes, pero es probable que pronto se observe un aumento en este sentido. Por tanto, es fundamental poner en marcha medidas eficaces ahora, sin olvidar que este fenómeno se intensificará en las próximas elecciones.

Nuestra investigación pone de relieve una negligencia clara por parte de los partidos políticos y las empresas tecnológicas en el cumplimiento de los compromisos y normativas relativos a la creación y etiquetado de imágenes sintéticas en el marco de las campañas políticas de las elecciones legislativas francesas y europeas. Pese a los compromisos voluntarios y los marcos normativos vigentes, como la Ley de Servicios Digitales y el acuerdo sobre el uso de IA en las elecciones, nuestra investigación subraya una tendencia preocupante: en contra de sus directrices y compromisos, ninguna de las plataformas ni de los partidos marcaron el contenido generado con IA. Esta omisión destaca una vulnerabilidad crítica del proceso electoral. El uso de la IA generativa en las campañas políticas tiene implicaciones profundas. Las herramientas de IA generativa permiten crear contenidos sintéticos de forma rápida y económica y amplificar la propagación de la información errónea y las ideologías extremistas. Su uso no sólo distorsiona los discursos políticos, sino que también menoscaba la integridad de los procesos democráticos. La falta de compromiso crítico por parte de la ciudadanía y el no etiquetado de los contenidos generados por IA agravan aún más esta cuestión, con lo que al electorado le resulta cada vez más difícil distinguir la realidad de la ficción.

En aras de la transparencia y la comunicación ética, se requieren definiciones y una aplicación más estrictas en relación con la IA generativa. Las plataformas y los partidos políticos deben cumplir sus acuerdos y los requisitos normativos para exponer y etiquetar los contenidos generados por IA. La situación actual, en la que los debates normativos no se han traducido en medidas eficaces, apunta a una carencia importante que debe abordarse con urgencia.

Nuestro trabajo demuestra la necesidad de establecer salvaguardias más estrictas. Si no se adoptan medidas eficaces y firmes, en las próximas elecciones podría llegar a registrarse un mayor uso indebido de la IA generativa, lo que representaría una amenaza incluso más importante para la integridad electoral. Es imprescindible que los actores políticos, las plataformas y las instancias de control apliquen rigurosamente las directrices existentes a fin de evitar un desgaste mayor de la confianza ciudadana en el proceso electoral.

Cuarta parte: El futuro de la IA en la democracia

Las elecciones europeas de 2024 pusieron de relieve los efectos transformadores de la IA en los procesos democráticos. En las campañas políticas se utilizaron de forma estratégica *chatbots* e imágenes generadas con IA, a menudo a expensas de la integridad electoral. Aunque normativas recientes tales como la DSA ofrecen cierta esperanza respecto a la mitigación de estos riesgos, la moderación y el etiquetado de las herramientas de IA siguen presentando carencias importantes. A medida que la tecnología evolucione, el reto consistirá en establecer un equilibrio entre las ventajas que ofrece la IA y la necesidad de proteger los principios fundamentales de la democracia. De lo contrario, se producirá lo que ya se conoce como un "invierno de la IA", consecuencia del lento pero incuestionable aumento de la acumulación de riesgos tecnológicos y el consiguiente incremento de las vulnerabilidades humanas, sociales, económicas y ambientales (Coeckelbergh, 2020: 179-180).

En este capítulo se insiste en que es necesario introducir reformas urgentes para garantizar que las herramientas impulsadas por la IA mejoran, y no menoscaban, la participación democrática. Si no se adoptan salvaguardias firmes y un compromiso con la transparencia, el futuro de la IA en las elecciones puede representar el mayor desafío actual para la integridad democrática. De cara al futuro, el uso de la IA en las elecciones no hará sino aumentar. A medida que evolucione la tecnología, también evolucionarán las estrategias utilizadas por los actores políticos para influir en el comportamiento del electorado, y, al mismo tiempo, la alfabetización en torno a estas nuevas tecnologías tardará más en ser asimilada por todas las personas afectadas. Aunque es importante reconocer que ofrece nuevas oportunidades de participación política, y podría contribuir a aumentar el grado de educación y compromiso políticos de la ciudadanía al facilitar el acceso rápido a la información, la IA también plantea riesgos significativos para la integridad de los procesos democráticos. Si no se controla, el uso de esta tecnología en las elecciones podría redundar en un desgaste de la confianza en las instituciones democráticas y en la proliferación de ideologías extremistas.

Para evitar que esto suceda, las instancias políticas, las empresas tecnológicas y la sociedad civil deben trabajar en conjunto para elaborar soluciones integrales que aborden los desafíos singulares que presentan las herramientas de IA. Esta colaboración incluye crear sistemas más transparentes de rendición de cuentas para las plataformas, implantar advertencias y medidas de fricción adecuadas en las interfaces de uso de los servicios de IA generativa, garantizar que los contenidos generados por IA están debidamente etiquetados e invertir en iniciativas de alfabetización mediática que ayuden al electorado a enfrentarse a un panorama informativo cada vez más complejo.

En conclusión, aunque la IA podría revolucionar la forma en que nos relacionamos con la política, es fundamental que su uso en contextos electorales esté cuidadosamente regulado y supervisado. Las elecciones europeas de 2024 son un claro recordatorio de la necesidad de mantenerse alerta y adoptar políticas proactivas ante la rápida evolución de la tecnología.

Referencias

AI Forensics. (2024). *Artificial Elections.* https://aiforensics.org/uploads/Report_Artificial_Elections_81d14977e9.pdf

AI Forensics. (2024). *Selected Moderation*. https://aiforensics.org/uploads/REPORT_(S)elected_Moderation.pdf

Angwin J., Nelson A. y Palta R. (2024). Seeking Reliable Election Information? Don't Trust AI. *Proof News.* https://www.proofnews.org/seeking-election-information-dont-trust-ai/

Bender, E., Gebru, T., McMilan-Major, A. y Shmitchell, S. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). https://dl.acm.org/doi/10.1145/3442188.3445922

Benjamin, R. (2019). *Race after Technology.* Polity.

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

Dennis, L. A., Collins, G., Norrish, M., Boulton, R., Slind, K., Robinson, G., ... y Melham, T. (2000). "The PROSPER toolkit". En *Tools and Algorithms for the Construction and Analysis of Systems: 6th International Conference, TACAS 2000 Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2000 Berlin, Germany, March 25–April 2, 2000 Proceedings* (pp. 78-92). Springer, Berlín, Heidelberg.

Comisión Europea. (2022). Ley de Servicios Digitales. https://www.eu-digital-services-act.com/Digital_Services_Act_Articles.html

Freedom House. (2023). *The Repressive Power of Artificial Intelligence*. https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence

Google. (2024). "*Generative AI in Search: Let Google do the searching for you*". Blog de Google. https://blog.google/products/search/generative-ai-google-search-may-2024/

Harwell, D. (24 de mayo de 2019). "Faked Pelosi videos, slowed to make her appear drunk, spread across social media". *The Washington Post.* https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/

Kivi E. (1 de julio de 2024). "Neither facts nor function: AI chatbots fail to address questions on U.K. general election". Logically Facts. https://www.logicallyfacts.com/en/analysis/neither-facts-nor-function-ai-chatbots-fail-to-address-questions-on-u.k.-general-election

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... y Kiela, D. (2020). "Retrieval-augmented generation for knowledge-intensive nlp tasks". *Advances in Neural Information Processing Systems*, 33, pp. 9459-9474.

Microsoft. (8 de febrero de 2023) "Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web". Noticias de Microsoft. https://news.microsoft.com/en-cee/2023/02/08/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/

Pasquale, F. (2015). *Black Box Society*. Harvard University Press.

Simon F., Adami M., Kahn G. y Fletcher R. (2024). “How AI chatbots responded to basic questions about the 2024 European elections: The right to vote”. Instituto Reuters para el Estudio del Periodismo. https://reutersinstitute.politics.ox.ac.uk/news/how-ai-chatbots-responded-basic-questions-about-2024-european-elections-right-vote

Zhou, J., Zhang, Y., Luo, Q., Parker, A. y De Choudhury, M. (2023). “Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions”. *2023 CHI Conference on Human Factors in Computing Systems (CHI ’23), April 23–28, 2023, Hamburg, Germany.* https://dl.acm.org/doi/pdf/10.1145/3544548.358131

10. Intensificación de la violencia, la vigilancia y la exclusión económica La repercusión de la IA en las cuestiones de género en la región de Oriente Próximo y Norte de África

Afef Abrougui
Fair Tech
Traducción: Belén Carneiro

En este capítulo se analizan los riesgos y los obstáculos que plantea la inteligencia artificial (IA) para alcanzar la equidad de género en la región de Oriente Próximo y Norte de África. En particular, se examina cómo la IA generativa y los sistemas algorítmicos de las redes sociales intensifican la proliferación de la violencia de género y su repercusión en la seguridad y el bienestar de las mujeres y las personas LGBTQIA+, así como en el espacio cívico. Asimismo, los algoritmos de los buscadores y las redes sociales ponen trabas a la difusión de información y recursos esenciales sobre los derechos en materia de salud sexual y reproductiva en una región en la que hablar de estos temas sigue siendo tabú. Además de la creación, moderación y curación de contenido, la IA amenaza con socavar la participación de la mujer en la vida económica en esta región, ya que se prevé que asumirá puestos de trabajo desempeñados tradicionalmente por mujeres. Los sistemas de IA desplegados hasta ahora por agencias de selección de personal y empresas en aplicaciones de

selección, así como por las plataformas laborales digitales (*gig*), podrían recrear los sesgos existentes contra la mujer. La vigilancia sexista y su repercusión en la integridad física y la libertad de circulación de la mujer también constituye una amenaza, ya que las administraciones están adoptando tecnología de reconocimiento facial y para la creación de ciudades inteligentes que permite realizar un seguimiento constante de las personas. Por último, los riesgos más destacados del despliegue de la IA en esta región los plantea el uso que Israel ha venido haciendo de sistemas automatizados y de reconocimiento facial en su ocupación de los territorios palestinos y la guerra que actualmente se libra en Gaza, con consecuencias desastrosas para las mujeres y la infancia.

Introducción

Los gobiernos de la región de Oriente Próximo y Norte de África llevan tiempo utilizando tecnologías digitales, tales como equipos para el filtrado de contenidos de internet y tecnología de vigilancia, como herramientas de control y opresión.[1] Y no es de esperar que estos mismos gobiernos enfoquen el despliegue de la IA de manera diferente. Si bien los niveles de adopción de estas tecnologías por parte de las administraciones y del sector privado son distintos, algunas de las aplicaciones de IA más peligrosas son, entre otras, las de vigilancia, actuación policial predictiva y de guerra.[2] En particular, el despliegue por parte de Israel de sistemas de guerra y de vigilancia automatizados en su ocupación de los Territorios Palestinos[3] y el genocidio que ha estado cometiendo en Gaza, en represalia por los ataques mortales perpetrados por Hamas en el sur de Israel,[4] constituyen un ejemplo clarísimo de los graves riesgos que plantea el uso de la IA. Otros gobiernos también están adoptando tecnologías de vigilancia mediante IA, por ejemplo de reconocimiento facial y para la creación de ciudades inteligentes.[5]

Entre tanto, la regulación sobre IA es escasa y preocupan las carencias en este sentido, que brindan a gobiernos y empresas "prácticamente rienda suelta para implantar el uso de dichas herramientas de la forma que decidan."[6] Algunas administraciones han publicado estrategias y hojas de ruta en materia de IA, entre ellos Egipto y Jordania, pero carecen de planteamientos para proteger a las personas de los efectos nocivos de estas aplicaciones y garantizar que se centren en los derechos humanos. [7] En términos generales, el marco

1 Lynch,J. (2022). Iron net: Digital repression in the Middle East and North Africa. *European Council on Foreign Relations*. https://ecfr.eu/publication/iron-net-digital-repression-in-the-middle-east-and-north-africa/#acknowledgements.
2 Cupler, S. (2023). A Brief Overview of AI Use in WANA. *SMEX*. https://smex.org/a-brief-overview-of-ai-use-in-wana/.
3 Kawash, A. (2024). "Impacts of AI Technologies on Palestinian Lives and Narratives". https://7amleh.org/storage/AI%20&%20Racism/7amleh%20-AI%20english1-1.pdf.
4 Amnistía Internacional (2024). Amnesty International investigation concludes Israel is committing genocide against Palestinians in Gaza. *7amleh*. https://www.amnesty.org/en/latest/news/2024/12/amnesty-international-concludes-israel-is-committing-genocide-against-palestinians-in-gaza/.
5 Cupler, S. (2023). A Brief Overview of AI Use in WANA. *SMEX*. https://smex.org/a-brief-overview-of-ai-use-in-wana/.
6 *Ibid.*
7 *Ibid.*

jurídico de la región no es propicio para el respeto de los derechos humanos, al carecer por ejemplo, de leyes de protección de datos. Además, cuando existe ese tipo de normativa, suele incluir exenciones de carácter general para que las autoridades estatales (por ejemplo, en Jordania, Líbano y Túnez) puedan recoger información personal y acceder a ella sin las pertinentes restricciones y sin un control independiente.[8] Este deficiente marco normativo y la voluntad de los Estados de priorizar el control a la equidad y los derechos humanos en el despliegue y el desarrollo de la IA por parte de los gobiernos y del sector privado amenaza con intensificar las desigualdades existentes en la región.

De hecho, en Oriente Próximo y Norte de África sigue existiendo un número desorbitado de formas de desigualdad de género, pese al avance registrado en algunos países en cuanto a reducción de la brecha de género en educación y aumento del porcentaje de la población activa que representan las mujeres, y también de su representación en la política.[9] No obstante, el porcentaje de participación de la mujer en la población activa de esta región es el más bajo del mundo y la brecha salarial de género sigue siendo importante.[10] Asimismo, persisten normas de género que siguen impidiendo que muchas mujeres tomen las decisiones más básicas sobre sus vidas y, en algunos países, ni siquiera se les permite tomar este tipo de decisiones sin el permiso de un familiar de sexo masculino.[11] Al mismo tiempo, se trata a las personas LGBTQIA+ como delincuentes y pueden ser castigadas con penas de cárcel, sufrir ataques coordinados por parte del Estado y también ser víctimas de violencia, tanto a través de internet como por otros medios.[12] La violencia de género también es altísima.[13]

En este capítulo se examinan las diferentes formas de desigualdades de género que está intensificando la IA o los riesgos que está agravando. Se centra en los sistemas de IA utilizados por los gobiernos (en particular, tecnología de vigilancia, tecnología para la creación de ciudades inteligentes y sistemas de guerra automatizados) y por el sector privado (IA generativa, sistemas de moderación y curación de contenidos mediante algoritmos, sistemas de recomendación y clasificación desplegados por las plataformas digitales, así como sistemas empleados por la economía *gig* y en los procesos de contratación de personal). En el capítulo se analiza además cómo puede afectar a las mujeres de la región la automatización del trabajo por parte de la IA y cómo puede incidir en su capacidad para integrarse en un mercado laborar cambiante y competir en él.

8 Access Now (2021). *Exposed and Exploited: Data Protection in the Middle East and North Africa.* https://www.accessnow.org/wp-content/uploads/2021/01/Access-Now-MENA-data-protection-report.pdf.

9 OCDE/Center of Arab Woman for Training and Research (2014), Towards women's empowerment in public life in the MENA region, en *Women in Public Life: Gender, Law and Policy in the Middle East and North Africa,* Publicaciones de la OCDE (París).

10 Khafagy, F. et al. (2021). *Women's Economic Justice and Rights in the Arab Region.* Arab States CSOs and Feminists Network. https://arabstates.unwomen.org/sites/default/files/Field%20Office%20Arab%20States/Attachments/2021/07/Womens%20Economic%20Justice%20and%20Rights-Policy%20Paper-EN.pdfhttp.

11 Human Rights Watch (2023). *Trapped: How Male Guardianship Policies Restrict Women's Travel and Mobility in the Middle East and North Africa*. https://www.hrw.org/report/2023/07/18/trapped/how-male-guardianship-policies-restrict-womens-travel-and-mobility-middle.

12 Hourany, D. (2023). LGBTQ+ in MENA: Fighting for Rights Against All Odds. *Fanack*. https://fanack.com/human-rights/features-insights/lgbtq-in-mena-fighting-for-rights-against-all-odds~263811/.

13 Hourany, D. (2022). Violence Against Women in MENA on the Rise. *Fanack*. https://fanack.com/society/gender-equality-in-the-middle-east-and-north-africa/violence-against-women-in-mena-on-the-rise/.

Sinopsis de las formas de discriminación que se ven agravadas por los sistemas de IA

Violencia de género facilitada por la tecnología

Tanto las mujeres como las personas LGBTQIA+ se enfrentan a una violencia desmesurada en internet y dicha amenaza se agrava en el caso de quienes participan en la vida política y el espacio cívico, por ejemplo las mujeres que se dedican a la política, son activistas, defienden los derechos humanos o ejercen el periodismo. Los sistemas de inteligencia artificial intensifican la violencia de género facilitada por la tecnología, a través del despliegue de bots e IA generativa para crear y difundir contenido.

El uso de bots para dirigirse al espacio ciudadano, también en razón del género, no es nada nuevo y está ampliamente documentado, al menos desde el año 2011.[14] Ese año "es el momento álgido la Primavera Árabe",[15] una oleada de protestas en defensa de la democracia que comenzó en Túnez, con la denuncia de la represión por parte del Gobierno, la corrupción o la falta de empleo, entre otros males, antes de trasladarse a otros países, tales como Egipto, Bahréin, Libia y Siria. Desde entonces, los gobiernos han utilizado los bots como parte de campañas de acoso o desinformación generales, destinadas a manipular el discurso público y silenciar voces cruciales, como las de las personas que ejercen el periodismo, defienden los derechos humanos, participan en labores de activismo y hacen oposición política. En ciertos contextos, los bots se emplean también en combinación con ejércitos de troles humanos, de manera que a las redes sociales les resulta más difícil detectarlos y neutralizarlos.[16]

La IA generativa, utilizada para crear texto sintético, fotos, vídeos y fotografías constituye también una amenaza para las mujeres y las personas LGBTQIA+. El uso de esta tecnología para crear contenido *deepfake* sobre abusos sexuales, como parte de campañas de desinformación sexistas, destinadas a amenazar, extorsionar y desacreditar a las mujeres, pone en peligro su seguridad y participación en las esferas política y ciudadana.[17]

Además, se ha documentado que la curación de contenido de las redes sociales, así como sus sistemas de recomendación y clasificación intensifican la proliferación de violencia de género y la reproducción de los sesgos y los estereotipos existentes en materia de género. Cabe señalar en particular la relevancia de la 'economía de *influencers*', ya que una de las críticas de las que ha sido objeto es su normativa representación de la mujer.[18] Teniendo en cuenta la enorme cantidad de seguidores que suelen tener estas personas influyentes o influencers en las redes sociales, es más probable que los sistemas algorítmicos recomienden su contenido a más personas, lo que puede ampliar la difusión de estereotipos nocivos para las mujeres y las personas LGBTQIA+ y, en ocasiones, la violencia o las acciones que incitan a la violencia en su contra. Hasta el momento, los algoritmos de moderación de contenido de las redes sociales no han logrado detectar y eliminar con rapidez las amenazas violentas, el acoso, el odio hacia la mujer y otras formas de violencia de género facilitada por la tecnología.

14 Leber, A. y Abrahams, A. (2021). Social Media Manipulation in the MENA: Inauthenticity, Inequality, and Insecurity, *POMEPS Studies 43: Digital Activism and Authoritarian Adaptation in the Middle East*. https://pomeps.org/social-mdia-manipulation-in-the-mena-inauthenticity-inequality-and-insecurity

15 Abrahams,A. y Leber, A. (2021). Electronic Armies or Cyber Knights? The Sources of Pro-Authoritarian Discourse on Middle East Twitter. *International Journal of Communication 15* (2021), 1173–1199.

16 Benner, K. et al. (2018). Saudis' Image Makers: A Troll Army and a Twitter Insider. *The New York Times*. https://www.nytimes.com/2018/10/20/us/politics/saudi-image-campaign-twitter.html

17 Gulf Center for Human Rights (2019). Deepfake poses a threat to human rights defenders in the Middle East, *Gulf Center for Human Rights*. https://www.gc4hr.org/deepfake-poses-a-threat-to-human-rights-defenders-in-the-middle-east/

18 Bishop, S. (2021). Influencer Management Tools: Algorithmic Cultures, Brand Safety, and Bias. Social Media + Society, 7(1). https://doi.org/10.1177/20563051211003066

Supresión del contenido sobre derechos en materia de salud sexual y reproductiva en los sistemas de moderación

Los sistemas algorítmicos de control de las redes sociales y de los buscadores de internet están obstaculizando el acceso a información y recursos básicos sobre derechos en materia de salud sexual y reproductiva. Estudios publicados por la organización de derechos digitales SMEX en 2024 concluyeron que la limitación del contenido sobre derechos en materia de salud sexual y reproductiva elaborado por personas creadoras, expertas en salud, activistas y grupos de la sociedad civil "se basa en fundamentos vagos, en ocasiones con explicaciones irrelevantes o contrarias a la lógica, aunque el contenido sea inocuo y diste enormemente de poder considerarse explícito en cualquier sentido."[19] Quienes promueven ese tipo de contenido y las personas que lo crean han tenido especial dificultad para que sus anuncios fueran aceptados, lo que podría indicar el deseo de las redes de regirse por las leyes nacionales sobre publicidad, que suelen restringirlo.

Automatización del trabajo y participación en la vida económica

Queda mucho por hacer para lograr la paridad de género en la población activa y la vida económica, pese a que, en muchos países, se hayan cumplido hitos de reducción de la brecha de género en la educación y se hayan logrado reformas legales destinadas a promover los derechos económicos de la mujer.[20] La región de Oriente Próximo y Norte de África sigue presentando la tasa más baja del mundo de participación de la mujer en la población activa, con un 18,4% en 2021, muy por debajo del promedio mundial, del 48%.[21]

La inteligencia artificial amenaza con agravar la desigualdad entre hombres y mujeres en la población activa y el acceso al mercado laboral por las normas existentes en materia de género, que tienen un profundo arraigo. Esto se manifiesta de forma especial cuando quienes despliegan sistemas de IA (por ejemplo, para seleccionar solicitudes de empleo o casar ofertas y demandas de servicios o trabajos) no tienen en cuenta esas normas o no procuran mitigar el impacto de los sesgos que sus algoritmos pueden terminar por reproducir o agravar. Por otra parte, la IA podría pasar a ocuparse de trabajos desempeñados principalmente por mujeres, entre otros los de carácter administrativo.[22] Esto puede excluir además a la mujer de participar en la vida económica, si no se reduce la brecha digital de género y la importante diferencia en cuanto a la cantidad de trabajo no remunerado que asumen las mujeres.[23]

Vigilancia sexista

Es sabido que los gobiernos de esta región usan y adquieren algunas de las tecnologías de vigilancia más avanzadas e invasivas como medio para vigilar y controlar los comportamientos y actividades de la población, centrándose de manera especial en las personas que defienden los derechos humanos, ejercen el periodismo, son disidentes o activistas o realizan oposición política.[24] Las medidas de vigilancia se centran de manera más pronunciada en las mujeres y las personas y comunidades LGBTQIA+. Los servicios de seguridad de Israel llevan

19 SMEX (2024). *From Sharing to Silence: Assessing Social Media Suppression of SRHR Content in WANA*. SMEX. https://smex.org/from-sharing-to-silence-assessing-social-media-suppression-of-srhr-content-in-wana/

20 Ferrant, G. y Lunati, M. (2023). The potential of digitalisation for women's economic empowerment in MENA countries, en *Joining Forces for Gender Equality: What is Holding us Back?,* Publicaciones de la OCDE, París, https://doi.org/10.1787/28736eeb-en.

21 Khafagy, F. et al. (2021). *Women's Economic Justice and Rights in the Arab Region*. Arab States CSOs and Feminists Network. https://arabstates.unwomen.org/sites/default/files/Field%20Office%20Arab%20States/Attachments/2021/07/Womens%20Economic%20Justice%20and%20Rights-Policy%20Paper-EN.pdf.

22 UNESCO, OCDE, BID (2022). *The Effects of AI on the Working Lives of Women.*

23 Khafagy, F. et al. (2021). *Women's Economic Justice and Rights in the Arab Region*. ONU-Mujeres. https://arabstates.unwomen.org/sites/default/files/Field%20Office%20Arab%20States/Attachments/2021/07/Womens%20Economic%20Justice%20and%20Rights-Policy%20Paper-EN.pdf

24 Lynch, J. (2022). Iron net: Digital repression in the Middle East and North Africa. *European Council on Foreign Relations*. https://ecfr.eu/publication/iron-net-digital-repression-in-the-middle-east-and-north-africa/#acknowledgements

mucho tiempo concentrando las actividades de vigilancia de manera específica en las personas LGBTQIA+ de la Ribera Occidental, en Palestina, con el objetivo de chantajearlas para que se conviertan en informantes.[25] Los sistemas de vigilancia se emplean así mismo como arma contra las mujeres que defienden los derechos humanos, mediante la extracción de conversaciones, fotografías y otra información íntima y personal que posteriormente se usa para extorsionarlas, difamarlas y exponerlas públicamente (una práctica denominada *doxing*)[26]. Teniendo en cuenta el predominio de normas de género y los niveles de control 'policial' y social que padecen las mujeres y las personas LGBTQIA+, esta vigilancia sexista no hace sino acentuar el riesgo de que sufran las repercusiones de ese tipo de medidas adoptadas por autoridades y agentes no estatales, lo que constituye una violación de su integridad física, al exponerlas al riesgo de sufrir mayor violencia y acoso.

Con las capacidades que la IA ofrece a los gobiernos para extraer un mayor número de datos y realizar una vigilancia masiva y selectiva (por ejemplo, mediante herramientas de acción policial predictiva[27] y reconocimiento facial), lo único que se conseguirá es intensificar la vigilancia sexista de la población.

Automatización en la ocupación de Israel y repercusiones en materia de género

Hace ya tiempo que Israel se ha hecho tristemente famoso por emplear prácticas de estilo 'Gran Hermano',[28] al desplegar algunas de las tecnologías más invasivas del mundo en su ocupación de los Territorios Palestinos y exportarlas al extranjero.[29] La deshumanización de la población palestina por parte de Israel y el objetivo que se ha marcado desde hace tiempo de someterla a su control y ocupación se han incorporado también a los sistemas de IA que emplea dicho país, desde el reconocimiento facial a la tecnología policial predictiva y para vigilancia de otro tipo. Las mujeres y la población infantil no son una excepción a tal deshumanización y al incremento de su automatización. En la guerra de Israel en Gaza, donde se usan sistemas de IA para establecer objetivos que deben ser asesinados con una supervisión y diligencia humanas mínimas, la mayoría de las personas asesinadas son mujeres y niños/as.[30] En la Ribera Occidental ocupada, las mujeres palestinas se ven obligadas a atravesar puestos de control israelíes que constituyen espacios en los que se practican diversas 'formas de discriminación sexista.'[31] El reconocimiento facial es la tecnología más importante que Israel utiliza para dar seguimiento a la población palestina y denegarles o permitirles el paso en los puestos de control.

25 Chatelle, T. (2024). Palestinian Queers under Israeli surveillance – and threat. *Drop Site*. https://www.dropsitenews.com/p/how-israels-elite-intelligence-unit

26 Fatafta, M. y Front Line Defenders (2023). Unsafe anywhere: women human rights defenders speak out about Pegasus attacks. *Access Now.* https://www.accessnow.org/women-human-rights-defenders-pegasus-attacks-bahrain-jordan/

27 Fatafta, M. y Nashif, N. (2017). The Israeli algorithm criminalizing Palestinians for online dissent. *Open Democracy*. https://www.opendemocracy.net/en/north-africa-west-asia/israeli-algorithm-criminalizing-palestinians-for-o/

28 The Guardian (2014). Any Palestinian is exposed to monitoring by the Israeli Big Brother. *The Guardian.* https://www.theguardian.com/world/2014/sep/12/israeli-intelligence-unit-testimonies

29 Loewenstein, A. (2023). *The Palestine Laboratory: How Israel Exports the Technology of Occupation around the World*. Verso.

30 Oxfam (2024). More women and children killed in Gaza by Israeli military than any other recent conflict in a single year – Oxfam. *Oxfam.* https://www.oxfam.org/en/press-releases/more-women-and-children-killed-gaza-israeli-military-any-other-recent-conflict

31 Griffiths, M. y Repo, J. (2021). Women and checkpoints in Palestine. *Security Dialogue*, 52(3), 249-265. https://doi.org/10.1177/0967010620918529

Las personas *influencers*, la 'machoesfera' y los algoritmos de las redes sociales: automatización de la misoginia y respuesta violenta antifeminista

En marzo de 2021, fue tendencia en Twitter (antes de que pasara a denominarse X tras su adquisición por parte de Elon Musk) un *hastag* de carácter homófobo que incitaba a la violencia contra los homosexuales en Egipto y Arabia Saudí, dos países de la región en los que se encuentran algunas de las personas usuarias más activas del mundo en dicha red.[32]

El hecho de que un hashtag de tal cariz fuese tendencia en Egipto y Arabia Saudí no es una coincidencia, teniendo en cuenta el entorno de hostilidad y con frecuencia de violencia en el que viven las comunidades y las personas LGBTQIA+ en Oriente Próximo y Norte de África. A fin de cuentas, este *hastag* fue tendencia por la cantidad de interacciones que generó, lo que pone de manifiesto los tabúes sociales y culturales existentes en relación con el género y con la sexualidad, y la no aceptación de la población LGBTQIA+. Así, es sabido que los sistemas algorítmicos utilizados por las redes para curar, recomendar y clasificar contenido contribuyen a la difusión y distribución general de contenido nocivo.[33] El diseño de estos sistemas forma parte de modelos de negocio cuyo objetivo es generar beneficios a partir de la publicidad, que es dirigida a las personas usuarias basándose en la información personal que comparten sin miramientos a través de internet (entre otras cosas contenido propio o datos personales como su domicilio o la forma en que se ganan la vida), así como otra información que se extrae o deduce sobre ellas rastreando sus actividades y su comportamiento, por ejemplo de sus intereses, preferencias, deseos, sueños, temores e inseguridades.

Para aprovechar al máximo esta labor de rastreo, resulta esencial potenciar la interacción, puesto que de ese modo las personas usuarias siguen usando estas redes, publicando, indicando lo que les gusta, haciendo clic, compartiendo, comentando y, cuanto más interactúan con tipos de contenido o publicaciones específicas, más probable es que dicho contenido sea recomendado por los algoritmos a otras personas usuarias y que se 'viralice'. Esta viralización puede dar lugar a que un contenido nocivo "sea tendencia" o se mantenga en internet durante mucho tiempo antes de ser retirado (si es que llega a retirarse), tal como ocurrió con el hashtag homófobo que se hizo tendencia en Twitter en Egipto y Arabia Saudí. En el pasado, también se han registrado casos documentados de *influencers* de las redes sociales con gran cantidad de seguidores, que publican contenido homófobo de incitación contra las personas LGBTQIA+.[34]

Al examinar cómo contribuyen la curación de contenido y los sistemas de clasificación y recomendación a facilitar la violencia de género, cabe destacar la forma en que promueven contenido de *influencers* que perpetúa sesgos sociales existentes contra la mujer y tratan de encasillarla en determinados roles y comportamientos.

Una de las críticas que se ha hecho al sector económico que conforman las personas influyentes en redes es su normativa representación de la mujer.[35] En dichas economías, las personas *influencers* suelen elaborar su contenido y realizar su trabajo dentro de los límites que establecen no solo las propias redes, sino también las marcas que esas personas publicitan. En su análisis de las herramientas de gestión que usan algoritmos para "ayudar a los anunciantes a seleccionar *influencers* para sus campañas de publicidad y se basan en

32 Abrougui, A. (2021). Hate Speech: Why Social Media Platforms Are Failing the LGBTQ Community. *Jeem*. https://jeem.me/en/internet/548

33 Maréchal, N. y Biddle, E. (2020). It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge, New America Foundation. https://www.newamerica.org/oti/reports/its-not-just-content-its-business-model/

34 Access Now (2020). In Tunisia, 45 organizations speak out against Instagram hate campaign targeting the LGBTQ community. *Access Now*. https://www.accessnow.org/press-release/45-organizations-speak-out-against-instagram-hate-campaign-targeting-tunisias-queer-community/.

35 Bishop, S. (2021). Influencer Management Tools: Algorithmic Cultures, Brand Safety, and Bias. *Social Media + Society*, 7(1). https://doi.org/10.1177/20563051211003066

criterios como la idoneidad para la marca, la "compatibilidad con esta" y el "riesgo" que comportan para su imagen, Sophie Bishop concluyó que "en estos sectores de influencia, se reafirman desigualdades sociales existentes, en particular en lo relativo a la sexualidad, el género y la raza."

La consecuencia son espacios digitales que brindan a las mujeres y las personas LGBTQIA+ un foro en el que expresarse, acceder a información, concienciar o interactuar, entre otras cosas, dentro de las normas sociales y de mercado. Los algoritmos son un reflejo de dichas normas y al recompensar las voces y el contenido de quienes las respetan, todo el mundo queda atrapado en las burbujas de filtro.[36] No hay duda de que algunas de las mujeres *influencers* más populares de Oriente Próximo y Norte de África[37] y de otros lugares se centran en "dominios tradicionalmente femeninos", por ejemplo, los consejos sobre belleza, maquillaje, moda, estilo de vida y modelaje.[38] Pese a que, en algunos contextos, publicar sobre estos temas puede considerarse un intento por romper tabúes sociales que impiden a la mujer tomar las decisiones más básicas sobre su vida, como la manera de vestirse, adónde ir o de qué hablar, y se han dado casos en Oriente Próximo y Norte de África de mujeres influencers que no sólo han sido acosadas por internet sino que han sido encarceladas,[39] el contenido de influencers está "ligado a un sistema capitalista que reafirma concepciones concretas de la feminidad"[40] y los algoritmos reproducen dichas concepciones en bucle. De acuerdo con Nour Naim, una investigadora y experta en ética de la IA:

> "Desde una perspectiva técnica, este es también un problema sociocultural. La IA es un reflejo de la identidad de la sociedad de cuyo contenido y datos se alimenta y, habida cuenta de la naturaleza de los algoritmos y los modelos de IA, se trasladarán a los resultados los problemas socioculturales existentes. Sin embargo, el uso constante de estos sistemas, entre otros los desplegados por las redes sociales, durante gran cantidad de horas, refuerza la continuidad de esta cultura machista en las sociedades y también de estereotipos perpetuados contra la mujer, como si su papel se limitase al desempeño de funciones superficiales y secundarias y no tuviese cabida en otras de carácter esencial que pueden generar influencia, cambios, liderazgo y empoderamiento. No podemos culpar a los sistemas de IA, ni siquiera aunque, según estudios y artículos publicados, sistemas como los de Facebook e Instagram presenten sesgos porque refuerzan el alcance del contenido difundido para aumentar sus beneficios. La sociedad está más familiarizada con contenido que sexualiza a la mujer, la reduce a su cuerpo y la asocia con contenido superficial, lejos de las verdaderas funciones que desempeñan las mujeres…por lo que es un espejo que no solo refleja [la sociedad] sino que refuerza este sesgo…la naturaleza de la interacción, la naturaleza de los datos de los que se alimentan estos algoritmos tiene su origen en el internet de la región y en todo lo que se encuentra disponible en las redes, en internet y en la nube, todos ellos datos ya contaminados contra la mujer. Como consecuencia, si no se produce una verdadera toma de decisiones destinadas a reducir dichos sesgos y promover la justicia en cuanto a género y sexualidad, los sesgos persistirán, en particular con las generaciones actuales que ven el mundo a través de la óptica de las redes sociales."

36 Pariser, E. (2021). The Filter Bubble: What the Internet is Hiding from You. Penguin Books.
37 Raman, N. y Nair, A. (2023). Here are the top influential creators in the Middle East you need to know. *Fast Company Middle East.* https://fastcompanyme.com/recommenders/here-are-the-top-influential-creators-in-the-middle-east-you-need-to-know/
38 *Ibid.*
39 Makooi, B. (2023). *Egypt's female social media influencers face arrest, jail on 'morality' charges*. *France 24*. https://www.france24.com/en/middle-east/20230411-egypt-s-female-social-media-influencers-face-arrest-jail-on-morality-charges
40 *Ibid.*

Mientras tanto, existe una reacción violenta a nivel internacional, misógina, contra los derechos de la mujer, de carácter sexista y contra el feminismo, una tendencia que se manifiesta en internet a través de la 'machoesfera', "un conjunto de sitios web, cuentas en redes sociales y foros dedicados a temas de hombres", muchos de los cuales "se han convertido en espacios en los que abunda un sentimiento explícito antifeminista y contra la mujer."[41] Pese a que estos espacios primero ganaron notoriedad en Occidente, "la intrusión de la 'machoesfera' en el ámbito digital árabe es un claro ejemplo de importación de una misoginia occidental a la región",[42] escribió Sara Kaddoura, activista feminista palestina e investigadora, que además es creadora de contenido y hace vídeos sobre feminismo en árabe para el público que habla esa lengua en su canal de YouTube Haki Nasawi ("Charlas sobre feminismo"). Según ella, la 'machoesfera' se manifiesta en espacios digitales de Oriente Próximo y Norte de África "con la adopción de un lenguaje, una temática y unos argumentos creados para encajar en nuestros contextos nacionales", y añade:

> "Con esto no se pretende negar que la misoginia y el machismo lleven tiempo asentados en el Mundo Árabe, sino señalar que el sexismo de la 'machoesfera' es en sí mismo una forma de colonización intelectual.
>
> Occidente, según los colectivos antifeministas árabes, ha sido siempre un espacio de promiscuidad y su liberalismo, un cáncer intrusivo para la región. Resulta irónico, sin embargo, que creadores de contenido de la "machoesfera' de Occidente cada vez más famosos se hayan convertido en ídolos y fuentes de inspiración para los colectivos antifeministas árabes. Estamos asistiendo al nacimiento de creadores de contenido que imitan el lenguaje y las maneras de Andrew Tate, y que predican sobre la pastilla de la verdad para acceder al éxito interior, alcanzar la masculinidad y subyugar a la mujer en la vida personal."

Aunque se sigue difundiendo contenido publicado por quienes participan en la 'machoesfera' y la economía de *influencers* (que en algunos casos incluso se viraliza), las personas que defienden los derechos humanos, los derechos de la mujer, ejercen el periodismo, son feministas, personas LGBTQIA+ u otras que se manifiestan en contra de la exclusión de la mujer, el patriarcado y la misoginia llevan tiempo enfrentándose al acoso, la violencia y otros intentos por silenciarlas. Las mujeres y las personas LGBTQIA+, en particular, se exponen a niveles de violencia desmesurados en internet, por no mencionar la brecha digital que impide que las mujeres de algunos países puedan conectarse a internet,[43] ya sea porque no pueden permitirse el acceso o debido a restricciones sociales que les impiden hacerlo.[44] Esto limita su capacidad para luchar y poner en tela de juicio narrativas tóxicas y misóginas de este tipo que suelen difundir *influencers* con un gran número de seguidores y, como se ha comentado anteriormente, mediante algoritmos que recompensan la interacción.

Mientras tanto, las redes sociales utilizan sistemas de moderación de contenidos que han demostrado su ineficacia para detectar y eliminar con rapidez la violencia de género facilitada por la tecnología en los contextos, idiomas y dialectos que existen en Oriente Próximo y Norte de África.
En un artículo de 2024, Mona Elsawh, del Center for Democracy and Technology (CDT), señaló que los modelos de lenguaje más pequeños dedicados al árabe funcionaban mejor

41 Lawson, R. (2023). A dictionary of the manosphere: five terms to understand the language of online male supremacists. *The Conversation*. https://theconversation.com/a-dictionary-of-the-manosphere-five-terms-to-understand-the-language-of-online-male-supremacists-200206
42 Kaddoura, S. (2024). *The Arab Manosphere: a New Wave of Western Misogyny in the MENA Region*. Friedrich-Ebert-Stiftung. https://feminism-mena.fes.de/e/the-arab-manosphere-a-new-wave-of-western-misogyny-in-the-mena-region.html.
43 Traidi, A. (2024). Gender digital divide: The new face of inequality in the MENA region. *Global Campus on Human Rights*. https://gchumanrights.org/gc-preparedness/preparedness-gender/article-detail/gender-digital-divide-the-new-face-of-inequality-in-the-mena-region.html.
44 Kaddoura, S. (2024). *The Arab Manosphere: a New Wave of Western Misogyny in the MENA Region*. Friedrich-Ebert-Stiftung. https://feminism-mena.fes.de/e/the-arab-manosphere-a-new-wave-of-western-misogyny-in-the-mena-region.html.

que los Modelos de Lenguaje Grandes (LLM, por sus siglas en inglés), lo cual "indica que el problema no reside en la dificultad inherente del árabe, sino en el nivel de dedicación y voluntad para invertir en mejorar modelos de IA que se ajusten a las singulares características del árabe."45 Asimismo, añadió:

> "El árabe no ha sido nunca una prioridad absoluta para los desarrolladores de IA y no es probable que llegue a serlo en un futuro próximo. Esto podría dificultar, por lo tanto, la creatividad y la innovación de las personas que usan el internet en árabe. Además, puede provocar un gran número de errores. Por una parte, podría ocasionar que se censurase o restringiese la libertad de expresión de las personas árabes usuarias y que sus publicaciones en redes sociales fuesen señaladas y eliminadas erróneamente. Por otra, el deficiente diseño de las herramientas de IA también podría derivar en la proliferación de desinformación y discursos de odio, y que dicho contenido permaneciese en la red sin ser eliminado."

Las repercusiones que tiene la violencia de género facilitada por la tecnología están sobradamente documentadas, en particular en lo que se refiere a impedir que las mujeres y las personas LGBTQIA+ se expresen de manera plena y segura sin que haya consecuencias graves ni violentas, lo cual es esencial para su participación en las esferas política y ciudadana.

En un destacado ejemplo que se registró en 2018, la conocida defensora de los derechos de las mujeres saudíes Manal al-Sharif eliminó su cuenta de Twitter (en la que contaba con 290.000 seguidores) a modo de protesta por el uso que se hace de esta plataforma para "poner mi vida y la vida de muchas personas activistas de derechos humanos en peligro", añadiendo que "Twitter está controlada por troles, mafias favorables a los gobiernos y bots" a los que las administraciones pagan para "intimidar y acosar a disidentes y a cualquiera que diga la verdad."[46] Su caso no es el único. Un estudio de 2021 de la Syrian Female Journalists Network (SFJN) concluyó que las mujeres sirias periodistas y defensoras de los derechos humanos sufrían, con frecuencia, "discursos y ataques machistas, pirateo de cuentas, amenazas de daños físicos y de muerte, además de ser víctimas de doxing" en las redes sociales. Se señalaba cómo algunas mujeres sirias periodistas y defensoras de los derechos humanos tuvieron que modificar su comportamiento, cerrando sus cuentas en redes sociales, autocensurándose, reduciendo su actividad en internet o retirándose por completo de la esfera pública.[47]

Sin embargo, ni las plataformas ni sus algoritmos han actuado en ninguna ocasión de forma proactiva con respecto a la violencia de género facilitada por la tecnología, lo que contribuye a que exista un entorno hostil para las mujeres y las personas LGBTQIA+ y provoca además su exclusión de la participación en la política y de la vida civil.

En una entrevista, Elswah, atribuyó estos problemas a tres factores: las fuentes de los datos de entrenamiento, la anotación de los datos y la forma en que se construyen los modelos. Se considera que muchos dialectos que se hablan en esta región son lenguajes de bajos recursos, lo cual significa que no existen datos de calidad suficiente para entrenar de manera apropiada los sistemas algorítmicos. El proceso de anotación de los datos también brinda oportunidades para que se incurra en errores. Según ella:

> "Cuando se obtienen los datos, se necesitan personas que los anoten, clasifiquen y etiqueten para poder comenzar a procesarlos realmente

45 Elswah, M. (2024). *Does AI Understand Arabic? Evaluating the Politics Behind the Algorithmic Arabic Content Moderation*. Cambridge, MA: Harvard Kennedy School.

46 "*Why I deleted my Twitter account*," Manal al-Sharif on YouTube, octubre de 2018, https://www.youtube.com/watch?v=8regaO3hl_g.

47 Abrougui, A. y Asad, R. (2021). *Digital Safety Is A Right. Syrian Women Journalists and Human Rights Defenders in the Digital Space: Risks and Threats, Syrian Female Journalists Network*. Stichting Female Journalists Network https://media.sfjn.org/en/digital-safety-is-a-right/.

> y crear el modelo y los anotadores. Es un trabajo muy tedioso y normalmente está mal pagado, e incluye sesgos porque las personas que anotan son seres humanos y, si contratas solo a hombres, por ejemplo, y les pides que anoten datos, presentarán algún sesgo."

La solución que han dado las redes para detectar contenido nocivo en idiomas diferentes al inglés ha sido implantar Modelos de Lenguaje Multilingües (MLM, por sus siglas en inglés).[48] A través de un entrenamiento con datos de varios lenguajes al mismo tiempo, los MLM "deducen conexiones entre ellos, lo que les permite descubrir patrones en lenguajes con mayores recursos y aplicarlos a los que tienen menores recursos."[49] Sin embargo, estos modelos no necesariamente han mejorado la moderación del contenido en Oriente Próximo y Norte de África de manera que garanticen una detección y eliminación adecuada y oportuna del contenido nocivo, garantizando al mismo tiempo la libertad de expresión y que la información no se vea gravemente afectada por una censura o eliminación errónea. Un estudio del CDT de 2023 identificó cuatro limitaciones en estos modelos: con frecuencia responden a texto traducido automáticamente, que contiene errores y no se corresponde con los lenguajes nativos, no funcionan bien en todos los idiomas y no tienen en cuenta ni reflejan los contextos de los hablantes del idioma nacional. Además, cuando surgen problemas, resultan difíciles de identificar y resolver.[50]

Por otra parte, la IA generativa y su uso para generar *deepfake* sobre abusos sexuales, cuyo objetivo son en mayor medida las mujeres, reducirán la seguridad de los espacios digitales. Por ejemplo, en Iraq, durante las elecciones parlamentarias de 2021, se extorsionó a una candidata usando *deepfake* sobre abusos sexuales, lo que la obligó a retirar su candidatura.[51] Pese a ello, las empresas tecnológicas no están adoptando las medidas adecuadas para resolver las deficiencias de sus algoritmos, ya sean los que usan para curar y recomendar contenido o los que emplean para moderar. De hecho, debido a los despidos masivos de personal de sus equipos de seguridad y fiabilidad,[52] se prevé que aumente la dependencia de los algoritmos, sin una correcta supervisión por personas humanas que moderen. Esta situación podría agravar, por lo tanto, los sesgos y las inexactitudes en relación con el contenido creado en los dialectos y las lenguas de Oriente Próximo y Norte de África.

"[Las empresas tecnológicas] están despidiendo al personal [que se dedica a moderar el contenido] y pregonando las virtudes del aprendizaje automático, por lo que ya han decidido: este es el futuro. En Estados Unidos, la realidad es que el personal dedicado a moderar contenidos afirma no querer hacer más ese trabajo…internamente el propio personal quiere una estructura automatizada, no quieren vivir la experiencia traumática que supone tener que ver ese horrible material," afirma en una entrevista Azza El Masri, estudiante de doctorado en Periodismo y Medios en la Universidad de Texas. Azza hizo hincapié en la necesidad de un cambio de estrategia por parte de los grupos de la sociedad civil que priorizan las capacidades de aprendizaje automático de estas redes.

48 Nicholas, G. y Bhatia, A. (2023). The Dire Defect of 'Multilingual AI Content Moderation. *Wired*. https://www.wired.com/story/content-moderation-language-artificial-intelligence/

49 Nicholas, G. y Bhatia, A. (2023). Lost in Translation: Large Language Models in Non-English Content Analysis". *Center for Democracy and Technology.* https://cdt.org/wp-content/uploads/2023/05/non-en-content-analysis-primer-051223-1203.pdf.

50 *Ibid.*

51 Al-Kaisy, A. (2022). *Online violence towards women in Iraq*. Elbarlement. https://elbarlament.org/wp-content/uploads/2022/03/Aida-2.pdf

52 Motyl, M. y Ellingson, G. (2024). The Unbearably High Cost of Cutting Trust & Safety Corners. *Tech Policy Press*. https://www.techpolicy.press/the-unbearably-high-cost-of-cutting-trust-safety-corners/

Supresión de contenido sobre derechos en materia de salud sexual y reproductiva por los sistemas algorítmicos de control

La población de los países de Oriente Próximo y Norte de África está encontrando obstáculos para disfrutar plenamente de sus derechos en materia de salud sexual y reproductiva, aunque existen diferencias entre los distintos países individuales.[53] En esta cuestión siguen influyendo mucho ciertos aspectos, entre ellos algunos tabúes, la estigmatización y las sensibilidades culturales en relación con el aborto y la anticoncepción, las infecciones y enfermedades de transmisión sexual y la capacidad de tomar decisiones informadas sobre el propio cuerpo.[54] Es esencial disponer de acceso a información objetiva, inclusiva y no estigmatizante. Sin embargo, al no existir programas de educación sexual ni poder tratar estos temas abiertamente en los entornos familiares, la "población más joven recurre a internet, a la pornografía en línea y a sus iguales para acceder a información sobre salud sexual y reproductiva, que suele ser inexacta y puede hacer peligrar las normas de igualdad de género."[55] Cabe señalar nuevamente que, pese a que internet es un recurso fundamental, la brecha digital y la brecha digital de género siguen siendo una realidad en algunos países. Además, algunos gobiernos censuran basándose en motivos políticos, sociales y religiosos,[56] lo cual influye en el contenido al que se puede acceder, entre otro el relativo a derechos sobre salud sexual y reproductiva. Aparte de estas dos trabas, se ha demostrado que los sistemas algorítmicos de control también eliminan contenido sobre estos derechos o recomiendan otro que no es objetivo o que resulta estigmatizante.

En un artículo de 2018 para Jeem, una organización regional feminista de medios, la autora Salma Mohamed relataba su experiencia como adolescente al buscar información sobre salud sexual y sexualidad en internet, en árabe, y haber encontrado resultados plagados de desinformación, estigmatización y edictos religiosos.[57] Años más tarde, como estudiante de Medicina, se le ocurrió buscar la misma información en inglés y los resultados fueron totalmente diferentes, pues no perpetuaban la estigmatización. Una vez más, en este caso, los sistemas algorítmicos de las plataformas digitales –buscadores de internet– son un reflejo de los datos con los que se entrenan.

Por otra parte, las personas que crean contenido, profesionales de la sanidad, activistas y pertenecientes a grupos de la sociedad civil que publican en redes sociales sobre derechos de salud sexual y reproductiva, entre otras aquellas que lo hacen para concienciar, padecen una constante censura y eliminación de su contenido, publicidad y cuentas. Las personas que publicaban en árabe tenían más probabilidades de ser censuradas que aquellas que lo hacían en inglés, lo que genera un nuevo obstáculo para acceder a información y recursos básicos a quienes sólo hablan árabe.58 Las plataformas utilizan masivamente la automatización para moderar el contenido y la publicidad. Un estudio de la organización de derechos digitales SMEX documentó varios ejemplos de retirada de contenido educativo sobre derechos sexuales y reproductivos por contravenir las políticas de las plataformas en materia de "contenido para adultos", lo que plantea dudas

53 Túnez, por ejemplo, es el único país de la región en el que es legal el aborto cuando se solicita durante el primer trimestre del embarazo. En el resto de la región, el acceso y las restricciones en este sentido varían. Aunque todos los países permiten el aborto cuando está en peligro la vida de la embarazada, en algunos es legal si corre peligro su integridad física o mental, en caso de malformaciones del feto o en caso de violación [véase: Maffi I, Tønnessen L. The Limits of the Law: Abortion in the Middle East and North Africa. Health Hum Rights. 2019 Dic;21(2):1-6. PMID: 31885431; PMCID: PMC6927385.]

54 Oraby D. Sexuality Education for Youth and Adolescents in the Middle East and North Africa Region: A Window of Opportunity. Glob Health Sci Pract. 2024 Feb 28;12(1):e2300282. doi: 10.9745/GHSP-D-23-00282. PMID: 38290752; PMCID: PMC10906548.

55 I*bid.*

56 Freedom House (2024). *Internet Freedom in the Middle East Remained Restricted in 2024*. https://freedomhouse.org/article/fotn-2024-middle-east-release

57 Mohamed, S. (2018). كيف ساهم المحتوى العربي المغلوط في تشويه حياتنا الجنسية؟ [How misinformation in Arabic contributed to distorting our sexual lives., *Jeem*. https://jeem.me/bodies/116.

58 SMEX (2024). From Sharing to Silence: Assessing Social Media Suppression of SRHR Content in WANA. *SMEX*. https://smex.org/from-sharing-to-silence-assessing-social-media-suppression-of-srhr-content-in-wana/.

sobre en qué medida entienden los algoritmos el contexto de esas publicaciones, en particular en árabe.[59]

De acuerdo con la investigadora Mira Nabulsi, que estudió las restricciones que experimenta el contenido sobre derechos en materia de salud sexual y reproductiva en la región, varias personas dedicadas a la creación a las que entrevistó pensaban que Meta estaba contribuyendo a "mantener un statu quo sumamente conservador en nuestras sociedades, que censura el cuerpo de la mujer e información importante pertinente para los derechos de las personas a decidir sobre su vida y su futuro."[60]

En este sentido, mientras los algoritmos se alimenten de datos estigmatizantes sobre derechos en materia de salud reproductiva y sexual, las plataformas seguirán agravando dichos estigmas y desigualdades, al adoptar una moderación sesgada del contenido con respecto a esta región y penalizar a quienes ponen en tela de juicio las normas y tabúes existentes. Esto resulta patente, por ejemplo, en las políticas publicitarias de las plataformas, que parecen fundamentarse en las leyes nacionales y el conservadurismo social de los países de Oriente Próximo y Norte de África, lo que intensifica la censura y las restricciones. Por ejemplo, X prohíbe "promocionar métodos anticonceptivos sin prescripción médica" en muchos países de la región.[61] YouTube, por otra parte, no permite "anuncios relacionados con el control de la natalidad ni con productos para la fecundidad" en 23 países, la mayoría de los cuales están en Oriente Próximo y Norte de África (17 países).[62] Meta posee una política publicitaria general que permite "anuncios que promuevan productos o servicios de bienestar o salud sexual y reproductiva, por ejemplo de anticoncepción y planificación familiar" siempre y cuando "no se centren en el placer sexual."[63] Las políticas de TikTok son similares.[64]

Los retrocesos que supone la automatización para la ya precaria participación de la mujer en la vida económica

La IA amenaza con asumir trabajos que suelen realizar las mujeres, como son las labores de secretariado, contabilidad y llevanza de libros y otras de carácter administrativo. Pese a que estos efectos adversos se materializarán principalmente en países de rentas altas,[65] en los de rentas medidas, algunos trabajos, como el que se realiza en un centro de atención telefónica, seguirán estando amenazados por la automatización. En esta región, los centros de atención telefónica proporcionan empleo a muchas personas de países con rentas medidas, como Túnez[66] y Marruecos,[67] algunas de ellas mujeres.

59 *Ibid*.
60 Nabulsi,M. (2024). Navigating Taboos: Exploring social media policies and SRHR content restrictions in WANA". *SMEX*. https://smex.org/wp-content/uploads/2023/03/MiraNabulsi-SRHR-Mariam-al-Shafei-Fellowship-2023.pdf.
61 "X Business. Healthcare". acceso del 15 de noviembre de 2024. https://business.x.com/en/help/ads-policies/ads-content-policies/healthcare.
62 "Healthcare and medicines". Políticas de Google sobre publicidad. Acceso del 15 de noviembre 2024. https://support.google.com/adspolicy/answer/176031?hl=en&ref_topic=1626336&sjid=13573517111321317776-EU#zippy=%2Ctroubleshooter-birth-control.
63 Meta. "About Meta's Health and Wellness advertising policy". Acceso del 15 de noviembre de 2024. https://www.facebook.com/business/help/2489235377779939?id=434838534925385
64 TikTok. "Adult Content". Políticas de TikTok sobre publicidad. Acceso del 1 https://ads.tiktok.com/help/article/tiktok-ads-policy-adult-content y "Healthcare and Pharmaceuticals," Políticas de TikTok sobre publicidad, acceso del 15 de noviembre del 15. https://ads.tiktok.com/help/article/tiktok-ads-policy-healthcare-pharmaceuticals
65 Gmyrek, P., Berg, J., Bescond, D. (2023). *Generative AI and jobs: A global analysis of potential effects on job quantity and quality*, Documento de trabajo de la OIT 96 (Ginebra, OIT). https://www.ilo.org/publications/generative-ai-and-jobs-global-analysis-potential-effects-job-quantity-and
66 Ahmed, S. (2024). Le télémarketing: Un secteur négligé malgré ses contributions cruciales à l'économie tunisienne, *La-Presse*. *https://lapresse.tn/2024/09/08/le-telemarketing-un-secteur-neglige-malgre-ses-contributions-cruciales-a-leconomie-tunisienne/*
67 TRTFrançais (2024). Intelligence Artificielle, le grand remplacement dans les centres d'appels marocains?. *TRT Français*. https://www.trtfrancais.com/actualites/intelligence-artificielle-le-grand-remplacement-dans-les-centres-dappels-marocains-17699021

Nagla Rizk, profesora de Economía de la American University of Cairo School of Business y administradora fundadora del Access to Knowledge Development Center, afirmó en una entrevista que "cuanto mayor es el nivel de competencias, más probabilidades hay de que la tecnología sea un medio facilitador. Sin embargo, conforme baja el nivel de competencias, sobre todo en los tramos medios, toda tarea que comporte una repetición es susceptible de ser realizada por una máquina."

La incidencia de la IA en la participación de la mujer en la población activa es resultado de sesgos con un profundo arraigo, que animan a la mujer a desempeñar ciertas funciones y puestos de trabajo o la circunscriben a estos. Con la automatización por parte de la IA de algunas de estas funciones y de los trabajos que suelen desempeñar las mujeres, se podría socavar su participación en el mercado de trabajo, en particular en el caso de las que no tienen acceso a las oportunidades laborales en unas condiciones de igualdad con los hombres que les permitan satisfacer las nuevas exigencias de un mercado laboral cambiante, debido a la cantidad de trabajo y de cuidados no retribuidos que deben asumir diariamente en comparación con ellos. En la región de Oriente Próximo y Norte de África, siguen predominando normas de género desiguales y la expectativa sigue siendo de manera generalizada que la mujer se haga cargo de los cuidados, realice las tareas domésticas o su trabajo retribuido se limite a determinados puestos y sectores.[68]

La brecha digital de género en ciertos casos impide además a la mujer desarrollar sus capacidades digitales para que sus competencias sigan siendo válidas para un mercado laboral cambiante. [69] En Oriente Próximo y Norte de África, las mujeres tienen un 12% menos de probabilidades que los hombres de utilizar internet, porque no se pueden permitir el acceso, carecen de las competencias digitales necesarias o debido a normas de género que limitan su presencia en internet y su acceso a este medio.[70] Por otra parte, incluso aquellas que llegan a acceder, suelen disponer de tiempo limitado para interactuar con las tecnologías de la información y las comunicaciones y seguir mejorando sus competencias, debido al trabajo doméstico, que asumen de forma primordial.[71]

Otra dificultad reside en que la discriminación contra la mujer se intensifica en las herramientas de IA utilizadas para procesos de contratación y en la economía *gig* en internet.

Con el aumento de empresas y agencias de selección de personal[72] [73] [74] y plataformas de oportunidades laborales que usan soluciones de IA, como LinkedIn y Bayt.comm, para filtrar las solicitudes de empleo, seleccionar a las personas candidatas a las que se entrevistará y finalmente contratarlas, preocupa que esto reduzca las oportunidades de las mujeres de ser contratadas.

Según Sarah Cupler, candidata a la obtención de un Doctorado en la Universidad de Melbourne, que investiga el uso de herramientas de toma de decisiones automatizadas por parte de la policía: "El proceso de recogida y el etiquetado influye sumamente en los datos, puede haber muchos sesgos en ellos y que pongan de manifiesto una discriminación histórica…Si una mujer tiene más probabilidades de tener que dejar su trabajo por presiones económicas, laborales o sociales tras quedarse embarazada, el algoritmo podría conside-

68 Nazier, H. (2019). *Women's Economic Empowerment: An Overview for the MENA Region*. Instituto Europeo del Mediterráneo. https://www.iemed.org/publication/womens-economic-empowerment-an-overview-for-the-mena-region/.

69 UNESCO, OCDE, BID (2022). *The Effects of AI on the Working Lives of Women*.

70 Traidi, A. (2024). Gender digital divide: The new face of inequality in the MENA region. *Global Campus on Human Rights.* https://gchumanrights.org/gc-preparedness/preparedness-gender/article-detail/gender-digital-divide-the-new-face-of-inequality-in-the-mena-region.html.

71 Rizk, N. (2020). Artificial Intelligence and Inequality in the Middle East: The Political Economy of Inclusion, in Dubber, M. D.; Pasquale, F. y Das, S. (eds), *The Oxford Handbook of Ethics of AI*. https://doi.org/10.1093/oxfordhb/9780190067397.013.40, acceso del 15 nov. 2024.

72 Maharat. Révolutionnez votre recrutement avec l'IA. Acceso del 16 de diciembre de 2024. https://maharat.ma/

73 Look Up Tunisie. Les Nouvelles Technologies dans le Recrutement. Acceso del 16 de diciembre de 2024. https://www.lookuptunisie.com/les-nouvelles-technologies-dans-le-recrutement/

74 Kader. Get to know Kader. Acceso del 16 de diciembre de 2024. https://www.kaderapp.com/en/about-us

rar que las mujeres tienen menos probabilidades de éxito en el mundo laboral. La IA puede ser un reflejo de la sociedad, en particular si se usa sin espíritu crítico, aunque pueda hacerse que parezca objetiva y neutral, y perpetuar estos problemas existentes."

"Ya tenemos este problema de sesgo contra la mujer en el empleo. Todo sistema de IA de la región se alimentará de esos datos y manifestará también dicho problema. Estoy segura de que así ocurre, con contadas excepciones, cuando se sabe que existe un sesgo en una institución concreta… si el sesgo existe realmente, lo habríamos visto reflejado en los mecanismos tradicionales de selección de personal", afirmó Naim. Explicó además que estos sesgos se agravan por la falta de representación de las mujeres en las empresas de IA, también en las que son proveedoras de ese tipo de soluciones. De hecho, pese a que en esta región ha aumentado el número de mujeres que se gradúan en estudios STEM (sigla en inglés que se refiere a ciencia, tecnología, ingeniería y matemáticas), su representación en la población activa sigue siendo desigual[75] y las mujeres tienen además muchas dificultades para acceder a puestos superiores y de nivel ejecutivo. Esto "les impide ocupar puestos superiores (por ejemplo en dirección, como consejeras delegadas y otros puestos ejecutivos), desde los que se ejerce una gran influencia en las empresas e instituciones. Ante estos obstáculos para que la mujer ascienda profesionalmente, se le está privando de la influencia y el poder que comportan esos cargos, lo que contribuirá a que los sesgos se perpetúen.''

En el trabajo mediante plataformas o trabajo *gig* a través de internet, los algoritmos también están agravando estas distorsiones sistemáticas.

Si tenemos en cuenta las elevadas tasas de desempleo, en particular entre mujeres y jóvenes de Oriente Próximo y Norte de África, el trabajo a través de plataformas brinda a muchas personas la oportunidad de participar en la vida económica y obtener ingresos.[76] Sin embargo, este tipo de trabajo presenta un "importante sesgo de género"[77] y perpetúa normas profundamente arraigadas en esta materia, así como los sesgos actuales contra la mujer.[78] Por ejemplo, el trabajo doméstico de la economía *gig*, entre otras las labores de limpieza y cuidado de la infancia, sigue dominado por mujeres.[79] Además, su ilusoria promesa de flexibilidad anima a las mujeres a realizar trabajo *gig* para poder ocuparse, al mismo tiempo, de las tareas domésticas y de cuidado no retribuidas, lo que afianza la desigualdad de género en relación con el trabajo no remunerado.[80]

Una de las preocupaciones que mencionaban las mujeres conductoras de Egipto en un estudio de 2018 es que los sistemas de calificación y los actuales procesos de selección no mitigan de manera efectiva las preocupaciones por la seguridad, en particular en lo que se refiere a los riesgos que constituyen para ellas los pasajeros de género masculino.[81]

Además, las mujeres perciben salarios inferiores. Rizk, que realizó un estudio sobre las mujeres que trabajaban en aplicaciones de movilidad y reparto a domicilio, puso como ejemplo la forma en que los algoritmos penalizan a las mujeres en cuanto a primas, sin tener en cuenta sus contextos y realidades socioculturales:

75 Ferrant, G. y Lunati, M. (2023). *The potential of digitalisation for women's economic empowerment in MENA countries*, en Joining Forces for Gender Equality: What is Holding us Back?, Publicaciones de la OCDE, París, https://doi.org/10.1787/28736eeb-en.

76 Rizk, N. (2020). Artificial Intelligence and Inequality in the Middle East: The Political Economy of Inclusion, in Dubber, M. D.; Pasquale, F. y Das, S. (eds), *The Oxford Handbook of Ethics of AI*. https://doi.org/10.1093/oxfordhb/9780190067397.013.40, acceso del 15 nov. 2024.

77 Siddiqui, Z. y Zhou, Y. (2021). How the platform economy sets women up to fail. *Rest of World*. https://restofworld.org/2021/global-gig-workers-how-platforms-set-women-up-to-fail/

78 Al-Kaisy, A. (2021). *Bias In, Bias Out: Gender and work in the platform economy*. IDRC. https://idl-bnc-idrc.dspacedirect.org/items/7d8e2f97-b1dd-49ad-9843-a0480f5f80eb.

79 Fairwork (2022). *Domestic Platform Work in the Middle East and North Africa*, Fairwork: https://fair.work/en/fw/publications/domestic-platform-work-in-the-middle-east-and-north-africa/

80 Entrevista con Nagla Rizk.

81 N. Rizk et al. (2018). A Gendered Analysis of Ridesharing: Perspectives from Cairo, Egypt. En: *Urban Transport in the Sharing Economy Era*. CIPPEC. https://www.cippec.org/wp-content/uploads/2018/09/UrbanTransport-completo-web_CIPPEC.pdf

"En el caso de la movilidad —esta es información que obtenemos de las personas sobre el terreno—, [las aplicaciones] basan las primas en el número de horas que se trabaja. Es en ese sentido en el que resulta peligroso, porque el algoritmo refleja los sesgos existentes sobre el terreno y los intensifica. Por lo tanto, si en su cultura las mujeres trabajan menos horas que los hombres porque cuidan de sus hijos e hijas, hacen las tareas domésticas... si los algoritmos y la máquina son un reflejo de lo que ocurre en la realidad, lo amplifican. Por lo tanto, en el caso de la movilidad, las mujeres quedan excluidas de forma inmediata de la obtención de primas."

Ciudades inteligentes, vigilancia sexista e integridad física

El desarrollo de capacidades de vigilancia automatizadas no hará más que agravar las amenazas que plantea la vigilancia en la región.[82]

"En la actualidad, estamos avanzando en la capacidad de realizar una vigilancia masiva con mucha más facilidad y también de tratar esos datos mucho más rápidamente, lo que tendría un impacto ingente en la sociedad civil", afirmó Cupler.

Los gobiernos de la región han estado ampliando sus capacidades de IA para fines de vigilancia. Aparte de Israel, la región del Golfo es la que más interés ha mostrado y más ha invertido en el despliegue de tecnología de reconocimiento facial y vigilancia de otro tipo.[83] Qatar, por ejemplo, desplegó una amplísima red de cámaras de vigilancia dotadas de capacidades de reconocimiento facial durante el Mundial de Fútbol de la FIFA de 2022.[84] En los Emiratos Árabes Unidos (EAU), la policía de Dubái se asoció con SAS, un proveedor de soluciones de IA, para que le facilitasen soluciones de actuación policial predictiva85 y en Abu Dhabi, la policía utilizó soluciones de aprendizaje automático y reconocimiento facial para predecir delitos y enviar coches patrulla a zonas consideradas de "alto riesgo".[86] Otros países de Oriente Próximo y Norte de África también están mostrando interés en estas soluciones. El municipio del Gran Ammán, en la capital de Jordania, anunció en 2023 que planea comenzar a usar tecnología de reconocimiento facial para "ayudar a mejorar la seguridad, reducir la delincuencia y para que la capital sea más eficiente."[87] Está proliferando el desarrollo de ciudades inteligentes y muchas administraciones planean invertir en este sector.[88] La Nueva Capital Administrativa de Egipto, por ejemplo, contará con una red de 6.000 cámaras de vigilancia, fabricadas por la empresa estadounidense Honeywell, que enviarán lo que recojan a un centro de mando y control que ejecuta una "avanzada analítica de vídeo para vigilar a las multitudes y los atascos de tráfico, detectar casos de robo, observar objetos o personas sospechosas y activar alarmas automatizadas en situaciones de

82 Business and Human Rights Resource Center (2024). *Keeping watch: Surveillance companies in Middle East & North Africa*. https://www.business-humanrights.org/en/from-us/briefings/mena-surveillance-2024/

83 S. Cupler (2023). "*A Brief Overview of AI Use in WANA*". SMEX. https://smex.org/a-brief-overview-of-ai-use-in-wana/

84 Zidan, K. (2022). The Qatar World Cup Ushers in a New Era of Digital Authoritarianism in Sports. *The Nation*. https://www.thenation.com/article/society/qatar-world-cup-surveillance/

85 Cupler, S. (2023). A Brief Overview of AI Use in WANA. *SMEX*. https://smex.org/a-brief-overview-of-ai-use-in-wana/.

86 Dawood, A. (2021). AI looks at historical data to predict future crimes for UAE's police force. *Mashable Middle East.* https://me.mashable.com/tech/15800/ai-looks-at-historical-data-to-predict-future-crimes-for-uaes-police-force

87 Alakaleek, H. (2023). Facial recognition technology usage in Jordan. *Jordan News.* https://www.jordannews.jo/Section-36/Opinion/Facial-recognition-technology-usage-in-Jordan-31219

88 Belaïd, F., Amine, R., Massie, C. (2024). Smart Cities Initiatives and Perspectives in the MENA Region and Saudi Arabia. En: Belaïd, F., Arora, A. (eds) *Smart Cities. Studies in Energy, Resource and Environmental Economics*. Springer, Cham. https://doi.org/10.1007/978-3-031-35664-3_16

emergencia."[89] Además de las cámaras, se hará un seguimiento de los residentes usando rastreadores de teléfonos móviles, puestos de control digital y puertas de control digital en las estaciones del transporte público. Según Waisová: "Los habitantes de la Nueva Capital Administrativa tienen escasas posibilidades de vivir de manera real y experimentar un desarrollo natural y orgánico de su sociedad. Esta Nueva Capital se ha convertido en un instrumento de segregación y exclusión, y en una fuente de desigualdad social, política y económica."[90]

La automatización agrava la vigilancia y su incidencia en la mujer y las personas LGBTQIA+, que ya padecen altos niveles de control por su género, expresión o identidad de género u orientación sexual. Puesto que la IA facilita que los gobiernos rastreen y recojan datos, en un contexto que carece de una sólida protección de la privacidad y una supervisión judicial independiente,[91] su integridad física y su capacidad para moverse libremente, ejercer sus libertades de pensamiento, opinión o sencillamente tomar decisiones sobre sus vidas y sus cuerpos será aún más difícil bajo la vigilancia constante del Estado. En Irán, aparecieron pruebas del uso del reconocimiento facial, el análisis del tráfico web, la geolocalización y otras herramientas de AI por parte del Gobierno para realizar actuaciones policiales y aplicar normas sobre el uso del velo por parte de las mujeres, así como para acabar con el movimiento por los derechos de la mujer.[92]

No resulta difícil imaginar que surjan más pruebas en el futuro sobre el despliegue de dichas tecnologías para seguir controlando a las mujeres, en particular en países que cuentan con políticas de 'tutela masculina' y restringen los movimientos y las libertades de la mujer (por ejemplo para viajar o sacarse el pasaporte) sin la aprobación de los denominados 'tutores varones', que suelen ser sus maridos si están casadas, o un padre, hermano, tío, abuelo o incluso un hijo en algunos casos.[93] Por ejemplo, en Jordania, Kuwait, Qatar y Arabia Saudí, los tutores varones y otros miembros de la familia pueden denunciar a mujeres ante la policía por "ausentarse" de sus hogares. En Bahréin, Irán, Kuwait, Omán, Qatar, Arabia Saudí y los EAU, las mujeres de las universidades estatales no pueden hacer viajes de campo ni quedarse en las residencias del campus ni abandonarlas sin el permiso de los tutores varones.[94]

La IA y la tecnología de las ciudades inteligentes solo facilitarán el rastreo de las actividades y los movimientos de las mujeres, lo que puede coartar aún más sus libertades y posiblemente pondrá en peligro su seguridad (por ejemplo, cuando se hace un seguimiento de las víctimas de abusos domésticos y son devueltas a la fuerza a sus abusadores). Como escribió Waisová sobre la Nueva Capital Administrativa de Egipto, la población "tendrá escasas posibilidades de vivir de manera real".[95]

89 Thomson Reuters Foundation (2023). FEATURE-CCTV cameras will watch over Egyptians in new high-tech capital. *Reuters*. https://www.reuters.com/article/business/media-telecom/feature-cctv-cameras-will-watch-over-egyptians-in-new-high-tech-capital-idUSL8N33I0DO/

90 Waisová Š. The Tragedy of Smart Cities in Egypt. How the Smart City is Used towards Political and Social Ordering and Exclusion. *Applied Cybersecurity & Internet Governance*. (1):1-10. doi:10.5604/01.3001.0016.0985.

91 Access Now (2021). "*Exposed and Exploited: Data Protection in the Middle East and North Africa*". https://www.accessnow.org/wp-content/uploads/2021/01/Access-Now-MENA-data-protection-report.pdf.

92 George, R. (2023). The AI Assault on Women: What Iran's Tech Enabled Morality Laws Indicate for Women's Rights Movements. *Council on Foreign Relations*. https://www.cfr.org/blog/ai-assault-women-what-irans-tech-enabled-morality-laws-indicate-womens-rights-movements

93 Human Rights Watch (2023). Trapped: How Male Guardianship Policies Restrict Women's Travel and Mobility in the Middle East and North Africa*. Human Rights Watch*. https://www.hrw.org/report/2023/07/18/trapped/how-male-guardianship-policies-restrict-womens-travel-and-mobility-middle.

94 *Ibid*.

95 Waisová Š. The Tragedy of Smart Cities in Egypt. How the Smart City is Used towards Political and Social Ordering and Exclusion. *Applied Cybersecurity & Internet Governance*. (1):1-10. doi:10.5604/01.3001.0016.0985

La automatización de la ocupación y el genocidio: las devastadoras consecuencias para las mujeres y la infancia de Palestina

Durante el genocidio en curso en Gaza por parte de Israel, la milicia israelí ha estado usando sistemas de IA para generar objetivos a los que matar y para cometer un 'domicidio' (la destrucción masiva de hogares). Una de las herramientas, denominada "Where's Daddy", se utiliza para rastrear objetivos y bombardearlos una vez que llegan a sus residencias familiares. Otra herramienta, llamada "Lavender", designó como "sospechosas" a decenas de miles de personas palestinas en Gaza y, según una investigación de +972 Magazine y Local Call, la milicia sabía que el sistema "en ocasiones señalaba a personas que únicamente tenían una vaga conexión con grupos militantes o que no tenían ninguna conexión en absoluto con estos."[96]

"El nivel de destrucción que hemos visto en Gaza solo es posible porque la tecnología de IA ha acelerado mucho más la toma de decisiones", señaló Cupler.

Esto tiene consecuencias devastadoras para la población civil de Gaza, ya que prácticamente un 70% de las personas fallecidas en este conflicto son mujeres y niños o niñas según datos de la ONU.[97] Análisis de Oxfam de septiembre de 2024 concluyeron que "en el último año, la milicia israelí ha matado a más mujeres y niños y niñas en Gaza que en el período equivalente de cualquier otro conflicto durante las dos últimas décadas,"[98] lo que subraya el impacto de la violencia contra la mujer y contra la infancia, como consecuencia de la deshumanización de la población palestina por parte de Israel y la falta de diligencia debida en la ejecución de la guerra, lo que incluye el uso de sistemas automatizados, para establecer como objetivos infraestructuras civiles tales como escuelas, hogares, hospitales y refugios. Umaiyeh Khammash, administrador del socio de Oxfam Juzoor, que presta apoyo a cientos de miles de personas en más de 90 refugios y puntos de atención sanitaria en toda Gaza,[99] señaló que las mujeres de Gaza "soportan una doble carga": "Muchas se han convertido de repente en la cabeza de familia de sus hogares, abriéndose paso para sobrevivir y cuidar en medio de la destrucción. Las madres embarazadas y en período de lactancia han encarado inmensas dificultades, también por la desintegración de los servicios de atención sanitaria".

En los puestos de control de la Ribera Occidental ocupada, Israel usa el reconocimiento facial como parte de una amplia red de cámaras de vigilancia que escanea las caras de la población palestina y las añade a las bases de datos de vigilancia, sin su consentimiento, para seguirlas constantemente, "como parte de un intento deliberado por parte de las autoridades israelíes de crear un entorno hostil y coercitivo".[100]

Toda persona de Palestina que intente atravesar estos puestos de control se enfrenta a una realidad plagada de "largos períodos de espera, interrogatorios invasivos y comprobaciones de identidad, así como a una amenaza constante de violencia."[101] En el caso de las mujeres, estos puestos de control representan además "imposiciones muy sexistas relacionadas con la (in)movilidad, la experiencia corporal y las relaciones de cuidado," según argumentan el autor Mark Griffiths y la autora Jemima Repo en un artículo en el que

96 Abraham, Y. (2024). 'Lavender': The AI machine directing Israel's bombing spree in Gaza. +972 *Magazine*. https://www.972mag.com/lavender-ai-israeli-army-gaza/.

97 E. Farge (2024). Gaza women, children are nearly 70% of verified war dead, UN rights office says. *Reuters*. https://www.reuters.com/world/middle-east/nearly-70-gaza-war-dead-women-children-un-rights-office-says-2024-11-08/

98 Oxfam (2024). More women and children killed in Gaza by Israeli military than any other recent conflict in a single year – Oxfam. Oxfam. *https://www.oxfam.org/en/press-releases/more-women-and-children-killed-gaza-israeli-military-any-other-recent-conflict.*

99 *Ibid*.

100 Amnistía Internacional (2023). Israel/OPT: Israeli authorities are using facial recognition technology to entrench apartheid. Amnistía. https://www.amnesty.org/en/latest/news/2023/05/israel-opt-israeli-authorities-are-using-facial-recognition-technology-to-entrench-apartheid/

101 Griffiths, M. y Repo, J. (2021). Women and checkpoints in Palestine. *Security Dialogue*, 52(3), 249-265. https://doi.org/10.1177/0967010620918529

se analizan las dimensiones sexistas de los puestos de control de Israel en relación con las mujeres de Palestina.[102] De hecho, concluyeron que la capacidad de las mujeres para travesar estos puestos de control se limita a sus funciones como cuidadoras (es decir, si acompañan a familiares que van a obtener tratamiento médico) o a motivos religiosos, lo que refuerza aún más las normas de género existentes que sitúan a los hombres como cabezas y jefes de familia de los hogares y a las mujeres, como cuidadoras.

Como Israel utiliza sistemas de IA en los puestos de control, esos sesgos y formas de discriminación no harán más que agravarse. Existe, por lo tanto, una necesidad de investigar en mayor medida las repercusiones del uso de la IA en las mujeres y las chicas de Palestina y otros lugares mientras continúe el genocidio en Gaza e Israel amplíe sus ataques a la población palestina de la Ribera Occidental y otros países como Líbano y Siria.

Conclusiones y recomendaciones

Este capítulo analiza las repercusiones pluridimensionales de la IA en la justicia de género en la región de Oriente Próximo y Norte de África. Examina de manera específica el papel de la IA generativa, los bots y los sistemas algorítmicos desplegados por las redes sociales en la proliferación de la violencia de género y los estereotipos nocivos contra las mujeres y las personas LGBTQIA+ en la región. Al mismo tiempo, las periodistas y las mujeres que defienden los derechos humanos, las activistas feministas, las comunidades LGBTQIA+ y otras personas que quieren contrarrestar estas narrativas, se enfrentan a censura y más violencia debido a modelos de lenguaje MLM que generan inexactitudes y no funcionan bien en los diversos contextos, idiomas y dialectos de Oriente Próximo y Norte de África.

En el caso de El Masri, una solución podría ser destinar más recursos a la creación de Modelos de Lenguaje Pequeños (SLM): “son igual de robustos y, de hecho, incluyen más contexto; se pueden crear más parámetros de contexto en un SLM que en un LLM [Modelo de Lenguaje Grande]. Además, se puede hacer de forma colaborativa.” Invertir en soluciones de este tipo localizadas y promovidas por la comunidad, de una manera inclusiva que garantice la participación de la mujer, las personas LGBTAIQ+, las minorías y las personas de diferentes especializaciones puede ayudar a superar los efectos nocivos de estas plataformas en relación con la moderación de contenido, en particular teniendo en cuenta que esos efectos se agravarán con el avance del desarrollo tecnológico y la reducción de la supervisión humana.

Ahora que la automatización amenaza con sustituir los puestos de trabajo o las funciones laborales que suelen ser repetitivas, la preocupación es que este fenómeno incidirá en mayor medida en los empleos ostentados por mujeres. Para que la participación de la mujer en la población activa no se vea mermada, es esencial que los diferentes grupos de interés de las administraciones, el sector privado y las organizaciones nacionales e internacionales prioricen las acciones de capacitación complementaria de la mujer y de todas aquellas personas con trabajos amenazados por la automatización. Además de diseñar y poner en marcha programas de actualización de competencias, también es necesario reducir la brecha digital de género y enfrentar las actuales normas de género que imponen a la mujer las responsabilidades vinculadas a tareas de cuidado no retribuidas, que restringen sus posibilidades para competir en condiciones de igualdad en un mercado laboral cambiante. Esto último constituirá una ardua batalla, ya que esas normas están muy arraigadas. Sin embargo, mediante la colaboración con las mujeres, los grupos de mujeres nacionales y las organizaciones de derechos de la mujer, es posible diseñar programas de actualización de competencias que tengan en cuenta sus necesidades y realidades. Y, lo que es más

102 *Ibid*.

importante, las empresas han de percibir incentivos para que sus plantillas puedan dedicar algunas horas de trabajo a la semana o al mes a mejorar sus competencias.

En Oriente Próximo y Norte de África, donde muchos países presentan elevadas tasas de desempleo, en particular entre mujeres y jóvenes, el trabajo a través de plataformas ofrece un sustento a muchas personas, entre otras cosas dándoles la oportunidad a las mujeres de participar en la vida económica y obtener sus propios ingresos. Sin embargo, este tipo de trabajo sigue estando dividido por las actuales normas de género y los algoritmos de las plataformas *gig* suelen terminar siendo un reflejo de los sesgos existentes contra la mujer. Por lo tanto, resulta esencial documentar en mayor medida los efectos de estas aplicaciones y sus algoritmos en las personas trabajadoras desde una perspectiva de género. También existe la necesidad de potenciar las iniciativas que recaben de manera proactiva la participación de personas trabajadoras de la economía *gig*, para conocer sus preocupaciones y necesidades, y para usar esos conocimientos con el fin de promover que las plataformas *gig* modifiquen sus políticas y erradiquen los sesgos de sus algoritmos.

Por último, el despliegue de la IA por parte de las administraciones, en particular de tecnologías de seguimiento, reconocimiento facial, sistemas de toma de decisiones automatizados, constituye una grave amenaza para los derechos humanos, el espacio cívico y la ciudadanía. Los más graves de estos riesgos han surgido del despliegue de ese tipo de sistemas por parte de Israel en su ocupación de los territorios palestinos, en particular durante la actual guerra en Gaza. Resulta urgente regular estas tecnologías basándose en el derecho internacional sobre derechos humanos, aunque los Estados podrían optar por no tener en cuenta ninguna de esas medidas. Por lo tanto, la gobernanza internacional debería combinarse con una presión sobre las empresas tecnológicas que son cómplices de esas violaciones (por ejemplo, los proveedores de tecnologías de reconocimiento facial y vigilancia y las empresas que prestan servicios de computación en la nube y aprendizaje automático, entre ellas Google y Amazon Web Services).[103] Se podría presionar para que se regulen las exportaciones, de manera que se restrinja la venta de tecnologías tales como las de proveedores de reconocimiento facial y se boicotee a las empresas que no respeten los derechos humanos. Debería adoptarse esta táctica más allá del contexto de la guerra en Gaza por parte de Israel, puesto que otros gobiernos de Oriente Próximo y Norte de África están aumentando su inversión en tecnologías de vigilancia mediante IA, en especial en reconocimiento facial y ciudades inteligentes, para intensificar el control y la opresión que perpetran. Gracias a la IA, a los Estados les resultará más fácil rastrear a las personas, recabar más datos sobre ellas y analizarlos. De ese modo, será aún más difícil eludir la vigilancia por parte del Estado, lo que afectará de manera desproporcionada a las mujeres y las personas LGBTQIA+, que ya sufren altos niveles de control y discriminación. Sus libertades más básicas se someterán a un seguimiento y rastreo constantes, lo que hará que puedan perder el poco margen de maniobra que han tenido hasta ahora para vivir de la manera más libre y auténtica posible.

103 Fatafta, M. y Leufer, D. (2024). Artificial Genocidal Intelligence: how Israel is automating human rights abuses and war crimes. Access Now. https://www.accessnow.org/publication/artificial-genocidal-intelligence-israel-gaza/

Recapitulación: Hilvanar las dimensiones y tejer otro modelo de inteligencia artificial

Carlos Bajo Erro
Oxfam Intermón

Una tecnología que está cargada de política

La inteligencia artificial es la tecnología que está detrás de algunas de las investigaciones médicas más innovadoras para diagnosticar, de manera temprana, poco invasiva y con seguridad, enfermedades ampliamente letales o para prever y planificar la reacción ante fenómenos climáticos extremos que amenazan con ser cada vez más frecuentes y destructivos. Pero también es la tecnología que está detrás de aplicaciones superfluas para elaborar memes visuales pretendidamente divertidos, vídeos tiernos de gatitos o de chatbots que funcionan como asistentes de compañía y que no pueden garantizar ni la sensatez de sus instrucciones, ni la seguridad de sus usuarios; igual que de las herramientas que se emplean para aumentar la credibilidad de las noticias falsas con materiales gráficos o vídeos manipulados, y también de las que están sacudiendo la difusión de material de contenido sexual, incluso, con imágenes no consentidas. Y esta diversidad complica mucho adoptar una posición frente a esta tecnología que se ha convertido en *boom* publicitario, esperanza vital y moda, todo al mismo tiempo.

Sin embargo, tal vez no se trate de posicionarse frente a una tecnología concreta, sino de analizar y tomar decisiones en función de los usos y también de las consecuencias.

La inteligencia artificial se inserta en las cadenas de producción internacionales que se derivan del proceso de globalización de las últimas décadas. Así, si conectamos en un mapamundi los puntos en los que se realiza algún proceso del ciclo de vida de la inteligencia artificial pocos territorios quedarán desconectados. Hablamos de evidenciar la relación de las cartografías en las que se extraen los materiales con los que construyen los microprocesadores, o los lugares en los que, con ciclos vitales muy cortos, se desecha toda la electrónica de los centros de datos; los espacios de los que se extraen los datos que se usan en el diseño de los modelos o en los que se ubican los centros de datos; aquellos en los que se producen los entrenamientos o en los que viven las personas que etiquetan o refinan datos y, en general, las personas repartidas por todo el mundo que realizan microtareas imprescindibles para la construcción de las herramientas basadas en inteligencia artificial.

El mapa podría completarse con las ubicaciones de las personas afectadas por esos desarrollos. Esta nueva capa representaría a las personas en cuyas vidas se cruzan sistemas algorítmicos, en la gestión de su trabajo, en la prestación de servicios públicos o en su consumo de ocio; o las que participan en el diseño, desarrollo o implementación de las herramientas, desde las que se encuentran en situación de mayor precariedad hasta los máximos directivos; pero también aquellas que han aportado datos, consciente o inconscientemente, usados para los procesos de entrenamiento, ya sean autores y autoras de obras escritas, verbales, gráficas o audiovisuales, usuarios de herramientas digitales o simplemente personas cuyas acciones han dejado rastro en una base de datos; y, de la misma manera, las que experimentan los efectos indirectos de alguno de los procesos, desde beneficiarios de una vacuna descubierta en una investigación sanitaria con inteligencia artificial, hasta vecinos de un centro de datos privados de agua o electricidad por las necesidades de la infraestructura digital.

Es evidente que la inteligencia artificial no ha inaugurado las mecánicas globales industriales, pero sí que se está convirtiendo en su máxima expresión, a través de las largas cadenas de suministro, de la deslocalización de sus procesos o de la microfragmentación de sus tareas, además del intento de universalizar la comercialización de algunos de sus productos. Tampoco ha inventado las prácticas de producción que algunas compañías emplean, como el extractivismo o un intenso ejercicio de lobby. El modelo hegemónico de desarrollo de la IA está propiciando una extremada acumulación de riquezas en las manos de los responsables de las pocas empresas que tienen la capacidad para desarrollar tecnología en estas condiciones, que además se traduce en acumulación de poder real y práctico a escala global. Estas derivas colocan en una situación comprometida a los actores del sector digital (incluidas las empresas) que defienden una inteligencia artificial que sea justa y equitativa y apuestan por un modelo de desarrollo de la tecnología que respete escrupulosamente los derechos fundamentales de los habitantes de todos los territorios y que no agrave las desigualdades, sino que suponga una aportación real al bienestar colectivo.

Impacto de la inteligencia artificial en el bienestar

Este documento pretende arrojar un poco más de luz sobre algunos de los espacios en los que la expansión de la inteligencia artificial está teniendo un impacto en las desigualdades, ya sea a través del agravamiento de algunos de los ejes de discriminación ya consolidados o de la generación de otros nuevos. Teniendo en cuenta cómo la presencia de la inteligencia artificial se va extendiendo a más y más ámbitos de nuestras vidas, se han escogido aquellas dimensiones en las que se considera que la influencia es más profunda o más significativa.

Además de evidenciar estas huellas de la inteligencia artificial en diferentes ámbitos que marcan la aspiración al bienestar de las personas, la suma de todas esas dimensiones transmite la complejidad del fenómeno y sus efectos. Por un lado, insistiendo en las múltiples capas de la existencia que se ven alcanzadas por la expansión de una tecnología como esta; y, por otro lado, reflejando las intersecciones y las interacciones entre muchas de esas capas.

Desde la perspectiva de los desequilibrios económicos que genera o agrava el actual modelo de desarrollo de la inteligencia artificial, Sofia Scasserra llama la atención sobre la reproducción de esquemas y dinámicas antiguas. La promesa de un nuevo mundo que acompaña la expansión de la inteligencia artificial no ha desmantelado la histórica división internacional del trabajo, que ha establecido a los países de la llamada “periferia” como proveedores de las materias primas que han propiciado el crecimiento de la economía

mundial y como consumidores finales de unos productos que se transforman en los países del centro, en los que se queda también el valor añadido.

En el caso del desarrollo de la inteligencia artificial, además, la producción exige un nivel de tecnificación y de inversión que es difícilmente alcanzable para esos países considerados periféricos. De manera que la producción de lo que la economista argentina llama la "IA industrial", refiriéndose a los procesos de transformación de datos masivos para dar lugar a los grandes modelos, es patrimonio de un grupo reducido de países del Norte Global. A los países de la periferia apenas les queda el espacio de lo que llama la "IA artesanal" que son los desarrollos más modestos y menos lucrativos, con un alcance apenas local y que siempre trabajan sobre la base de la IA industrial.

El segundo papel del Sur Global en este modelo de desarrollo vuelve a ser el de ser proveedora de la materia prima necesaria para el proceso industrial. Se trata de datos y mano de obra que se despojan del Sur Global, en gran medida, siguiendo esquemas extractivistas y neocoloniales. Por ello, Scasserra emplea una imagen muy sugerente que, al mismo tiempo, es una advertencia real: "No podemos permitir un nuevo siglo de saqueo de recursos. No volvamos a ser Potosí".

Con otra imagen reveladora, "no es una nube, es una nave industrial", Ana Valdivia recuerda la importancia del relato para desentrañar el impacto ecosocial del desarrollo de esta tecnología. Las narrativas, especialmente las que ha impuesto la industria tecnológica, juegan un papel fundamental a la hora de abordar la que se va revelando como la profunda huella medioambiental de la inteligencia artificial. Por ello, en la divulgación de estos efectos negativos para el planeta y para la vida hay que empezar por desarticular discursos que se difundieron como realidades, a pesar de ser solo promesas. El augurio de una aliada determinante contra el cambio climático en la inteligencia artificial, de momento, solo se ha traducido en una enorme voracidad en recursos naturales, energía, agua y, cada vez más, territorio.

En el mismo sentido, las narrativas han alimentado una imagen de inmaterialidad, obviando la realidad material de la infraestructura digital sobre la que se apoya esa capacidad de computación. Lo que pasa entre una orden a una herramienta basada en inteligencia artificial y una deslumbrante respuesta no es magia. Son un conjunto de procesos electrónicos que discurren a través de miles de kilómetros de gruesos cables de fibra óptica protegidos por recubrimientos plásticos y metálicos o de toneladas de metal orbitando alrededor de la Tierra, para llegar a enormes edificios de hormigón ocupados por servidores y procesadores que requieren cantidades ingentes de electricidad para mantenerse activos 24 horas al día, siete días a la semana y de agua para mantener una temperatura adecuada, en una competencia por los recursos con las comunidades locales. Unas cantidades que, por otro lado, apenas pueden estimarse o inferirse con complicadas operaciones, porque la mayor parte de las compañías se resisten a ofrecer información clara, completa y transparente.

El camino que toma el actual modelo de desarrollo de la inteligencia artificial entra en un conflicto frontal con la más básica exigencia del mantenimiento de la vida cuando se aborda la connivencia entre algunas de las empresas del sector tecnológico con la industria armamentística. El conflicto se agrava si se tiene en cuenta que, de manera progresiva, es la tecnología de consumo la que se está empleando y adaptando a los usos militares, para aumentar la capacidad letal de los ejércitos. Este abordaje transmite una realidad que puede resultar redundante, pero que encierra un dramatismo insostenible, la de la "deshumanización" de la guerra: El efecto de la proliferación de armas autónomas y de herramientas que actúan sin la necesidad de la interacción humana para acabar con la vida.

En esta siniestra colaboración entre la industria militar y el sector tecnológico, de nuevo, se imponen las narrativas para tratar de justificar un resultado contrario a la sostenibilidad de la vida. En este caso, el argumento es el de objetividad y precisión, el del robot que analiza

y no se equivoca, que no titubea y actúa. Aunque los resultados de estas experiencias, hasta el momento, distan mucho de ser quirúrgicos y abundan unos errores de cálculo con consecuencias que no tienen vuelta atrás. Sí que es más difícil de cuestionar, otro de los principales argumentos, el de la reducción de costes. Sin embargo, necesita un matiz, se trata de una reducción de costes económicos, ya que los costes humanos no dejan de aumentar en los actuales conflictos armados en los que se están ensayando estas herramientas letales. Esta connivencia está tensando las costuras del sector tecnológico, precisamente en su dimensión humana. La de las y los trabajadores que recuerdan los códigos éticos de algunas de estas empresas, que consagraron sobre el papel que su investigación y desarrollo de inteligencia artificial tenía algunas líneas rojas, una de las principales, que nunca perjudicaría o dañaría la vida.

Es incontestable que la inteligencia artificial generativa se está empleando ampliamente en la creación de contenidos y en la producción de conocimiento. De nuevo, en este aspecto emerge el desequilibrio global en la producción de tecnología y la concentración de esas capacidades en unos pocos actores, que en su mayoría son corporaciones del Norte Global que refuerzan el proceso de acumulación de riqueza y de poder, para incorporar otra capa a las desigualdades. La base de los modelos de inteligencia artificial es el lenguaje y esa concentración ha hecho que el inglés haya sido la lengua preeminente en el desarrollo de estas herramientas. La advertencia, en este sentido es clara y contundente: no incorporar la diversidad lingüística y cultural en este proceso se traduce en un empobrecimiento de los conocimientos que se están produciendo y en un obstáculo de muchas comunidades para reconocerse en esa producción que está llamada a ocupar un importante espacio en el futuro (si no ya, en el presente).

Esta incorporación de la diversidad cultural y lingüística no consiste, sin embargo, en hacer que esos grandes modelos sean accesibles en diferentes idiomas, esa es solo una maniobra empresarial para extender su comercialización; ni siquiera en introducir en el entrenamiento contenidos y datos originados en diferentes idiomas. Como explica Pelonomi Moiloa, la incorporación de la diversidad cultural y lingüística en la inteligencia artificial pasa por desarrollar modelos desde las estructuras de lenguas diversas, de manera que todo el modelo se impregne de las particularidades de cada contexto y construya sobre la riqueza de esa diversidad.
Además de la producción de contenidos, la gestión de tareas y, especialmente, la gestión del trabajo es una de las funcionalidades más extendidas de las herramientas basadas en inteligencia artificial o mecanismos algorítmicos. Por ello, un impacto fundamental en el mundo del trabajo es el que supone la gestión algorítmica en el sector de la economía de plataforma. Para intentar entender su alcance es necesario vislumbrar sus consecuencias. De nuevo, emergen las narrativas intencionadas que tratan este fenómeno como una nueva oportunidad laboral o como un valor positivo de flexibilidad y autogestión. Este discurso invisibiliza el conflicto que se abre en la relación de esas condiciones de trabajo con los derechos laborales aparentemente consolidados.

Las investigaciones evidencian que la opacidad de los algoritmos coloca en una situación de indefensión a trabajadores y trabajadoras que buscan una manera de ganarse la vida en condiciones difíciles y que se ven obligadas a aceptar una precarización que pone en jaque los derechos laborales básicos. La duración de las jornadas de trabajo, el desconocimiento de la remuneración o la arbitrariedad de la asignación de tareas confluyen con la hipervigilancia y con amenazas a la intimidad y a la privacidad. De nuevo, una imagen sugerente para ilustrar esta situación es la que proponen Villarreal y Pérez de la Mora cuando hablan de la búsqueda por parte de los trabajodores del *jackpot*, el gran premio, como si de un juego de azar se tratara mientras el algoritmo les da las tareas básicas para mantener su interés y su actividad.

Los servicios públicos, más específicamente, en la administración de las herramientas de protección social, también han recurrido a la gestión algorítmica. Vuelve a hacerse necesario separar la paja del grano, las narrativas de las realidades. La implantación de estas

herramientas se produce con un argumentario que apela a la eficiencia, la reducción de costes innecesarios y la optimización de unos recursos que lamentablemente son limitados. En la práctica, una buena parte de estas herramientas, al menos, las que han acabado haciéndose más populares por sus fallos y sus graves consecuencias, se orientan exclusivamente a la persecución del fraude, desvelando la introducción de prejuicios racistas o machistas, en una lógica de sospecha que se cierne sobre las personas que, en situación de vulnerabilidad, acceden a esos mecanismos de protección social.

Resulta relevante que los algoritmos de este tipo que se han hecho populares han salido a la luz después de deficiencias detectadas en su funcionamiento a través de investigaciones de organizaciones de la sociedad civil. Algo que ya da una idea del clima de opacidad que, además, se ha desvanecido cuando las consecuencias de esos errores han sido especialmente graves. En general, el funcionamiento de estos mecanismos aumenta la indefensión de las personas que están siendo controladas por esos sistemas, que pertenecen a colectivos en situación de vulnerabilidad, porque la interlocución se complica y porque, a menudo, se impone la presunción de infalibilidad de la máquina.

La capacidad de ejercer eficazmente los derechos de participación política también aparece afectada por las características de algunas herramientas basadas en inteligencia artificial que pueden adulterar los procesos democráticos. Las investigaciones de algunas organizaciones de la sociedad civil como AI Forensics, han detectado escenarios de riesgo para la democracia en el contexto europeo. Entre las diversas situaciones destacan tres que aparecen probadas. La primera es la incorrección de las informaciones ofrecidas por algunos *chatbots*, precisamente en contextos electorales, y que amplificaban narrativas que constituyen un riesgo sistémico para la democracia. La segunda es la insuficiencia de los esfuerzos de moderación de contenidos, por ejemplo, según los idiomas en los que se difundan esos contenidos. Y la tercera es el uso de herramientas de inteligencia artificial generativa para producir contenidos que apoyan las noticias falsas aumentando su verosimilitud y su alcance.

Dos de los ejes de discriminación más extendidos y, al mismo tiempo, más arraigados son, por un lado, la raza, etnia o lugar de origen; y, por otro, el género. No se puede perder de vista que son condiciones estructurales las que subyacen a los impactos negativos de las herramientas basadas en inteligencia artificial.

En el caso del racismo, bajo este paraguas se identifican prácticas que solo se normalizan en marcos con una mayor presencia de personas racializadas, como ocurre con el uso de herramientas que prácticamente se reservan para la gestión de la migración, porque se considera que en otros contextos la amenaza a los derechos fundamentales las hace inasumibles. En otros casos, es la práctica la que nos dice qué aplicaciones, que no han probado su eficacia o en las que se han detectado fallos recurrentes, resultan especialmente perjudiciales para personas racializadas, como ocurre en el uso de mecanismos algorítmicos en el ámbito de la seguridad o en la gestión de los servicios públicos y, especialmente, de las herramientas de protección social, con un impacto directo en el ejercicio de sus derechos. Por otro lado, la economía política que dicta el ciclo de vida de la inteligencia artificial, desde la extracción de materiales críticos, hasta la implantación de herramientas está marcada por las raíces racistas de los desequilibrios entre los Sures y los Nortes globales. Y en el caso de la discriminación de género, la inteligencia artificial generativa y los sistemas algorítmicos empleados por las redes sociales aumentan la propagación de la violencia de género, los contenidos misóginos, el odio y los estereotipos dirigidos contra las mujeres y las comunidades LGBTIQ+. Investigaciones como la de Afef Abrougui muestran cómo estas circunstancias dificultan la participación social y política en los contextos más hostiles y obstaculizan los debates sobre los derechos sexuales y reproductivos, propiciando la perpetuación de esas discriminaciones.

De la misma manera, se está detectando un impacto de género en el espacio laboral, donde la inteligencia artificial ocupa con más facilidad puestos de trabajo feminizados como los de administración, precarizando a las mujeres y contribuyendo a la división de género tradicional del trabajo. Y, en el espacio político hay casos de violencia sexual mediante *deep fakes* a políticas que incluso tuvieron que abandonar su carrera. Los efectos sobre las mujeres se agravan, por ejemplo, en las situaciones de hipervigilancia que propicia la consolidación de las ciudades inteligentes y permiten reforzar otras conductas de control. Lo mismo ocurre en los casos de conflictos armados, en los que la introducción de herramientas basadas en inteligencia artificial se hace sentir especialmente entre las mujeres y los niños y las personas LGTBIQ+, donde las herramientas de vigilancia pueden tener efectos devastadores.

Multidimensionalidad e interseccionalidad

A pesar de que cada uno de los capítulos de este volumen se ha centrado en una dimensión concreta, la mayor parte de los análisis y evaluaciones que contienen estos apartados ya apuntan a la multidimensionalidad e interseccionalidad de diferentes capas, a las relaciones que se establecen entre estos ámbitos y cómo muchos de ellos se refuerzan entre sí, se amplifican o se complementan y en todo caso interactúan para mostrar la complejidad del fenómeno. La comprensión del impacto de la inteligencia artificial en las vidas de las personas y en su bienestar en ningún caso puede pasar por una mirada simplista ni superficial. La influencia en los múltiples ejes de discriminación (los que se mencionan específicamente aquí y muchos otros para los que no ha habido espacio) y en las desigualdades exige tiempo y profundidad para ser evaluada.

Después de repasar los análisis de estos capítulos se hace evidente, por ejemplo, que el impacto negativo de las herramientas basadas en inteligencia artificial para una mujer migrante en Europa se ve atravesado por la dimensión de raza y de género, pero también por la dimensión que tiene que ver con las amenazas a la democracia, la que hace referencia a la aportación a la industria armamentística y, tal vez, la que se refiere a los mecanismos de protección social. Es evidente que podría haber impactos positivos, también, pero esos no agravan las desigualdades y no son objeto de la misma preocupación.

De la misma manera el conductor de una plataforma de transporte de las afueras de Ciudad de México podría verse cruzado por la dimensión relacionada con el trabajo digno y la gestión algorítmica del trabajo, pero también por el impacto ambiental de los centros de datos que se están instalando en su entorno y por los desequilibrios económicos globales. Y la lista podría continuar. Podría ser un ejercicio interesante imaginar perfiles aleatorios en todo el mundo y pensar cuáles de estas dimensiones del impacto de la inteligencia artificial afecta a sus vidas, desde una civil en Ucrania, hasta un campesino en el sudeste asiático o una música que rapea en una lengua nacional en África…

Además, existen ejes de desigualdad en los que no se ha podido profundizar en este documento. Por ejemplo, falta explorar en profundidad el impacto de la IA en los cuerpos, desde una mirada especialmente anticapacitista, también *queer* y feminista, y anticolonial y antirracista, aunque hay aspectos de estas relaciones en los que sí se profundiza a lo largo del texto.

¿Y si el modelo de desarrollo de la inteligencia artificial fuese otro?

Las reflexiones sobre las diferentes dimensiones de impacto de la inteligencia artificial en el bienestar de las personas que se comparten en este volumen destacan las amenazas y los efectos negativos, pero no pretenden enmendar la tecnología en sí misma. De hecho, hay un esfuerzo por proponer alternativas y soluciones, que destilan una certeza: el problema no es la inteligencia artificial, sino el modelo concreto de desarrollo de esta tecnología que se ha convertido en el hegemónico.

Las sugerencias van orientadas a cuestionar los principios que impactan en el agravamiento de las brechas de las desigualdades, pero las investigadoras y expertas van perlando los capítulos de propuestas de transformación que podrían enmendar los vicios del modelo.

Entre esos apuntes aparecen, por ejemplo, la posibilidad de inspirarse en el regionalismo que en América Latina se ha puesto en marcha ante otros retos y en la tradición de trabajo alrededor de los bienes comunes, como ejes para organizar mecanismos de gobernanza más igualitarios. Frente a la evidencia de que el actual modelo tiene unas exigencias ambientales insostenibles, aparece la propuesta de abordar la utilidad de los sistemas algorítmicos desde un análisis crítico, es decir, decidir cuáles de los algoritmos, cuáles de las herramientas basadas en inteligencia artificial ofrecen realmente un beneficio para las vidas, teniendo en cuenta el alto costo que presentan. Esa sería una solución más o menos inmediata a la espera de que se reformule el abordaje de la tecnología en su conjunto desde perspectivas que pongan en el centro la vida y la sostenibilidad y no el lucro.

Para otras de las amenazas, las propuestas tienen que ver con reforzar y mejorar el aparato regulatorio, concretamente, hacerlo más eficiente ante la flexibilidad que han mostrado estas herramientas para pasar por debajo de los radares y eludir los controles. En este sentido, las autoras coinciden en la necesidad de una mejora en la regulación que puede enmarcar la transferencia de conocimientos y las interacciones entre la industria tecnológica y la armamentística, así como garantizar que el Derecho Internacional Humanitario responde a unas nuevas modalidades de conflicto bélico. Algo parecido ocurre con los derechos laborales, sobre los que es necesario que la regulación garantice las conquistas consolidadas y responda a las amenazas que llegan desde una pretensión intencionada de revisar todo el marco laboral.
En relación con las amenazas para la democracia, todo apunta a que existen mecanismos suficientes para garantizar formalmente su protección, sin embargo, los hechos revelan algunas dudas sobre su eficiencia o sobre su capacidad de aplicación. Las investigadoras han detectado la necesidad de una mayor vigilancia y un margen mayor para la intervención en los casos en los que se evidencian los usos perversos de esas herramientas. Sin ir más lejos, advierten que la Digital Services Act (DSA) requiere a las grandes plataformas y a los grandes buscadores que evalúen los riesgos asociados a sus servicios, pero eso no siempre ocurre.

El refuerzo de las capacidades de las comunidades y la promoción de su participación aparece también como un contrapeso a algunos de los impactos negativos detectados. En el caso de la necesidad de introducir la diversidad lingüística esta sería una alternativa ideal y además, como apunta Moiloa, abre la puerta a un paradigma que beneficia no sólo a los hablantes de lenguas subrepresentadas, sino también a los ecosistemas tecnológicos globales al proporcionar modelos diversos y sostenibles para el futuro. Ocurre algo parecido en la lucha contra la discriminación racial. Este último es un marco en el que se platea la necesidad de que las comunidades tengan las competencias para producir herramientas adaptadas a sus realidades, pero también los recursos necesarios para hacerlo, lo que requiere una redistribución y romper el monopolio del acceso a la financiación de unos pocos actores. Frente a la amenaza de la complicidad entre industria tecnológica y militar también las comunidades tienen importancia, en este caso, las comunidades de trabajadores

cuya naturaleza podría reforzarse para que sean el primer círculo de contención de usos maliciosos.

Y, finalmente, la trasparencia algorítmica también se presenta como un antídoto recurrente ante las amenazas reales. Transparencia en el caso de las relaciones entre sector tecnológico y armamentístico; en el caso de la gestión algorítmica del trabajo; o en el caso de las herramientas que gestionan la protección social.

De esta manera, las investigadoras parecen proponer una receta en la que abunda la creatividad, para cambiar el paradigma e imaginar nuevas maneras de gobernanza; regulación, para garantizar el respeto de los derechos y evitar que la ciudadanía pierda el control de la tecnología en favor de intereses particulares; participación, para diversificar las miradas y propiciar un desarrollo adaptado a las diferentes necesidades de las personas y respetuoso con sus particularidades; y transparencia, para ejercer un control ciudadano continuo y construir un modelo sostenible en el tiempo.