

Opresión, desigualdad e inteligencia artificial*

Yolanda Martínez Suárez**

Recibido: 18.06.2024 — Aceptado: 16.07.2024

Titre / Title / Titolo

Oppression, inégalité et intelligence artificielle
Oppression, Inequality and Artificial Intelligence
Oppressione, disuguaglianza e intelligenza artificiale

Resumen / Résumé / Abstract / Riassunto

En un tiempo de profundos cambios en nuestra manera de vivir y convivir, como los operados por la digitalización actual, la filosofía política se presenta como una herramienta que nos permite aproximarnos a la revolución tecnológica de la inteligencia artificial desde una perspectiva crítica y emancipatoria. Desde una posición situada, en este artículo, se reflexionará sobre las desigualdades ante las que nos sitúa el cambio, apuntando las posibilidades. Para ello, se aplicarán las cinco categorías operacionales de Young (explotación, marginación, carencia de poder, imperialismo cultural y violencia) a la inteligencia artificial, lo que nos permitirá diagnosticar las injusticias, y señalar las posibles soluciones. Desde una perspectiva feminista se pondrá en cuestión la supuesta neutralidad de la IA y se analizarán los efectos, intencionados o no, de la digitalización de la sociedad. Concluyendo con la necesidad de proponer una serie de cambios que nos aproximen a una digitalización más inclusiva.

À l'heure des profonds changements dans la manière dont nous vivons et coexistons, tels que ceux induits par la numérisation actuelle, la philosophie politique est présentée comme un outil qui nous permet d'aborder la révolution technologique de l'intelligence artificielle d'un point de vue critique et émancipatoire. À partir d'une position située, cet article réfléchira aux inégalités auxquelles le changement nous confronte, en soulignant les possibilités. À cette fin, les cinq catégories opérationnelles de Young (exploitation, marginalisation, désresponsabilisation, impérialisme culturel et violence) seront

* Esta investigación ha contado con el apoyo del Proyecto Reproducción Biológica, Reproducción Social y Esfera Pública (PID2020-115079RB-I00) financiado por MCIN/AEI/10.13039/501100011033

** Universidade de Santiago de Compostela

ont appliquées à l'intelligence artificielle, ce qui nous permettra de diagnostiquer les injustices et d'indiquer des solutions possibles. Dans une perspective féministe, nous questionnerons la neutralité supposée de l'IA et analyserons les effets, voulus ou non, de la numérisation de la société. Nous conclurons sur la nécessité de proposer une série de changements qui nous rapprocheront d'une numérisation plus inclusive.

At a time of profound changes in the way we live and coexist, such as those brought about by current digitalisation, political philosophy is presented as a tool that allows us to approach the technological revolution of artificial intelligence from a critical and emancipatory perspective. From a situated position, this article will reflect on the inequalities that the change places us in front of, pointing out the possibilities. To this end, Young's five operational categories (exploitation, marginalisation, disempowerment, cultural imperialism and violence) will be applied to artificial intelligence, allowing us to diagnose injustices and point to possible solutions. From a feminist perspective, we will question the supposed neutrality of AI and analyse the effects, intended and unintended, of the digitalisation of society. It will conclude with the need to propose a series of changes that will bring us closer to a more inclusive digitalisation.

In un'epoca di profondi cambiamenti nel nostro modo di vivere e coesistere, come quelli portati dall'attuale digitalizzazione, la filosofia politica si presenta come uno strumento che ci permette di affrontare la rivoluzione tecnologica dell'intelligenza artificiale da una prospettiva critica ed emancipatoria. Da una posizione situata, questo articolo rifletterà sulle disuguaglianze che il cambiamento ci pone di fronte, evidenziandone le possibilità. A tal fine, le cinque categorie operative di Young (sfruttamento, emarginazione, esautorazione, imperialismo culturale e violenza) saranno applicate all'intelligenza artificiale, consentendoci di diagnosticare le ingiustizie e di indicare possibili soluzioni. Da una prospettiva femminista, metteremo in discussione la presunta neutralità dell'IA e analizzeremo gli effetti, voluti e non voluti, della digitalizzazione della società. Si concluderà con la necessità di proporre una serie di cambiamenti che ci avvicinino a una digitalizzazione più inclusiva.

Palabras clave / Mots-clé / Keywords / Parole chiave

Opresión, emancipación, algoritmo, contexto, violencias, Young, Haraway.

Oppression, émancipation, algorithm, contexte, violence, Young, Haraway.

Oppression, emancipation, algorithm, context, violence, Young, Haraway.

Oppressione, emancipazione, algoritmo, contesto, violenza, Young, Haraway.

1. Introducción

En el prólogo a la *Ontología del teléfono Móvil* de Maurizio Ferraris, Umberto Eco nos recuerda que «No siempre la transformación coincide con la emancipación» (Eco, 15). Han pasado ya dieciocho años desde que el semiótico italiano escribiera esta advertencia y, en este tiempo, la digitalización ha avanzado hasta el punto de provocar cambios profundos en nuestra manera de vivir y convivir. La gran crisis de 2008 supuso un punto de inflexión en nuestra conciencia de la interacción con las tecnologías. Las reverberaciones de la cita de Eco se corporizaron desde el colapso de Lehman Brothers y todo lo que vino después: la precarización del trabajo en la línea de la feminización que ya apuntaba Haraway (1995) en los noventa, las protestas de la ciudadanía, encabezadas por el activismo climático y la revolución de los jóvenes en la segunda década de los 2000, la pandemia de Covid-19, la guerra en Ucrania, el conflicto de Israel-Palestina, etc. Las tensiones sociales ante las crecientes desigualdades emergieron demandando respuestas, y pasando a primer plano del debate el problema de las injusticias y la igualdad. Como pone de manifiesto Lucía Velasco, la puntilla de la crisis asendada por la pandemia hizo evidente que: «la automatización, la inteligencia artificial y la robótica prometen un mayor crecimiento económico, pero también pueden incrementar la desigualdad en y entre las naciones» (18).

Las desigualdades eclosionaron y no solo entre los agentes políticos de primer orden, sino también entre los grupos poblacionales divididos por diferentes categorías de pertenencia (sexo-género, raza, clase, origen, formación, etc.) y entre los propios individuos.

En su diagnóstico, Velasco (2021), apunta las tres grandes transformaciones que trae la tecnología: la automatización de los procesos, el boom de las plataformas digitales y la digitalización masiva de desarrollos, que comportan respectivamente, para el mundo laboral, la sustitución de humanos por máquinas en la realización de determinadas tareas, el reemplazo de la interacción física por la virtual y la reorganización de los modelos laborales. Antes de analizar sus implicaciones en lo que a conformación de las desigualdades se refiere, voy a ampliar el radio de la esfera laboral a la política. Si extendemos a esta los factores de la tecnologización de la sociedad podríamos añadir otras tres grandes transformaciones: la minería de datos, la formación e información para la toma de decisiones y el control y dirección de procesos políticos mediante la automatización. Estas transformaciones tienen incidencia sobre la forma de entender la libertad, la participación y la responsabilidad política, la igualdad, la justicia y la autoridad o el propio poder político.

2. La perspectiva filosófico-política: un lugar estratégico ante la transformación

Si bien existen campos específicos de reflexión como la filosofía de la tecnología, que acierta en comprender la tecnología no solo como un medio sino también como un fin (Coeckelbergh, 2023), urge una reflexión desde la filosofía política. La motivación es incontestable: las tecnologías digitales afectan a todos los aspectos de la vida (Velasco, 2021), lo que tiene consecuencias sobre el vivir bien. Dicho con Mark Coeckelbergh (17): «la IA es política de cabo a rabo». Ahora bien, a pesar de esta evidencia, las aproximaciones a la IA están realizándose sobre todo desde una perspectiva ética, y no filosófico-política. Así, por ejemplo, la moratoria que invocan los magnates de Silicon Valley

(22/03/2023) está focalizada en las consecuencias individuales de los avances tecnológicos, haciendo más necesaria que nunca una recuperación de la ontología de la tecnología que, al modo de la que hizo Ferraris con el móvil, nos sitúe ante el carácter profundamente social y político de las tecnologías, ahora en su versión digital de última generación. Los efectos sociopolíticos de la IA y la robótica se diluyen en los relatos contemporáneos como azucarillos en el café. Urge apuntar hacia el dulzor de este elemento para poder obtener una radiografía certera, una historia o planteamiento del potencial y actualidad de las tecnologías, que no deje todo el sabor en manos del grano de café y el agua, ignorando ingredientes fundamentales, como la altitud o el pH del suelo donde crece y brota, los agentes que lo siembran y recolectan, la cafetera o lo que se ha tomado antes. Sobre todo, habida cuenta de los ritmos de vértigo que tales tecnologías conllevan. La reflexión ética y política va necesariamente rezagada respecto de las revoluciones digitales (Eubanks, 2021) y así lo recoge la solicitud de moratoria, generada por la alarma ante la velocidad de la evolución de la IA y el poder poco comprensible, predecible y controlable de esas nuevas mentes digitales que ya están aquí, según su propia justificación. Concretamente, en la carta abierta para pausar los experimentos de IA más poderosos se indica:

AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research and acknowledged by top AI labs. As stated in the widely-endorsed Asilomar AI Principles, Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources. Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control. (Bengio et al, 2023).

Si bien no vamos a detenernos en las aproximaciones a la IA desde la perspectiva ética, y por lo tanto no desgranaremos la propuesta de moratoria, sí conviene destacar el carácter de inevitabilidad y de ritmo propio de la IA, incluso frente a la voluntad de sus propios creadores, que el texto alimenta y que se refleja en la siguiente senten-

cia: «Should we risk loss of control of our civilization?». Para responder a la pregunta cabe situar antes al sujeto de esta. ¿Quiénes son o somos los que deberíamos evitar la pérdida de control de nuestra civilización? Recordemos que la moratoria ha sido promovida entre otros por magnates de la tecnología con participación directa en sus desarrollos, como Elon Musk, propietario inicial de la empresa que creó Chat GPT, Steve Worniak, fundador de Apple o Yoshua Bengio, de Element AI¹, lo que nos recuerda al discurso tecnomilenarista de Vernor Vinge y su profecía, enunciada en los noventa, sobre la singularidad tecnológica. Alimentar la lógica dicotómica que nos sitúa como humanos naturales frente a la artificialidad de las tecnologías puede entrañar el riesgo de la no asunción de responsabilidades (de Salvador y Martínez, 2023) por parte de los promotores y entrenadores de la IA que, como observamos en la carta abierta referenciada, apelan a la responsabilidad ética humana pero no asumen un rol particular en la misma, en tanto creadores.

Frente a esos discursos dicotómicos, nos resulta de mayor utilidad aquí la propuesta de Donna J. Haraway del *cyborg* como elemento híbrido entre lo natural y lo artificial, sensible a las categorías de opresión y compatible con la perspectiva política, que podría abrir la puerta a las posibilidades.

Un cyborg es un organismo cibernético, un híbrido de máquina y organismo, una criatura de realidad social y también de ficción. La realidad social son nuestras relaciones sociales vividas, nuestra construcción política más importante, un mundo cambiante de ficción. Los movimientos internacionales feministas han construido la «experiencia de las mujeres» y, asimismo, han destapado o descubierto este objeto colectivo crucial. Tal experiencia es una ficción y un hecho político de gran importancia. La liberación se basa en la construcción de la conciencia, de la comprensión imaginativa de la opresión y, también, de lo posible. (Haraway, 253).

La propuesta de Haraway nos recuerda la parte humana de la IA, contrarrestando así el determinismo y apelando

¹ Varios autores se han mostrado críticos con el «teatro de la ética» de la industria tecnológica: un intento de lavado de cara mediante la invocación de la ética» (Paniagua, 186) y han alertado del riesgo de que «los poderosos lobistas de las grandes empresas de tecnología acaben marcando las reglas» (Paniagua, 269), señalando varios casos previos a la carta de la moratoria de 2023.

do a la asunción de responsabilidad, a la vez que invoca el carácter construido de lo humano, y de la política por extensión. Dicho con Monasterio: «de lo que no nos hemos dado cuenta es que a la vez que hemos moldeado a las máquinas, las mismas máquinas nos han moldeado a nosotros» (Monasterio, 260). Los cruces entre natural y artificial, humano-máquina, son iluminados por la metáfora de Haraway, resistiéndose a otros enfoques menos responsables y críticos. Como aquellos que al focalizar la atención sobre los usos incurren en desviar la atención sobre la propia herramienta, lo que, en última instancia, puede contribuir a reforzar el discurso de la neutralidad de la tecnología, o de ausencia de ideología de sus creadores y procesos de conformación. Con este tipo de declaraciones, se invisibiliza la autocrítica, que Vinge sí asumió al finalizar su relato distópico con la afirmación: «Y, sin embargo, nosotros somos sus iniciadores» (en O'Connell, 86). Esa aproximación compromete la responsabilidad humana en el proceso tecnológico, elemento de reflexión obligada desde una perspectiva filosófico-política.

De nuestra capacidad para escrutar las tecnologías depende la posibilidad de minimizar los daños y reenfocar el proceso hacia un mundo más justo e igualitario. Las posibilidades de la revolución son infinitas². Especialmente desde la emergencia del dispositivo conversacional de IA, Chat GPT, en noviembre de 2022, se han multiplicado los discursos, tanto alarmistas como tecno-optimistas, que señalan el fin de la humanidad o el progreso sin límites de una humanidad mejorada. Pero solo tras la voluntad de mirar con acierto se puede enfocar de un modo realista y práctico la cuestión. La clave, como siempre, está en hacia donde mirar. Y más en contextos donde se reflexiona a contrarreloj. Las tesis de la neutralidad de las tecnologías apuntalan la implantación de los avances digitales, como la IA y la robótica, en una dirección de refuerzo del statu quo, sin cuestionar la base sobre la que se edifica y que reproduce. En una línea opuesta, aquí, nos

² Como recuerda Lucía Ortiz de Zárate (2023), los sistemas de IA son un conjunto diverso de tecnologías que, si bien tienen en común su funcionamiento a través de algoritmos inteligentes, tienen diferentes aplicaciones y finalidades. Su potencial transformador sin precedentes le viene dado, precisamente, por el uso de los algoritmos inteligentes y las grandes cantidades de datos.

aproximaremos desde las tesis críticas con la revolución digital, en tanto esta consolida e incrementa las desigualdades, las injusticias y la vulnerabilidad de determinados grupos oprimidos. Nos detendremos también en la capacidad de transformación que tienen las tecnologías³, y que ya Haraway destacó en los ochenta en su *Manifiesto para Cyborgs*. Porque «La lucha política consiste en ver desde las dos perspectivas a la vez, ya que cada una de ellas revela al mismo tiempo tanto las dominaciones como las posibilidades inimaginables desde otro lugar estratégico» (Haraway, 263). La reflexión filosófico-política sobre la IA es, pues, un enfoque necesario para abrir la factibilidad de una transformación emancipadora. Desde esta perspectiva podemos intentar imaginar esas posibilidades de las que habla Haraway siguiendo la advertencia de Eco. Al identificar las tensiones y los problemas presentes en la transformación del mundo, que las tecnologías ya han mostrado inevitable, la Filosofía Política permite aproximarla a su carácter emancipador.

3. La posición situada: contexto e identificación de las opresiones

Dentro de los posibles enfoques que se dan desde la perspectiva política, aquí tomaremos la teoría crítica feminista como marco para abordar las grandes cuestiones ante las que nos sitúa la necesidad de vivir juntos, en la era de la revolución digital. Y lo hacemos con la «convicción de que el feminismo comporta la perspectiva privilegiada, no exclusiva ni excluyente, para reflexionar sobre la justicia y la igualdad en el mundo contemporáneo» (Agra, 14). Así, frente a los enfoques filosóficos que parten de la neutralidad de la tecnología y de los algoritmos, procederemos poniendo en cuestión la supuesta neutralidad de la IA y analizando los efectos, intencionados o no, de la digitalización de la sociedad.

³ Coincidimos con Ortiz de Zárate, en que la inteligencia artificial es «un grupo de tecnologías disruptivas capaz de provocar cambios profundos en nuestras sociedades» (Ortiz de Zárate, 7).

Los nuevos avances tecnológicos suelen ir acompañados de una forma peligrosa de pensamiento mágico: la curiosa presunción de que una revolución de nuestras herramientas conllevará inevitablemente borrón y cuenta nueva con el pasado. La metáfora del asilo digital pretende ni más ni menos que oponerse a este borrado de la historia y del contexto al hablar de tecnología y desigualdad. (Eubanks, 219).

Para aproximarnos a la IA desde un enfoque filosófico-político crítico, que permita una transformación emancipadora, debemos tener en cuenta varios elementos. Como señala Eubanks en la cita anterior, la clave está en la historia y el contexto. Así, debemos reparar en el contexto de creación y en el de implementación de las herramientas digitales, las propias herramientas de alta tecnología, el lenguaje, los datos y los sujetos – que las crean, seleccionan y reciben. Parafraseando a Harold Lasswell (1985), quien pensó la formulación original para el proceso comunicativo de relación entre productores y audiencia, habría que preguntarse quién crea qué, para quién, por qué medio y con qué efectos⁴. Esto es, y salvando las diferencias con los medios de comunicación clásicos, habrá que prestar atención no solo a la investigación de «análisis de control» (los programadores de la herramienta, los diseñadores de los algoritmos, pero también los sujetos que recopilan y seleccionan los datos, las personas que entrenan los modelos), al «análisis de contenido» (los mensajes y el lenguaje), al «análisis de medios» (los propios modelos de IA, y los procesos de diseño, creación y selección y entrenamiento), al «análisis de audiencias» (los usuarios) sino también al «análisis de los efectos» intencionados o no, que tiene la IA en nuestra vida en común.

En este texto, bajo la premisa de la necesaria revisión de la relación de los humanos con las máquinas, nos centraremos en los efectos, y para ello partiremos de los sujetos, sin dejar por ello de mencionar el resto de los elementos: fundamentalmente datos, algoritmos y hardware (Ortiz, 2023). Ya que los sujetos, bajo la óptica política, son inseparables de su historia y red de relaciones. Partiendo de la base de la construcción social de la realidad (Berger & Luckman, 2001) y del conocimiento situado (Haraway, 1995), una aproximación crítica al proceso de modelización exige

que se describa su contexto. En este sentido cabe destacar, como recoge Coeckelbergh (2023), varias líneas de análisis, en el marco temporal que vinculan el pasado con el presente y el presente con el futuro. Los enfoques históricos permiten explicar los procesos de opresión y su reproducción, tanto desde la óptica que sostiene que las formas de opresión actuales son la continuación de formas históricas de opresión y prácticas discriminatorias concretas, como desde la perspectiva que mira hacia las injusticias presentes y alerta sobre su posible y probable transformación en opresiones futuras de mayor alcance, sobre todo debido al potencial difusor⁵ de las tecnologías digitales. Entendemos opresión en el sentido amplio y plural que defiende Iris Marion Young, en *La Justicia y la política de la diferencia* (2000), al categorizarla bajo la teoría de las cinco caras de la opresión. De este modo, y como la propia autora pretende, se evitan las exclusiones, simplificaciones y reduccionismos que dificultan el diagnosticar la condición de oprimidos de determinados individuos o grupos. Young señala que «en tanto grupo las mujeres están sujetas a la explotación en razón del género, a la carencia de poder, al imperialismo cultural y a la violencia» (Young, 112).

Siguiendo la visión de esta autora podríamos aplicar la categorización de las cinco caras de opresión a la IA y diagnosticar así los sesgos y las injusticias que esta impone al colectivo de las mujeres, en el que me voy a centrar, en el nuevo contexto.

He propuesto las cinco caras de la opresión -explotación, marginación, carencia de poder, imperialismo cultural y violencia- como la mejor manera de evitar tales exclusiones y reducciones. Dichas formas de opresión funcionan como criterios para determinar si individuos y grupos están oprimidos, más que como una teoría completa sobre la opresión. Creo que estos criterios son objetivos. Ellos proporcionan un medio para refutar la creencia de alguna gente de que su grupo está oprimido cuando no lo está, así como un medio para persuadir a otras personas de que un grupo está oprimido cuando dudan acerca de ello. Cada criterio puede ser operacionalizado; cada uno de dichos criterios puede ser aplicado a través de la evaluación de la conducta observable, las relaciones de estatus, las distribuciones, los textos y otros elementos culturales (Young, 111).

⁴ ¿Quién dice qué en qué canal a quién y con qué efecto? (Lasswell, 1979).

⁵ Como apunta Ortiz (7), de la IA se espera que pueda aumentar «los niveles de eficacia, eficiencia y personalización de servicios (públicos y privados) ya existentes».

Las categorías son operacionales y al permitirnos observar «las similitudes y superposiciones que se verifican en las formas de opresión de grupos diferentes» (Young, 111) contribuyen, en primera instancia, a identificar los grupos oprimidos, esto es: las injusticias, si las hubiera y, en segundo lugar, a corregirlas puesto que, al visibilizar los nexos históricos y el carácter estructural de las formas de opresión concretas, abren la puerta a su eliminación. Así, su aplicación a la IA podría ayudarnos a avanzar en el carácter emancipador de estas tecnologías.

Para comenzar, tendríamos que detenernos en la explotación, la primera categoría propuesta por Young, a partir de la contratación de mujeres en las periferias para entrenar a la IA o para realizar algunas de las tareas más ingratas y duras del proceso de modelización. Muchas de las personas encargadas de visionar durante horas textos e imágenes, que atentan contra la sensibilidad humana, para darle indicadores de toxicidad a modo de ejemplo a la IA, enseñándole qué contenidos debe excluir, son mujeres, mal pagadas, sin derecho a servicios psicológicos y de cuidado para superar el estrés postraumático que le provoca su tarea⁶. En esta categoría, además de a las *ghost workers*, habría que incluir también todo el trabajo no remunerado sobre el que se erige la producción y recopilación de datos que nutre la IA, que en la era de la feminización del trabajo tiene sello femenino. Como ya advirtió Haraway hace más de tres décadas: «mientras la oficina automatizada se convierte en la norma incluso en países con abundante oferta de trabajo, la feminización del trabajo se intensifica» (Haraway, 287). Feminización que ya la filósofa norte-americana ligaba a subempleo estructural y vulnerabilidad. Rasgos de la precariedad acentuados y globalizados ahora por el efecto de la pandemia de Covid-19. Y de la crisis consiguiente, proceso que ha contribuido al vaciamiento de los centros, esto es de la clase media, con las consecuencias que de ello se derivan para el contrato social (Velasco, 2021). Y, ampliamos

⁶ *Time* ha publicado en enero de 2023 un reportaje titulado «Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic», documentando las condiciones de los trabajadores de Kenia, Uganda e India que contribuyen a que Chat GPT4 no reproduzca los fallos de su antecesor: GPT3 en lo que a racismo, sexismo y violencias se refiere. Véase: <https://time.com/6247678/openai-chatgpt-kenya-workers/>.

aquí, para el contrato sexual, que diría Pateman (2020). El sesgo de género se refleja en las cifras de ocupación por sexo-género de los empleos más «tele-trabajables», masculinos por adscripción. Precisamente, Young propone la corrección de las instituciones y prácticas, así como de la división del trabajo para eliminar el desequilibrio entre lo que ella denominó beneficiado/explotada, y que en este caso serían los programadores, un mundo de hombres⁷, y las reproductoras y educadoras del algoritmo explotadas en los centros de computación del mal llamado tercer mundo. Ellos son los programadores y testadores, quienes toman las decisiones, así como los clientes de los productos de IA y ellas las que limpian el algoritmo. Una vez más se hace buena la expresión del hecho a imagen y semejanza de. Así, «La estructura de los ciberespacios y su funcionamiento, tienen mucho que ver con quien los crea y el sector tecnológico es eminentemente masculino y heterogéneo» (Velasco, 42-43).

Continuando con la sumaria aplicación de la propuesta de Young a la IA, la segunda cara de la opresión, la marginación, emerge en la forma de interacción con las tecnologías, dando lugar a la exclusión digital. Los grupos con privaciones de oportunidades y capacidades tecnológicas cuentan en sus filas con un elevado porcentaje de mujeres. Solo hay que hacer una aproximación con perspectiva de género⁸ a las brechas de acceso y uso de las tecnologías, en general, y de las más avanzadas en particular, y repasar la interseccionalidad de marcas o acumulación de brechas (Martínez, de Salvador y de Salvador, 2015; Martínez y de Salvador, 2019), acentuadas por la pandemia (Velasco, 2021). Las mujeres «tenemos de facto menos acceso a la tecnología digital y menos participación en los campos que ahora triunfan como la ingeniería o las matemáticas»

⁷ El extremo se presenta en las empresas de IA formadas exclusivamente por hombres. Este es el caso de la nueva xAI, compañía creada por Elon Musk, donde las mujeres son formalmente excluidas del objetivo de «comprender la verdadera naturaleza del universo». Para más información véase <https://www.lavanguardia.com/tecnologia/20230714/9106465/elon-musk-presenta-xai-empresa-inteligencia-artificial-pmv.html>.

⁸ Queda para una aproximación posterior un análisis pormenorizado con perspectiva de género de las cinco áreas de competencia digital que establece la Comisión Europea, a saber: la alfabetización en información y datos, la comunicación y colaboración, la creación de contenidos digitales, la seguridad y la resolución de problemas. Si bien los resultados de estudios anteriores en la misma línea muestran unas coincidencias que evidencian el refuerzo de las exclusiones de género más que su corrección.

(Velasco, 112), ya que como recuerda Miren Gutiérrez (52): «menos de una de cada cinco personas que se gradúan en informática son mujeres y la industria de algorítmica emplea una proporción aún menor de mujeres que el resto del sector tecnológico». En este sentido, para paliar la exclusión, se han propuesto soluciones como la formación en tecnologías a los miembros de los grupos históricamente minorizados. Su participación en el proceso tiene una serie de consecuencias correctoras sobre la exclusión de las mujeres del espacio público-político, al permitir su incorporación como usuarias, lo que puede hacer audibles sus voces. Si, para Young, «las personas marginales son aquellas a las que el sistema de trabajo no puede o no quiere usar⁹» (Young, 94), un análisis de la IA en clave colonialista podría señalar las tensiones y conflictos que se presentan y cuyo resultado ofrecería una radiografía de los grupos excluidos por su falta de competencias digitales, y las oportunidades o ejercicio de libertades que estas personas perderían. En este sentido, varios relatos apuntan a la discriminación que pueden, y de facto ya empiezan a sentir, las personas excluidas del uso de la IA frente a las que sí pueden incorporar sus desarrollos para mejorar sus vidas o simplemente para preservarlas (mediante el acceso al trabajo). Como sintetiza Velasco: «El mercado laboral se está rompiendo, segmentando, generando realidades de primera con empresas de trabajadores cualificados; y vidas de segunda, más vinculadas a las pymes, con trabajos poco estables, menos productivos y abocados a la precariedad o al desempleo» (Velasco, 47). Hemos visto en la anterior categoría, como la ausencia de formación puede provocar un desequilibrio en la participación como programadoras de los algoritmos y una sobre-presencia en los puestos marginales, las que hemos llamado limpiadoras del algoritmo. En esta cara de la oposición cabría reflexionar también en el contexto tecnológico sobre «los derechos básicos a la privacidad, el respeto y la elección individual» (Young, 96). Porque si bien las mujeres acceden menos como sujetos a la IA, lo hacen más como objeto. Como estereotipos negativos,

⁹ Entiéndase este «usar» en sentido positivo, en tanto agentes, y no como objetos a explotar, analizado en la primera categoría y que veremos a partir de la reflexión sobre los datos.

lo que se refleja en los datos tóxicos que las compañías intentan limitar, pero también en patrones subterráneos de oposición y otras manifestaciones de los sesgos de género.

Los datos¹⁰ es uno de los tres elementos principales de la IA y como tal uno de los mayores cajones de construcción de sesgos de género. Así lo apunta Ortiz de Zárate (2023), quien señala la sobrerrepresentación de los hombres en las bases de datos sobre las que se edifica la IA, lo que refleja la exclusión de las mujeres de la esfera pública mediática, así como de los espacios de poder. Pero además de la infrarrepresentación de las mujeres, cabe prestar atención a su re-presentación negativa. La dinámica de trabajo de la compañía de «IA ética» Sama prueba la no neutralidad de la poderosa tecnología conversacional en este segundo sentido. Los empleados de esta empresa, socio de subcontratación de OpenAI en Kenia, trabajan etiquetando contenidos de «abuso sexual», «discurso de odio» y «violencia» que afecta fundamentalmente a las mujeres. La necesidad de limpiar de toxicidad¹¹ los datos que puede acabar ingiriendo la IA nos sitúa ante varias cuestiones, alguna de las cuales ya han sido apuntadas aquí: ¿Quién define la toxicidad? ¿Qué entra dentro del cajón toxicidad? ¿Quién limpia y con qué consecuencias para su salud? ¿Bajo qué parámetros limpia? ¿Cómo se gestionan esos residuos? Preguntas que, en último término, se enmarcan en la tensión entre la autoridad y la verdad, de la que tradicionalmente se ha venido ocupando la filosofía política, lo que también arroja reflexiones sobre las relaciones jerárquicas de poder y las opresiones.

La tercera categoría presentada por la filósofa norteamericana, la carencia de poder, íntimamente ligada con la anterior, toma cuerpo en la era de la IA en el debate político entre tecnocracia y democracia¹². Abundando en la categoría anterior, Young divide a los grupos sociales oprimidos y excluyentes, en profesionales y no profesionales, lo que trasladado a la cuarta revolución digital serían esos «perfiles de alta cualificación, especialmente tecnoló-

¹⁰ Ortiz de Zárate (2023) señala los datos, los algoritmos y el hardware como los tres principales elementos que integran la IA y en cuyo seno se localizan los sesgos de género.

¹¹ Según documenta *Time*, Sama tiene contratos con Google, Meta o Microsoft, lo que desmiente la excepcionalidad de la práctica.

¹² Véase el capítulo que Coeckelbergh (2023) dedica a esta discusión.

gicos y que puedan tele-trabajar» (Velasco, 55) frente al «nuevo proletariado digital que está en riesgo de pobreza» (Velasco, 66). O dicho con Petrone (241): «nos encontramos ante el nacimiento de nuevas formas de trabajo que muchas veces absorben a aquellos trabajadores [y, fundamentalmente, trabajadoras] que no pueden encontrar otro empleo (más cualificado)». Como establece Young, los carentes de poder, tienen «poca a ninguna autonomía laboral, dispone[n] de pocas oportunidades para la creatividad y no utiliza[n] casi criterios propios en el trabajo, no tiene[n] conocimientos técnicos, ni autoridad, se expresa[n] con dificultad especialmente en ámbitos públicos o burocráticos, y no impone[n] respeto» (Young, 99). La falta de estatus tecnológico comporta una serie de consecuencias que redundan en la conformación de desigualdades en el marco de la lógica estructural que estudió Young. De ahí que la división del trabajo sea la causa y también la solución de esta línea de injusticias. En todo caso no debemos leer la falta de autonomía solo en términos individuales sino también en términos políticos, en tanto amenaza a la democracia en la línea del capitalismo de la vigilancia, la concentración de poder y el control paternalista y autoritario (Coeckelbergh, 2023). Ya que, actualmente, «los derechos de las mujeres dependen [también] de la justicia algorítmica» (Gutiérrez, 58).

Como hemos visto, al operacionalizar las dos primeras caras de la opresión, propuestas por Young, el hecho de que la IA suponga una falta de poder para las mujeres, esto es mayoritariamente las excluya de su gestión, no implica una falta de relación. La IA necesita a las mujeres por cuestiones de estrategia, «mercado» y «control», diríamos, extendiendo la tesis de Petrone (2022) de la geopolítica a los grupos. Las mujeres, como los grupos oprimidos, no interesan a la IA cuando se trata de discutir los temas y tomar las decisiones sobre los modelos, pero sí son necesarias como mano de obra barata para alimentar, cuidar y limpiar los datos, es decir como cuidadoras, o limpiadoras, de la IA.

Frente a las tres categorías anteriores, cuyo origen es la división social del trabajo, la cuarta cara de la opresión no se subsana con una redistribución de las tareas sino con la corrección de la perspectiva de la sociedad, que como de-

nuncia Young, excluye a otros grupos, estereotipándolos e invisibilizándolos. El imperialismo cultural, en el contexto de la IA, afecta a los colectivos no solo mediante las representaciones sino también y sobre todo en el nivel de las ausencias. Si bien por una parte se denuncian las presentaciones estereotipadas y el uso de lenguaje racista o sexista, que las empresas de IA ética se afanan en localizar y eliminar, por otra parte, se apunta la falta de representación de determinados colectivos en determinadas posiciones, en la medida en que esta reproduce el sistema de injusticias y desigualdades. Piénsese por ejemplo en la infra-representación de las mujeres en posiciones de poder político y científico o en la práctica ausencia de personas no occidentales en los modelos. Varias voces críticas empiezan a hablar de “colonialismo digital” o neocolonialismo para referirse a las relaciones que reedita la IA de explotación de Occidente sobre los sures, apuntando a las empresas de entrenamiento en África frente a las sedes de la programación en Silicon Valley (Coeckelbergh, 2023) o a las nuevas formas de explotación (Petrone, 2022). En el discurso positivo sobre los sujetos que representa/construye la IA está incluida la contracara de la representación negativa de los «otros», como señala Teun A. Van Dijk (2008) en su análisis del discurso que da cuenta de la dominación étnica y el racismo discursivo, entre otras manifestaciones de la opresión. Si las experiencias de vida de determinados grupos aparecen fagocitadas por la IA solo en determinados roles, asociados a conducta criminal y marginalidad, determinadas interpretaciones propias de grupos oprimidos no conformarán nuestro futuro. Y el grupo dominante, los grandes programadores y magnates de la tecnología, seguirán construyendo «las diferencias que exhiben algunos grupos como carencia y negación» (Young, 2000: 105). Los modelos de IA suelen y de hecho aprenden de casos anteriores para generar modelos automatizados. Su dependencia de secuencias de datos históricos, construidos por los grupos dominantes según la lógica del nosotros-otros enunciada, contribuye a perpetuar los estereotipos, los sesgos y las desigualdades estructurales, reificando así las opresiones.

Una IA inclusiva tendría que ir más allá de la revisión de los discursos explícitos de odio y la expresión de las

violencias físicas y verbales, para detectar las exclusiones y abrir así la puerta a la inclusión de los otros para que puedan tener experiencias de vida propias y autónomas. Los sesgos además de en el algoritmo y los datos van insertos también en el lenguaje y en los equipos y las propias compañías, así como en el hardware (Ortiz, 2023). La opresión algorítmica afecta a la capacidad de tomar decisiones y configura las relaciones de poder a nivel local y global. De ahí que urja preguntarse ¿Con qué datos se entrenan las inteligencias de automatización que toman las decisiones? ¿De qué sesgos adolecen esos datos? ¿Sobre quien toma decisiones la IA? ¿Hay grupos sobrerrepresentados en las sanciones gubernamentales o corporativas privadas? ¿Quién testa los modelos? ¿Cómo se corrigen los sesgos? ¿Quién establece el nivel de justicia e igualdad? ¿Cómo entiende o define la igualdad y la justicia? ¿Convendría aplicar la discriminación positiva a los algoritmos?

La corrección del contexto social que ampara la amenaza, restando libertad y dignidad, es la propuesta de solución, en la línea de la justicia, que Young plantea para su quinta y última cara de la opresión. La violencia, caracterizada como posibilidad, recordando al metafórico estado de guerra de Hobbes, ahora circunscrito a los miembros de determinados grupos y no a todos los hombres en el estado de naturaleza, es un rasgo sistemático, social, cuyos agresores y víctimas serían en la esfera de la IA, cuyos agresores y víctimas serían los mismos que en la esfera no tecnológica, contexto que sirve de base para la construcción de nuevas máscaras para viejos caracteres. La mera amenaza de la violencia del algoritmo es retratada detalladamente por Virginia Eubanks (2021) en su *Automatización de la Desigualdad*. Herramientas de tecnología avanzada para supervisar y castigar a los pobres. La politóloga norteamericana muestra como el asilo digital se adapta a la coyuntura particular de nuestro tiempo», pero sigue criminalizando a los mismos grupos poblacionales: los pobres y los colectivos racializados. El principal problema de esta categoría de opresiones es la falta de transparencia de los procesos y los modelos. El acceso a los algoritmos es muy oscuro lo que dificulta desentrañar los tentáculos y mecanismos que ponen en marcha y reproducen las violencias. Las consecuencias no intencionales de deter-

minados grupos o procesos hacen todavía más compleja su identificación. Esto es: «Reconocer el sesgo puede ser difícil cuando no hay referencias explícitas a criterios tales como el género, la raza, etc., en los datos de entrenamiento» (Coeckelbergh, 64). Cuando nos enfrentamos al universal, construido bajo la abstracción y la supuesta neutralidad –recuérdese aquí el hecho a imagen y semejanza de, antes mencionado–, es complejo identificar el sesgo y las consecuencias de la justicia e igualdad del algoritmo. Quizás, implementar el uso de herramientas como la ley de la inversión –que refuerza la posición situada, el contexto concreto–, o localizar sobre- e infrarrepresentaciones en el nivel de control y rastreo podría ayudar en la identificación de agresores/víctimas. Visibilizando así la dinámica en la que «los hombres tienen más probabilidades de ser autores, y las mujeres tienen más probabilidades de ser víctimas» (Velasco, 43). De este modo, las cinco caras de la opresión, de acuerdo con Young, serviría a su propósito de identificar injusticias, al ponerle rasgos, variables o marcas de pertenencia a grupos, a sus víctimas. Negando así la coincidencia estadística y apuntando el carácter estructural de las injusticias o desigualdades.

4. Consideraciones finales

Como hemos visto, la búsqueda de las conexiones causales y no meramente circunstanciales es fundamental para acertar en las cuestiones problemáticas que afectan al vivir juntos. Con el objetivo de reenfocar la cuestión, de mirar al contexto y el desarrollo histórico que subyace a toda construcción, en este texto, he apuntado hacia la articulación de los grandes debates de la política con las tecnologías digitales desde una perspectiva crítica feminista. Reparando los sesgos, mediante la implementación de la inteligencia social humana, se contribuye a la lucha contra las opresiones (Coeckelbergh, 2023). Las tensiones entre humano/máquina y entre la incertidumbre/cálculo son inevitables en la articulación entre política/tecnología. Porque, como ficciona Olga Ravn en la novela distópica *Los empleados*, los autómatas «Son más robustos, y la posibilidad de actualizar el programa hace que los

datos puedan ser almacenados y transferidos en enormes cantidades» (Ravn, 117), sin embargo «No comparto la postura, muy extendida entre varios de mis compañeros, de que la única solución eficaz sería suprimir la parte humana de la tripulación. A lo mejor son precisamente los humanos el componente de caos que mantiene vivo el mundo» (Ravn, 110).

Como he mostrado, el enfoque filosófico-político, desde una perspectiva realista, pero también desde una óptica normativa, se presenta como el lugar necesario para poder reflexionar sobre las tensiones entre la libertad y la automatización, entre la desobediencia y la aplicación, entre las emociones racionales y el cálculo. De un enfoque crítico bien afinado –que deposite la responsabilidad y la posibilidad de imaginación y cambio en manos de las personas y no del determinismo tecnológico– depende la identificación de las cuestiones que realmente están en juego, abriendo así la puerta a una transformación emancipadora. Hasta que llegue el momento de la singularidad absoluta (entendida esta como la capacidad de las máquinas de tomar decisiones y actuar de manera totalmente autónoma) si es que llega, se necesita a las personas para establecer el marco en el que convivir. Y para ello la aplicación de la teoría de las cinco caras de la opresión de Young a la IA se ha revelado especialmente útil para la identificación de las manifestaciones de la desigualdad de sexo-género. Si bien podría hacerse la misma aproximación a partir de cualquier otra categoría, desde la interseccionalidad, para testar la participación y sus efectos sobre determinados colectivos y sobre la propia sociedad en su conjunto.

Podemos concluir, parafraseando a Young, que las mujeres como grupo, así como los individuos que se enmarcan en la feminización del trabajo, están sujetas a la explotación en razón del género, a la marginación, a la carencia de poder, al imperialismo cultural y a la violencia en el contexto de la digitalización de la sociedad. Lo que exige activar estrategias de redistribución, reconocimiento y representación tales como las enumeradas en la operacionalización de las cinco caras al escenario de la digitalización y la IA, esto es: la revisión de la división social del trabajo en todas las ocupaciones relacionadas con el *big data*, con

especial atención a los *ghost workers*, la revisión de la corrección de ciertas prácticas e instituciones como la propia limpieza del algoritmo o las empresas solo masculinas, la redistribución de las tareas de diseño, testado y limpieza de la IA, la implementación de formaciones específicas en tecnologías para colectivos marcados que lime las brechas de acceso y uso y, sobre todo, la transformación en la perspectiva social y la corrección del contexto, para posibilitar la mayor participación en igualdad de condiciones, esto es para favorecer la representación y minimizar los efectos negativos de la revolución digital a la que asistimos, reenfocando, así, el proceso hacia un mundo más justo e igualitario. En tanto, como hemos visto, la participación de las mujeres, así como de todos los colectivos oprimidos, en el proceso, tiene una serie de consecuencias correctoras sobre la exclusión en el espacio público-político. Esto es, la injusticia y desigualdad digital y la general se retroalimentan. La clave de la solución está, como diría Donna J. Haraway, en la hibridación, en diluir las dicotomías. O, al menos, en establecer una distribución arbitraria de las personas en función de sus categorías de pertenencia a cada uno de estas posiciones. En definitiva, urge cuestionar las parejas que identifica Young: beneficiado/explotada, profesionales/no profesionales, nosotros/otras, agresor/víctima. O dicho en términos de *big data* e IA: diseñador/limpiadora de toxicidad, ingenieros informáticos/proletarias digitales, hombres blancos/mujeres racializadas, agresor/víctima. Solo así, habitaremos un mundo donde la revolución tecnológica se separe de las viejas y nuevas lógicas de opresión y traiga una transformación realmente emancipatoria.

Bibliografía

- Agra Romero, María Xosé ; *Olvidar a Clitemnestra? Sobre justicia e igualdad*. Santiago de Compostela: Universidade de Santiago de Compostela Editora, 2016.
- Berger, Peter L. y Luckmann, Thomas. *La construcción social de la realidad*. Buenos Aires: Amorrortu editores, 2001.

- Coeckelbergh, Mark. *La filosofía política de la inteligencia artificial*. Madrid: Cátedra, 2023.
- De Salvador Agra, Saleta y Martínez Suárez, Yolanda. «De injusticias algorítmicas y antilenguas digitales». *Periodismo, ciudadanía y política en el escenario digital*. Ed. María Isabel Míguez-González y Alberto Dafonte-Gómez. Madrid: Dykinson. 2023, pp. 178-195.
- Eco, Umberto «El teléfono móvil y la verdad». ¿Dónde estás? *Ontología del teléfono móvil*. Ed. Maurizio Ferraris. Barcelona: Marbot Ediciones, 2008, pp. 13–15.
- Eubanks, Virginia. *La automatización de la desigualdad. Herramientas de tecnología avanza-da para supervisar y castigar a los pobres*. Madrid: Capitán Swing, 2021.
- Bengio, Yoshua et al [en línea]: «Pause Giant AI Experiments: An Open Letter», en Future of Life Institute (22/03/2023), <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> Recuperado el 29 de febrero de 2024.
- Gutiérrez, Miren. «Algoritmos y desigualdad». *Revista de Occidente*, 502, 2023, pp. 47–59.
- Haraway, Donna J. «Manifiesto para cyborgs: ciencia, tecnología y feminismo socialista a finales del siglo XX». *Ciencia, cyborgs y mujeres. La reinención de la naturaleza*. Ed. Donna J. Haraway. Madrid: Cátedra, 1995, pp. 251–311.
- Lasswell, Harold. «Estructura y función de la comunicación en la sociedad». *Sociología de la comunicación de masas*. Ed. Miquel de Moragas Spá. Barcelona: Gustavo Gili, 1979, pp. 158–172.
- Martínez Suárez, Yolanda y de Salvador Agra, Saleta. «Digital snails?: Shuar women and mobile communication in Ecuador». *Gendered power and mobile technology: intersections in the Global South*. Ed. Caroline Wamala Larsson y Laura Stark. London/NY: Routledge, 2019, pp. 166-177.
- Martínez Suárez, Yolanda, de Salvador Agra, Saleta y de Salvador González, Xabier. «Triplemente marcadas: desconexiones comunicativas en la Amazonía sur ecuatoriana». *Cuadernos.info*, 36, 2015, pp. 89–107. <https://doi.org/10.7764/cdi.36.716>
- Monasterio, Aníbal. «Automatizando la toma de decisiones morales: inteligencia artificial y mejora humana». *Más que humanos. Biotecnología, inteligencia artificial y ética de la mejora*. Ed. Francisco Lara y Julián Savalescu. Madrid: Tecnos, 2021, pp. 255–281.
- O’Connell, Mark. *Como ser una máquina*. Madrid: Capitán Swing, 2019.
- Ortiz De Zárata Alcarazo, Lucía. «Sesgos de género en la inteligencia artificial». *Revista de Occidente*, 502, 2023, pp. 5–20.
- Paniagua, Esther. *Error 404 ¿Preparados para un mundo sin internet?* Madrid: Debate, 2021.
- Perrigo, Billy [en línea]: «Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic», en *Time* (18/01/2023), <https://time.com/6247678/openai-chatgpt-kenya-workers/> Recuperado 29 de febrero de 2024.
- Petrone, Francesco. «Nuevas tecnologías y viejas formas de opresión». *Robótica, ética y política. El impacto de la superinteligencia en el mundo de las personas*. E. Norbert Bilbeny. Barcelona: Icaria Editorial, 2022, pp. 239- 267.
- Ravn, Olga. *Los empleados*. Barcelona: Anagrama, 2023.
- Sanuy, Ada [en línea]: «Elon Musk presenta xAI, su proyecto de IA formado exclusivamente por hombres», en *La Vanguardia* (14/07/2023), <https://www.lavanguardia.com/tecnologia/20230714/9106465/elon-musk-presenta-xai-empresa-inteligencia-artificial-pmv.html> Recuperado el 29 de febrero de 2024.
- Van Dijk, Teu A. *Dominación étnica y racismo discursivo en España y América Latina*. Barcelona: Gedisa editorial, 2008.
- Velasco, Lucía. «No es “sólo un insulto”. La vida digital y la violencia contra las mujeres». *Re-vista de Occidente*, 502, 2023, pp. 34–46.
- Velasco, Lucía. *¿Te va a substituir un algoritmo?* Madrid: Turner Libros, 2021.
- Young, Iris Marion. *La justicia y la política de la diferencia*. Madrid: Cátedra, 2000.

