Capítulo 5

Fuck the algorithm: Navegando la promesa tecnológica y el impacto social de la IA

Marisol Flores-Garrido

https://doi.org/10.61728/AE24001052

La inteligencia artificial (IA) ha irrumpido con fuerza en tareas tan numerosas como diversas, suscitando una fascinación que puede oscurecer sus límites reales. Desde revolucionar el diagnóstico médico hasta simplificar nuestras tareas diarias, la IA no solo ha enriquecido nuestra vida cotidiana, también ha desatado optimismo y confianza en sus posibilidades. En ese sentido, podemos considerar a la IA como una tecnología carismática, que deriva gran parte de su poder a través de la posibilidad o promesa de acción: importa no solamente lo que sus herramientas son, sino la forma en que afectan nuestra imaginación con las promesas de lo que podrían hacer (Ames, 2019).

Los logros de la IA pueden conducirnos a su adopción precipitada en contextos donde su utilidad no es tan clara, persiguiendo utopías tecnológicas que ignoran las limitaciones de estas herramientas. Ante esta situación, es necesario cultivar una postura crítica tanto en los desarrolladores de sistemas de IA como en la sociedad en su conjunto, quien debe defender el derecho a decidir cómo y cuándo implementar estas tecnologías.

Por esta razón, se vuelve crucial reimaginar colectivamente usos y limitaciones de la IA. Nuestras expectativas y visiones sobre esta tecnología no solo guían la investigación y la inversión en proyectos, también validan sus aplicaciones y moldean nuestra actitud colectiva hacia ellas. Ajustar nuestras expectativas con una actitud crítica que distinga entre mitos y realidades puede complicarse por el hecho de que la IA existía en nuestro imaginario colectivo mucho antes de su formalización como campo de estudio. En consecuencia, muchos sistemas de IA surgen en un espacio ya habitado por numerosos mitos sobre sus capacidades. De alguna manera, las narrativas de autómatas e instrumentos exentos de debilidades humanas, que prometen superarnos apoyados en el poder de los cálculos numéricos y la razón absoluta, preceden a las herramientas mismas. Esto, combinado con estrategias desbordadas de mercadotecnia, puede dificultar la apreciación de los límites de la IA y de sus dimensiones políticas, sociales e, incluso, materiales.

Dada la promesa de la IA de insertarse en todos los dominios del conocimiento y, potencialmente, en cada aspecto de nuestra existencia, se vuelve imperativo fomentar una reflexión crítica sobre sus alcances. En este capítulo, propongo tomar como caso de estudio el episodio de un algoritmo utilizado en el Reino Unido en 2020 para ajustar calificaciones

estimadas de los exámenes *A-levels* (de Advanced Level qualifications). A través de un análisis teórico que recupera las nociones de sesgo algorítmico, complejidad y resistencia, se utilizan estos ejes para reflexionar sobre la necesidad de construir una mirada crítica hacia los sistemas de IA y de reconfigurar el discurso en torno a ellos para hacer evidentes sus posibilidades, sus limitaciones y las estructuras de poder que los hacen posibles.

El algoritmo de la Ofqual

En el año 2020, la pandemia de COVID-19 y las medidas de confinamiento impuestas por el gobierno británico enfrentaron a los estudiantes de bachillerato con una situación sin precedentes: la imposibilidad de realizar los exámenes A-levels, cruciales para el acceso a la educación superior. Esta situación obligó al sistema educativo a crear una estrategia que permitiera terminar el ciclo académico y evitara cualquier discontinuidad en el acceso a la educación universitaria.

Ante la urgencia de una solución, la Oficina de Regulación de Calificaciones y Exámenes (Office of Qualifications and Examinations Regulation u Ofqual, por sus siglas) propuso una medida extraordinaria: que los docentes realizaran una estimación de las calificaciones de cada estudiante, denominada "calificación evaluada por el centro" o CAG por su siglas en inglés, y establecieran un ranking entre quienes tenían estimaciones similares. Posteriormente, estas calificaciones proyectadas alimentarían un algoritmo diseñado para ajustar las notas basándose en el rendimiento histórico de cada centro educativo.

Específicamente, el algoritmo de la *Ofqual* estimaba las calificaciones de los exámenes considerando tres factores: (1) la distribución histórica de las calificaciones de los centros de los tres años anteriores (2017-2019); (2) la clasificación de cada estudiante dentro de su propio centro en una asignatura concreta, basada en las CAG; y (3) los resultados de exámenes anteriores de cada estudiante por asignatura. Con toda esta información, el algoritmo realizaba un análisis de la distribución de las notas a lo largo de los años, asignando calificaciones en función de la posición relativa de cada estudiante dentro de su centro. Por ejemplo, si un alumno se encontraba en la mitad inferior de la lista de clasificación, su nota sería aproxi-

madamente igual a la obtenida en años anteriores por otro alumno de la misma categoría. La intención era replicar, de manera artificial, la distribución de calificaciones de años anteriores, manteniendo consistencia con el desempeño histórico de las instituciones.

Preservar una cierta "normalidad" estadística era muy importante para la Ofqual, que partió de la premisa de que utilizar datos históricos para ajustar las calificaciones sería más preciso y más justo que confiar únicamente en las evaluaciones estimadas por docentes. Existía la percepción de que los profesores, por su cercanía y compromiso con el alumnado, podrían tender a una generosidad excesiva en sus calificaciones y conducir a una inflación de las mismas. Esto implicaría un aumento considerable en el número de estudiantes con calificaciones destacadas y, a su vez, desbordaría la capacidad de admisión de las universidades. En este contexto, el algoritmo se concibió como un mecanismo correctivo frente a lo que se consideraba un sesgo implícito en las CAG, buscando equilibrar un proceso esencialmente humanizado y, por ende, sujeto a las complejidades de la subjetividad docente.

No obstante, la táctica adoptada por la *Ofqual* rápidamente mostró fallas, en particular en el efecto desproporcionado que el algoritmo ejerció sobre los alumnos de escuelas públicas en relación con sus pares de instituciones privadas. La confianza en este sistema automatizado, fundamentada en la suposición de una posible sobrevaloración de las calificaciones por parte del profesorado, condujo a un resultado inesperado y perjudicial: una depreciación general de las notas finales en comparación con las CAG, afectando aproximadamente al 40 % de los estudiantes y perjudicando en mayor proporción a quienes provenían de una institución pública. El ajuste no solo truncó las posibilidades de acceso universitario de numerosos jóvenes, también intensificó las brechas ya existentes dentro del sistema educativo dejando en evidencia la fragilidad de un enfoque que, en su intento de buscar objetividad, terminó por profundizar las inequidades.

La determinación de calificaciones por parte del algoritmo desató una ola de manifestaciones en distintas regiones del país, que se extendieron durante varias semanas. Ante la creciente presión social, el 17 de agosto de 2020 el gobierno tomó la decisión de restablecer las calificaciones conforme a las estimaciones originales realizadas por los docentes, desestimando

los ajustes del algoritmo con la única excepción de aquellos casos en los que la nota asignada por el sistema automatizado superaba a la CAG. Esta medida representó un reconocimiento de las limitaciones y el impacto negativo de la dependencia en soluciones algorítmicas para asuntos trascendentes, como la educación, subrayando la necesidad de priorizar el juicio humano y de reconocer las limitaciones de predicciones basadas en datos.

Examinar este caso puede arrojar luz sobre aspectos fundamentales que deben tomarse en cuenta para desarrollar una perspectiva crítica ante la toma de decisiones mediada por algoritmos. En un contexto donde el uso de los sistemas de IA se expande rodeado de un discurso de exactitud, objetividad y capacidad para superar los sesgos y las imperfecciones vinculadas a la toma de decisiones humana, es imperativo cuestionar y examinar estas promesas, especialmente en escenarios marcados por gran complejidad.

Perspectiva crítica frente al poder del algoritmo

a. Sesgo algorítmico

El sesgo en algoritmos ha sido objeto de atención en la comunidad de la IA en los últimos años. En una investigación emblemática, Buolamwini y Gebru (2018) evidenciaron cómo la tecnología de reconocimiento facial exhibía variaciones en su precisión según subgrupos fenotípicos, revelando un rendimiento deficiente particularmente en mujeres de raza negra. El fenómeno se vinculó a las deficiencias presentes en los conjuntos de datos utilizados para entrenar estos sistemas de visión. A partir de este estudio se han identificado y analizado múltiples instancias de sesgo algorítmico, abarcando desde sistemas de recomendación que privilegian a los hombres en la selección laboral hasta herramientas de diagnóstico médico asistido por computadora que demuestran precisión variable según la etnicidad del paciente.

Las discusiones contemporáneas en torno al sesgo algorítmico lo conceptualizan como decisiones de un sistema de IA influenciadas por el uso de información que debería ser considerada irrelevante o, inversamente,

por la omisión de datos pertinentes (Turner, 2019). En el caso del algoritmo de la *Ofqual*, la indignación y rechazo suscitados estuvieron fuertemente ligados a la percepción de sesgo e injusticia en sus resultados. La dependencia de datos históricos relacionados con el desempeño previo de las escuelas condujo a situaciones en las que, por ejemplo, si ningún estudiante de una determinada escuela había alcanzado la calificación más alta en años anteriores, se tornaba casi imposible que alguien de ese mismo centro educativo lo lograra en el año en curso. Al estar diseñado para replicar patrones pasados, el algoritmo configuraba un escenario limitante para ciertos grupos, cerrando de antemano posibilidades y perpetuando desigualdades históricas.

Además, el algoritmo asignaba mayor relevancia a las CAG provenientes de grupos con menos de 15 estudiantes en una asignatura específica, dentro de un establecimiento educativo determinado. Esta característica del diseño presupone que las evaluaciones realizadas por los docentes en contextos de menor cantidad de estudiantes son más fiables, posiblemente debido a una relación más estrecha y un conocimiento más profundo sobre los alumnos. Esta premisa introdujo un sesgo importante: estudiantes de instituciones más pequeñas se vieron desproporcionadamente favorecidos por la inflación de calificaciones estimadas en comparación con aquellos de centros más grandes. Esta situación contribuyó a reforzar desigualdades preexistentes, pues las escuelas públicas suelen tener un mayor número de estudiantes por aula. Como resultado, los análisis revelaron que la proporción de calificaciones altas (A* y A) en los colegios privados incrementó con el algoritmo en 4.7 puntos porcentuales, más del doble que el aumento registrado en las escuelas públicas de enseñanza general (Porter, 2020).

El algoritmo de la *Ofqual* ilustra cómo la automatización de decisiones frecuentemente mantiene el status quo, replicando desigualdades sociales preexistentes. Como se ha evidenciado en investigaciones previas, las fallas en los algoritmos, incluyendo sesgos, tienden a afectar desproporcionadamente a aquellos grupos ya en situación de vulnerabilidad, exacerbando su riesgo de sufrir consecuencias adversas (Eubanks, 2018). Este fenómeno puede atribuirse, en parte, a que los equipos encargados del diseño algorítmico suelen ser homogéneos y carecen de la formación y la diversidad

necesaria para anticipar cómo ciertos aspectos de su diseño pueden impactar negativamente a distintos grupos sociales. Esta homogeneidad deriva en lo que se conoce como ceguera de privilegio: una incapacidad para reconocer y comprender las dificultades y desventajas que enfrentan otros colectivos debido a su posición económica, racial, de género, entre otras, en contraste con aquellos que detentan posiciones de privilegio (D'ignazio y Klein, 2020).

El algoritmo de la *Ofqual* se suma a la lista de sistemas algorítmicos que han sido señalados por presentar sesgos. La proliferación de este tipo de errores ha fomentado, dentro del campo de las ciencias computacionales, un área de investigación enfocada en desarrollar estrategias para "corregirlos". Esto se hace, principalmente, a partir de la implementación de criterios de "equidad" o "fairness", fundamentados en principios de probabilidad y estadística (Wang et al., 2022). Dichos criterios buscan asegurar que los resultados de un sistema no estén influenciados por atributos "irrelevantes" que podrían conducir a decisiones injustas, tales como el género, la etnicidad o la condición socioeconómica, a través de la definición de objetivos específicos que orientan el espacio de soluciones del modelo hacia resultados que se consideran imparciales.

Lamentablemente, los criterios de equidad algorítmica presentan limitaciones significativas. Determinar cuál de las decenas de propuestas es la más adecuada para un caso particular requiere un análisis exhaustivo y meticuloso. Además, existe un límite obvio en la capacidad de los criterios técnicos para definir y medir la justicia, especialmente cuando esto pasa por la corrección de sesgos en datos históricos y la variabilidad de normas culturales y sociales que definen lo justo en contextos específicos.

El sesgo algorítmico es un asunto de gran relevancia y su abordaje es esencial dentro de la comunidad que desarrolla sistemas de IA. Sin embargo, es igualmente crucial enriquecer el análisis de la toma de decisiones automatizada con una perspectiva más amplia, que reconozca que, a menudo, el sesgo es solo el reflejo de problemas más arraigados y complejos, vinculados con las dimensiones políticas, culturales, históricas y sociales que se entretejen en los datos, los algoritmos y la tecnología en su conjunto.

b. Complejidad

Además del sesgo evidente en el algoritmo de la *Ofqual*, se destaca un problema estructural más profundo en su diseño: la noción de que es posible prever, con base en una acumulación de datos, los resultados que personas alcanzarán en un examen aún no realizado. Tal como fue concebido, el algoritmo no solo ignoraba, sino que efectivamente cerraba las puertas a estudiantes talentosos de escuelas con un bajo historial académico, asumiendo que el esfuerzo individual era improbable, o no significativo, y podría ser subsumido en las estadísticas. Este caso señala una cuestión crucial que merece ser abordada en el debate actual: en el auge de la IA están surgiendo sistemas que pretenden predecir el comportamiento humano. Estos sistemas, más allá de sus posibles fallas de diseño, parten de una concepción imaginativa problemática, asumiendo que la complejidad humana puede ser reducida a modelos predictivos basados en datos.

Dicha perspectiva resuena con la visión cartesiana que anhela certeza, orden y previsibilidad en la comprensión del mundo, una aproximación que presupone la posibilidad de definir con claridad las conductas y decisiones humanas, obviando la ambigüedad, continuidad y fluidez que definen nuestra existencia (Birhane, 2021). Contrario a la noción de seres estáticos, somos entidades dinámicas, en constante evolución a través de nuestras interacciones y el entorno que nos rodea. Así, nos encontramos perpetuamente en formación, en un proceso de transformación que desborda las capacidades predictivas de cualquier algoritmo. En este contexto, la estrategia de la *Ofqual*, y de cualquier sistema similar en objetivo, choca con la realidad de nuestra indeterminabilidad.

Subestimar la impredictibilidad humana no solo revela una comprensión limitada sobre el alcance y las restricciones de la IA, también representa un riesgo importante. Al transformarse de herramientas que meramente describen patrones en los datos a herramientas que prescriben y moldean la realidad, estos sistemas inciden directamente en la conformación de nuestro entorno. Legitimar los resultados del algoritmo de *Ofqual* contribuye, de facto, a perpetuar un futuro anclado en desigualdades históricas, automatizando una realidad que mantiene la desventaja para los menos privilegiados.

Un diseño responsable de sistemas de IA debe reconocer que la tecnología configura no solo herramientas, sino prácticas y posibilidades futuras. Al asignar calificaciones superiores a estudiantes de escuelas privadas, se está configurando un futuro donde estas personas obtienen un acceso privilegiado a la educación superior, reduciendo simultáneamente las oportunidades para aquellos en situaciones iniciales menos favorecidas. Así, los sistemas de IA no solo determinan trayectorias; también orientan hacia determinados futuros y cierran las puertas a otros.

Clasificar y predecir en este tipo de modelos, más que un mero asunto de funciones numéricas o matemáticas, tiene un impacto directo en la construcción del orden social. Este orden, fundamentado en datos históricos, tiende a perpetuar y reforzar prácticas y normativas del pasado. En consecuencia, los modelos de IA –con sus tareas de clasificación, ranking, agrupamiento, predicción– incorporan una dimensión política y moral innegable, que debe ser reconocida y abordada con plena responsabilidad.

c. Resistencia

Cuestionar la justicia en el funcionamiento de ciertos sistemas de IA y poner en tela de juicio su autoridad representa un desafío nada trivial, marcado por al menos tres aspectos relevantes en el contexto contemporáneo.

En primer lugar, el análisis de los sistemas de IA nos confronta con diversas formas de opacidad. Según plantea Jenna Burrell (2016), esta opacidad puede manifestarse de varias maneras: como una estrategia intencionada por parte de corporaciones o entidades estatales propietarias de los sistemas; como resultado de la complejidad técnica que exige un alto grado de especialización para comprender el código y el funcionamiento del sistema; o como la dificultad asociada a la traducción de decisiones algorítmicas, optimizadas en espacios de alta dimensionalidad, a un marco de razonamiento humano que facilite su interpretación. Esta opacidad en la toma de decisiones mediada por IA obstaculiza la identificación de errores y la atribución de responsabilidades a las decisiones humanas subyacentes.

En segundo lugar, los sistemas de IA se enmarcan dentro de discursos que los exaltan por su precisión, exactitud y su fundamento en verdades respaldadas por datos y matemáticas avanzadas. Frecuentemente, se les

presenta como soluciones libres de los sesgos que ensombrecen la toma de decisiones humanas. Desafiar las conclusiones de un algoritmo implica, para empezar, cuestionar estos discursos y examinar críticamente tanto los mecanismos operativos del sistema como las dinámicas de poder que influyen en su desarrollo y funcionamiento.

Este proceso de indagación también está ligado a una comprensión detallada de los principios fundamentales del sistema, que permita identificar sus limitaciones, vulnerabilidades y riesgos. Por ejemplo, las métricas de "exactitud" comúnmente usadas para describir modelos de IA deben ser entendidas en su contexto disciplinario específico, como indicadores de la correlación entre los resultados de un algoritmo y sus datos de entrenamiento, y no como una medida de su capacidad para predecir el futuro o reflejar fielmente la realidad. Estos conceptos, más allá de su significado técnico, pueden ser empleados retóricamente para influir en la percepción pública de la tecnología. De manera similar, comprender en profundidad términos como "predicción", "aprendizaje", "equidad" y "entrenamiento" en el ámbito de la IA es clave para desmantelar marcos discursivos y formular cuestionamientos críticos que permitan una evaluación rigurosa de las herramientas.

En tercer lugar, los resultados generados por sistemas de IA pueden ser percibidos de manera fragmentada por los usuarios finales y esto complica la identificación de tendencias generales, el entendimiento del comportamiento global del sistema y el reconocimiento de patrones que puedan sugerir injusticias. En el contexto del algoritmo de la *Ofqual*, la interacción y comunicación entre estudiantes, especialmente aquellos que ya habían establecido conexiones previas a la pandemia, fue crucial para detectar un sesgo vinculado directamente al origen escolar.

Además, la controversia en torno al uso del algoritmo se convirtió en un tema central en el debate público del Reino Unido. En este proceso, fue fundamental la manera en que se configuró el imaginario colectivo respecto al algoritmo, conforme la ciudadanía intentaba comprender su funcionamiento y compartía sus hallazgos. La visibilización de los efectos del modelo empleado por la *Ofqual* emergió de un esfuerzo colectivo; numerosas personas se involucraron en la problemática y cuestionaron la autoridad del algoritmo, no necesariamente por haber sido afectadas

directamente por los exámenes, sino en solidaridad a las protestas y en respuesta a la atención mediática que estas generaron. Así, el diálogo en redes sociales –que en Twitter popularizó el uso del hashtag *#fuckTheAlgorithm*– y otros espacios digitales catalizó la organización de manifestaciones callejeras, fomentando la emergencia de un sentimiento colectivo de resistencia frente a la autoridad del algoritmo y de rechazo a una herramienta percibida como injusta. Las movilizaciones sociales, en última instancia, llevaron al gobierno a ofrecer a los estudiantes la opción de retener sus CAG, aceptar la nota modificada por el algoritmo, o presentar los exámenes en una fecha posterior, demostrando el poder de la acción colectiva en la redefinición de políticas y prácticas tecnológicas.

Este caso ilustra posibles respuestas a los desafíos planteados al comienzo de esta sección

Primero, ante la opacidad de los sistemas de IA, se hace imperativo abogar por la transparencia y la explicabilidad, especialmente en aquellos sistemas que tienen el potencial de impactar significativamente en la vida de las personas. Para esto es fundamental reconocer que no todos los sistemas de IA tienen el mismo propósito o naturaleza. Siguiendo la clasificación de Narayanan (2019), los sistemas pueden diferenciarse en funciones de percepción (como el reconocimiento de imágenes), automatización de criterios (como la clasificación de correos electrónicos) y toma de decisiones que afectan directamente a las personas. La capacidad de explicar las decisiones del sistema, junto con la transparencia en sus mecanismos, datos y construcción, y las estrategias para su auditoría, deben estar alineadas con su propósito y alcance.

En el caso del algoritmo de la *Ofqual*, cuando se conocieron las noticias sobre las calificaciones asignadas no existía un procedimiento de apelación claro. De hecho, el procedimiento era muy complicado y los estudiantes tenían que pagar para apelar sus calificaciones. Al igual que el propio modelo, las deficiencias del procedimiento de apelación podían afectar de forma desproporcionada a los estudiantes de entornos socioeconómicos más bajos.

En situaciones donde los algoritmos tienen la capacidad de influir en vidas humanas, es inadmisible que las decisiones se tomen sin un análisis

riguroso, que incorpore múltiples perspectivas y una descripción detallada de todas las partes del proceso y que facilite así una auditoría por especialistas y un escrutinio abierto. Conscientes de que ningún algoritmo es infalible, la *Ofqual* debería haber anticipado la posibilidad de resultados injustos para algunos estudiantes e implementado un mecanismo adecuado para realizar aclaraciones y apelaciones. La gestión de esta situación evidencia que no solo hubo problemas metodológicos, como el que originó sesgo algorítmico, sino también (especialmente) epistémicos. Depositar total confianza en un algoritmo demuestra un malentendido sobre las capacidades de estos sistemas en contextos complejos y variables.

En segundo lugar, es clave desmantelar el mito de la perfección que a menudo rodea a la IA, promovido en gran medida por estrategias de mercadotecnia. Se necesita reconocer que estos modelos no erradican la incertidumbre en la toma de decisiones, solo la transforman; las nuevas formas de incertidumbre pueden estar ligadas a distintos factores del proceso y demandan un análisis cuidadoso. A medida que aumenta la complejidad de la situación que se intenta modelar, crece la certeza de que cualquier modelo propuesto tendrá errores, pues capturar la plenitud de la complejidad real es una tarea que escapa a las posibilidades de cualquier modelo. Esto no resta valor a los modelos, pero sí subraya que sus aplicaciones estarán siempre condicionadas por las suposiciones y simplificaciones adoptadas durante su desarrollo, restringiendo su utilidad a escenarios específicos. La sociedad tiene el derecho a estar informada sobre los posibles errores de un algoritmo y sobre la forma en que estos podrían manifestarse. Más aún, la sociedad debe retener el derecho a decidir cuándo la incertidumbre de las decisiones humanas es preferible a una injusticia automatizada envuelta en un aura de infalibilidad.

En este sentido, es interesante observar que el algoritmo de la *Ofqual* fue ideado con el objetivo de mitigar el sesgo asociado a las evaluaciones subjetivas de los docentes, un sesgo que, de hecho, era real y tuvo consecuencias. Al anularse la implementación del algoritmo y retornar a las CAG sí se produjo un incremento en las notas, lo que sí generó desafíos logísticos para las universidades al tener que acomodar a un número mayor de estudiantes admitidos de acuerdo con altas calificaciones en los A-levels. Un ejemplo de esto fue la Universidad de Durham, que se vio

forzada a ofrecer incentivos para que los estudiantes pospusieran su ingreso hasta el 2021.

El problema que el algoritmo buscaba solucionar era genuino, pero su aplicación no logró crear un entorno libre de sesgos. En realidad, el algoritmo introdujo un nuevo tipo de sesgo, que en muchos casos afectó especialmente a estudiantes de estratos socioeconómicos más bajos. El caso ilustra la complejidad asociada a la búsqueda de soluciones algorítmicas para problemas sociales: las simplificaciones inevitables en el modelo y el diseño del algoritmo no erradicaron el sesgo: lo transformaron, creando desigualdades alternativas. Más aún, lo hicieron partiendo de una fantasía, pues los algoritmos no pueden, ni este ni otros escenarios, predecir de manera certera el futuro de las personas. La idea de "predicción" en el aprendizaje automático hace referencia a cálculos matemáticos y difiere significativamente de su interpretación en el imaginario popular. Esta discrepancia puede llevar a malentendidos sobre las capacidades reales de la tecnología algorítmica.

Finalmente, es fundamental enfatizar la importancia de ejercer nuestros derechos a la explicabilidad, a la rendición de cuentas por parte de quienes diseñan y operan los sistemas de IA, y a tomar decisiones informadas sobre la pertinencia de implementar estas tecnologías en contextos específicos. Esto requiere fomentar estructuras de organización colectiva que nos permitan reflexionar sobre la influencia de los sistemas de IA en nuestra cotidianidad y desarrollar estrategias de acción para influir en el proceso de adopción, adaptación o rechazo de estas herramientas tecnológicas. Solo así podremos asegurarnos de que la evolución y aplicación de la IA se alinee con los valores y objetivos compartidos por nuestra sociedad, y no solamente con los de un grupo reducido que se encuentra en el poder.

Conclusiones

Las expectativas en torno a la IA desempeñan un papel central en la configuración de los sistemas y en sus aplicaciones prácticas, actuando como intermediarias entre los diversos niveles de implementación de esta tecnología y las comunidades. El exceso de entusiasmo por las capacidades de la IA ha generado la difusión de promesas que superan ampliamente

las capacidades reales de la tecnología. En ocasiones, estas expectativas desmedidas conducen a problemas epistémicos, como en el caso de algoritmos propuestos para identificar delincuentes basados en rasgos faciales (Wu et al., 2016), o como se ha discutido en este capítulo, para predecir calificaciones justas en exámenes que estudiantes no han presentado.

Frente a todas las mitologías sobre las capacidades de los sistemas de IA, es fundamental volver al cuestionamiento crítico de estas herramientas y de su aplicación en contextos específicos. Los sistemas de IA, especialmente aquellos basados en aprendizaje automático, poseen una capacidad extraordinaria para identificar patrones en los datos. Al aproximarnos de manera adecuada a estos patrones, podemos abrir la puerta a la reflexión sobre dinámicas históricas, examinar con mayor detenimiento las relaciones entre variables o plantear nuevas hipótesis que nos ayuden a comprender mejor el mundo que nos rodea. Sin embargo, fracasar en reconocer que simplemente estamos identificando patrones puede llevarnos a terrenos confusos, como la creencia errónea de que es posible predecir el comportamiento humano.

El caso del algoritmo de la *Ofqual* subraya la necesidad de lineamientos claros para la supervisión de sistemas de IA en todas sus fases: diseño, entrenamiento y análisis de resultados. Esto implica promover la transparencia, la explicabilidad y la responsabilidad, mitigando riesgos, especialmente para los sectores más vulnerables.

Además, es esencial que busquemos influir en la percepción de la sociedad sobre el funcionamiento de los sistemas de IA. Al lograr un mejor entendimiento de estas herramientas, podremos trascender los marcos discursivos predominantes y cuestionar adecuadamente sus usos e implicaciones en diversos escenarios. De esta manera, podremos desafiar la autoridad de estos sistemas, rechazar su uso cuando sea necesario, y reimaginar sus posibilidades, alejándonos del utopismo tecnológico en favor de aplicaciones realistas, equitativas y beneficiosas para la colectividad.

Referencias

Ames, M. G. (2019). *The charisma machine: The life, death, and legacy of one laptop per child*. Mit Press.

A-levels and GCSEs: *Free exam appeals for schools in England*. (2020, 15 de agosto). BBC News. https://www.bbc.com/news/uk-53787938

Birhane, A. (2021). The impossibility of automating ambiguity. *Artificial Life,* 27(1), 44-61.

Buolamwini, J. y Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. En *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society*, 3(1), 2053951715622512.

D'ignazio, C. y Klein, L. F. (2020). *Data feminism*. MIT press.

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

Narayanan, A. (2019). *How to recognize AI snake oil*. Arthur Miller Lecture on Science and Ethics.

Porter, J. (2020, 17 de Agosto). *UK A-level results algorithm biased amid coronavirus pandemic*. The Verge. https://www.theverge.com/2020/8/17/21372045/uk-a-level-results-algorithm-biased-coronavirus-covid-19-pandemic-university-applications

Turner, J. y Turner, J. (2019). *Controlling the Creators. Robot Rules: Regulating Artificial Intelligence* (pp. 263-318).

Wang, X., Zhang, Y. y Zhu, R. (2022). *A brief review on algorithmic fairness. Management System Engineering*, 1(1), 7.

Wu, X. y Zhang, X. (2016). *Automated inference on criminality using face images*. arXiv preprint arXiv:1611.04135, 4038-4052.