


Arte y diseño crítico como desencadenantes de reflexión social para una inteligencia artificial más legible, inclusiva y responsable¹

Lara Sánchez Coterón
Universidad Complutense de Madrid (España) ✉ 

<https://dx.doi.org/10.5209/TEKN.94934>

Recibido: 07 de marzo de 2024 • Aceptado: 22 de abril de 2024 • OPR

ESP Resumen. El artículo reflexiona sobre las implicaciones éticas, políticas y las relaciones de poder que produce la integración de los sistemas computacionales de inteligencia artificial en nuestra sociedad. Analiza cómo desde las prácticas artísticas y de diseño crítico se acometen estas cuestiones, identificando ciertos puntos de intervención cruciales sobre las diferentes fases de los procesos integrales de la Inteligencia Artificial (IA). Proponemos que este tipo de proyectos, que visibilizan dinámicas habitualmente opacas o enigmáticas, pueden activar a nivel social otras formas de concebir y comprender estas tecnologías y generar nuevos planteamientos sobre cómo y para qué utilizarlas.

Palabras clave: creación artística; diseño de sistemas; justicia social; política cultural.

ENG Art and critical design as triggers for social reflection on a more legible, inclusive and responsible artificial intelligence

ENG Abstract. The article reflects on the ethical and political implications and the power relations produced by the integration of artificial intelligence computational systems in our society. It analyses how artistic and critical design practices address these questions, identifying certain leverage points to intervene in the different phases of the integral processes of Artificial Intelligence (AI). On this basis, we propose that this type of projects, which make usually opaque or enigmatic dynamics visible, can activate at a social level other ways of conceiving and understanding these technologies and generate new approaches on how and what to use them for.

Keywords: artistic creation; cultural policy; social justice; systems design.

Sumario. 1. Introducción. 2. Una IA legible, inclusiva y responsable. 3. Creación, algoritmos y datos, datos, datos. 4. Prácticas artísticas críticas sobre diferentes fases de los procesos integrales de la IA. 5. Conclusiones y Perspectivas. 6. Declaración sobre el uso de LLM. 7. Referencias.

Como citar: Sánchez Coterón, Lara (2024) Arte y diseño crítico como desencadenantes de reflexión social para una inteligencia artificial más legible, inclusiva y responsable. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 22(1), 41-49. <https://dx.doi.org/10.5209/tekn.94934>

1. Introducción

Al cierre del seminario Deepfakes: ficción, política y algoritmos, realizado en el contexto los Encuentros de Diseño y Cultura Digital #EDCD_03, en 2019 en

Medialab Prado Madrid, quedaron abiertas una serie de ideas de trabajo y reflexión en torno a la posibilidad de fortalecer desde las prácticas artísticas y desde el diseño crítico, acercamientos para

¹ #EDCD con la colaboración del proyecto de investigación *Interacciones del arte en la tecnosfera. Creación de soportes e instrumentos para una reflexión crítica, pública y transversal* (HAR2017-86608-P), proyecto del grupo de investigación UCM 970588-Prácticas artísticas y formas de conocimiento contemporáneas, financiado por el Ministerio de Ciencia, Innovación y Universidades.

una inteligencia artificial más legible, inclusiva y responsable, en definitiva, más justa. En el seminario también se abordaron cuestiones teóricas como las diferentes posibilidades de establecer lo que es verdad en la coyuntura actual de uso de las inteligencias artificiales (en adelante IAs), los acercamientos contemporáneos que desde los proyectos artísticos se están realizando sobre estas tecnologías y acciones como el taller introductorio a usos prácticos de inteligencia artificial en el ámbito de la creación de imágenes (García Angulo et al., 2019). Estos encuentros multidisciplinares trataron de generar procesos de conocimiento y sensibilización en relación con los modelos de aprendizaje automático o *Machine Learning* (ML) para poder crear diálogo en torno a determinadas situaciones de dominio, autoridad y preponderancia que subyacen a estas tecnologías.

Ya en 1995 Stephen Wilson trazaba el panorama creativo del arte creado con inteligencia artificial en su artículo titulado *Artificial Intelligence Research as Art* (Wilson, 2005). El autor reconocía que la relación de los artistas con la IA iba más allá de los límites técnicos y situaba estos trabajos en la esfera de la investigación artística que se interpela sobre la naturaleza del ser humano, la propia naturaleza de la inteligencia misma, los límites de las máquinas y nuestros límites como creadores de artefactos. Tres décadas más tarde, en el contexto de una sociedad altamente datificada y con procesos de IA que atraviesan las vidas de gran parte de la población, los artistas continúan estando pendientes y siendo el motor de esos aspectos sobre la naturaleza y los límites de lo humano y lo artificial, y las implicaciones éticas, políticas y las relaciones de poder que produce la integración de estas tecnologías en nuestra cotidianidad.

El presente artículo revisa las prácticas artísticas y de diseño crítico que, más allá de la generación de imaginarios IA como resultado del placer de la sorpresa, la predicción y la inmediatez, buscan hacer visibles las decisiones de diseño en los procesos de conceptualización e implementación de las IAs en relación con diferentes fases y las lagunas terminológicas y conceptuales en los conjuntos de datos de entrenamiento, además del papel clave de estos conjuntos de datos en relación al funcionamiento de los algoritmos. La investigación busca ratificar la importancia de las conversaciones transdisciplinares en torno al arte, la ética y las políticas de las tecnologías mediadas por la IA, así como el modo en que estos diálogos pueden contribuir a los procesos de diseño de las diferentes fases que constituyen esta materia. Se trata de un trabajo de carácter exploratorio, que se centra en valorar las posibilidades que el arte y el diseño crítico tienen de explicitar los vínculos entre los sistemas de IA y cuestiones estructurales de poder al discutir las siguientes cuestiones: ¿Pueden el arte y el diseño crítico enriquecer la reflexión social sobre las diferentes fases de los procesos integrales de la IA? ¿Son capaces las prácticas artísticas que trabajan con o sobre procesos de inteligencia artificial funcionar como agentes de cambio y acercarnos a modos más conscientes y consecuentes sobre la creación y los usos de estas tecnologías? Al abordar estas preguntas,

exploramos diferentes enfoques interdisciplinares conectados y proponemos:

- Reflexionar, desde las humanidades, sobre la ideología solucionista como un enfoque que presenta la IA como un dispositivo que se ejerce de forma irreflexiva y aproblemática en algunos aspectos.
- Valorar, como se propone desde estudios en psicología, los sesgos que podemos identificar en las IAs y los datos que las alimentan y cómo estos retornan y condicionan las interpretaciones que los usuarios hacen sobre la información.
- Introducir en el debate posturas que desde el arte, el diseño crítico y los movimientos colectivos y sociales proponen principios para un acercamiento crítico a las tecnologías que resultan muy interesantes de cara a crear unas tecnologías descentralizadas, más equitativas, legibles, inclusivas y responsables.

Planteamos estos acercamientos teniendo en cuenta que, tal y como se describe desde las ciencias de la computación, de partida los modelos de inferencia cognitiva de las IAs no tienen capacidad de generar conjeturas ni hipótesis (Larson, 2021). En ese sentido estamos hablando de automatización de procesos y de medios predictivos no comparables a la inteligencia humana, sino concebidos desde ella.

La investigación aborda los sistemas computacionales de inteligencia artificial como artefactos culturales que ejercen un poder importante y que modelan continuamente nuestra cultura colectiva. En estos modelos computacionales las dinámicas de poder son opacas e interrelacionadas, y en los que esa suerte de cognición distribuida a menudo oculta a las personas las consecuencias de sus acciones. La cuestión que aquí se aborda es ¿Pueden las prácticas artísticas con y sobre la IA poner en evidencia estas dinámicas sirviendo como espejos que revelan la trayectoria actual de nuestra sociedad? ¿Puede esa visibilización desde las artes y el diseño crítico hacer que seamos más conscientes de la necesidad de repensar algunos de estos planteamientos en vez de darlos por hecho?

2. Una IA legible, inclusiva y responsable

Para empezar a entender los procesos de inteligencia artificial es necesario conocer su anatomía integral de desarrollo. Esta puede dividirse de manera general en tres etapas, que incluyen diseño, desarrollo e implementación y una serie de sub-etapas que componen esas tres fases y que abarcan desde la concepción, hasta la producción de cualquier iniciativa de IA (De Silva y Alahakoon, 2022).

Primera fase, de diseño, realizada desde el ámbito de la ciencia de los datos que incluye: Definición del problema. Identificación y comprensión de la situación que se quiere abordar con la IA y el establecimiento de objetivos claros y medibles. Recopilación y preparación de datos. Recolección de datos relevantes para el problema en cuestión. Limpieza y preprocesado de los datos con el fin de eliminar ruidos y asegurar la calidad de los resultados. Exploración y análisis de datos para comprender su distribución y características. Identificación de patrones

y tendencias que puedan ser relevantes para el problema.

Segunda fase, de desarrollo, desde el contexto de las ciencias de aprendizaje automático (ML) que comprende: Selección de algoritmos. Elección del algoritmo de aprendizaje automático o las técnicas de IA más adecuadas para abordar el problema. Consideración de factores como la naturaleza de los datos, el tamaño del conjunto de datos (*datasets*) y los objetivos del proyecto. Entrenamiento del modelo. Utilización de datos de entrenamiento para ajustar los parámetros del modelo y mejorar su rendimiento. Evaluación del modelo utilizando conjuntos de datos de validación para evitar el sobreajuste. Evaluación del rendimiento del modelo utilizando datos de prueba independientes. Ajustar y afinar el modelo según sea necesario.

Tercera fase, de implementación, desde el área de la ingeniería de ML que abarca: Implementación e integración del modelo entrenado en un entorno de producción o en la aplicación específica para la cual fue diseñado. Monitoreo y mantenimiento: Registro y evaluación del rendimiento del modelo en tiempo real. Realización de actualizaciones y ajustes según sea necesario para mantener la eficacia del modelo a lo largo del tiempo.

Es curiosa y sintomática la retórica de resolución de problemas que exuda la naturaleza de estas descripciones y que entronca con la ideología solucionista vinculada a las nuevas tecnologías. Esta corriente de pensamiento, surgida desde el urbanismo y desarrollada en el contexto de las grandes empresas tecnológicas de Silicon Valley, entiende el desarrollo de objetos, dispositivos, datos y operaciones inteligentes que resuelven problemas de definición clara, concisa y solución definitiva. En ese contexto solucionista la inteligencia artificial es ejercida de forma aproblemática y acrítica. Preconiza además la inquestionabilidad de la solución técnica y nos sitúa en un territorio que la filósofa Marina Garcés (2017) denomina de 'las inteligencias delegadas'. Pero, ¿podemos subrogar la toma de decisiones de problemas sociales complejos a las máquinas? Pues de hecho lo hacemos. En su artículo 'AI systems as state actors' Kate Crawford y Jason Schultz (2017) detallan las preocupantes lagunas que genera el uso de las IAs en la toma de decisiones desde las administraciones:

Los gobiernos recurren cada vez con más frecuencia a proveedores externos para que les proporcionen las arquitecturas algorítmicas de los servicios públicos, como las prestaciones sociales y las evaluaciones de riesgos penales. Así, cuando se les ha cuestionado, muchos gobiernos estatales han negado cualquier conocimiento o capacidad para comprender, explicar o remediar los problemas creados por los sistemas de IA que han adquirido a terceros (Crawford y Schultz, 2019, p. 1941).

El problema, en este tipo de situaciones, es que la IA es técnicamente incapaz de percibir el contexto. Antoinette Rouvroy (2020), investigadora académica de ciencias legales en la University of Namur (Bélgica), defiende en su texto 'Algorithmic governmentality' escuchar a las personas que desarrollan

las IAs y no la narrativa de la industria, que ha sido asumida por la clase política, y que la utilizan para abdicar de su responsabilidad. En demasiadas ocasiones, al frente de la reproducción de este modelo ideológico tecnoliberal, hay actores como empresas, estados o instituciones que intervienen como gestores en la optimización tecnológica de las comunicaciones y la obtención y procesamiento de datos. Estas entidades se erigen como 'autoridades interpretativas' con una posición auto-validada para decidir cuáles, de los millones de datos que circulan en las redes sociales, se reconocen por ejemplo como peligrosos para el orden social (Gindin, Cingolani y Rodríguez-Amat, 2021).

Más allá de los sesgos que estas 'autoridades interpretativas' pueden integrar en los modelos de IA (instituciones, gobiernos, corporaciones), la propia IA parece condicionar la toma de decisiones que podemos llegar a hacer las personas. Esta suerte de bucle pernicioso está siendo investigado desde la psicología experimental por un grupo de científicas de la Universidad de Deusto. Un estudio realizado por Lucia Vicente Holgado y Helena Matute (2023) sugiere que las personas podemos heredar errores sistemáticos de las IAs y replicar a partir de las recomendaciones que obtenemos de estos modelos decisiones erróneas y sesgadas. En ese sentido, desde el ámbito de las artes y la cultura surgen espacios con posibilidad de intervención para incidir y desconfigurar estos sesgos en bucle a la par que contrarrestar la tremenda influencia que las grandes corporaciones, en las que delegamos nuestros datos, tienen sobre la orientación, los intereses y los objetivos últimos de la investigación y el desarrollo de la IA (Alsina, 2022). Sin embargo, estas corrientes de debate que generan atención en torno a nuestras relaciones con la IA, se están llevando a cabo más desde lo académico que desde lo social y más desde lo corporativo que desde la política en los escenarios gubernamentales (Gindin, Cingolani y Rodríguez-Amat, 2021).

La ética y la transparencia han de ser consideraciones importantes en todas las etapas del proceso integral de la IA. En esta investigación, al hilo de indagaciones anteriores en las que hemos utilizado conceptos y terminología de la científica experta en análisis de sistemas Donatella H. Meadows (2009; Sánchez Coterón, 2020), buscaremos reconocer las intervenciones sobre determinados puntos críticos a lo largo de los procesos integrales de creación de la IA como son la recopilación y preparación de datos y los entrenamientos del modelo. En todos los casos, independientemente de la naturaleza del contexto, es importante poder explicar cómo funciona la toma de decisiones del modelo, pero en la mayoría de las ocasiones estas explicaciones no suelen ser transparentes ni públicas. A ese respecto resulta interesante tener en cuenta los principios que definen los valores de movimientos sociales, con un acercamiento crítico a las tecnologías, como las propuestas para una web descentralizada (Ishikawa Sutton y Ryan, 2024) o el decálogo de prácticas del Critical Engineering Working Group (Oliver, Savičić y Vasiliev, 2021). Desde estas iniciativas se sostiene que las tecnologías deben permitir poner los intereses de las personas por delante de otro tipo de conveniencias, defendiendo por encima de todo la seguridad, privacidad y autodeterminación de las personas,

poniendo el foco no solo en el atractivo de la implementación sino en los modos de influencia y sus efectos específicos. Instan a la coexistencia y la interoperabilidad de diferentes tecnologías, desaconsejando 'jardines amurallados' y valorando el código fuente abierto como elemento fundamental de unas tecnologías abiertas e integradoras que fomenten las relaciones entre iguales, en lugar del control jerárquico y desequilibrios de poder. Proponen minimizar la vigilancia y la manipulación del comportamiento de las personas, optimizar los beneficios sociales y capacitar a los individuos para determinar cómo y por qué se utilizan sus datos. La idea es que incluso cuando las personas que diseñan y desarrollan IAs para que sean versiones mejoradas del ser humano, inteligencias más racionales, más eficientes e incansables, existe la posibilidad de concebir modos alternativos de desarrollo tecnológico que rechacen las lógicas raciales, coloniales, heteropatriarcales que mantienen y perpetúan las jerarquías sociales y la desigualdad (Vora y Atanasoski, 2019).

Un ejemplo claro y recurrente es el sesgo operativo que presenta el reconocimiento facial. Durante años, los expertos han dejado claro que esta tecnología tiene dificultades para reconocer con precisión los rostros negros, los géneros no masculinos y las diferentes razas. El reconocimiento facial parece funcionar mejor con rostros blancos y masculinos cisgénero. Joy Buolamwini, poetisa de código ghanesa-estadounidense, doctora en informática en el MIT especializada en sesgos algorítmicos y presidenta de la organización The Algorithmic Justice League, se da cuenta mientras está realizando su tesis de máster, en 2016, que los sistemas de reconocimiento facial no detectan su rostro y comienza a buscar de qué otros modos afecta la tecnología artificial a las minorías (Buolamwini, 2016). Los algoritmos y las tecnologías de inteligencia artificial discriminan por raza y género en ámbitos cotidianos como el acceso a la vivienda, las oportunidades profesionales, la atención sanitaria, el acceso al crédito, la educación e incluso cuestiones legales. El documental *Coded bias* (Kantayya, 2020) recoge todos estos aspectos sobre los sesgos de los sistemas de visión computacional.

La neutralidad de estos artefactos culturales, los sistemas computacionales de inteligencia artificial,

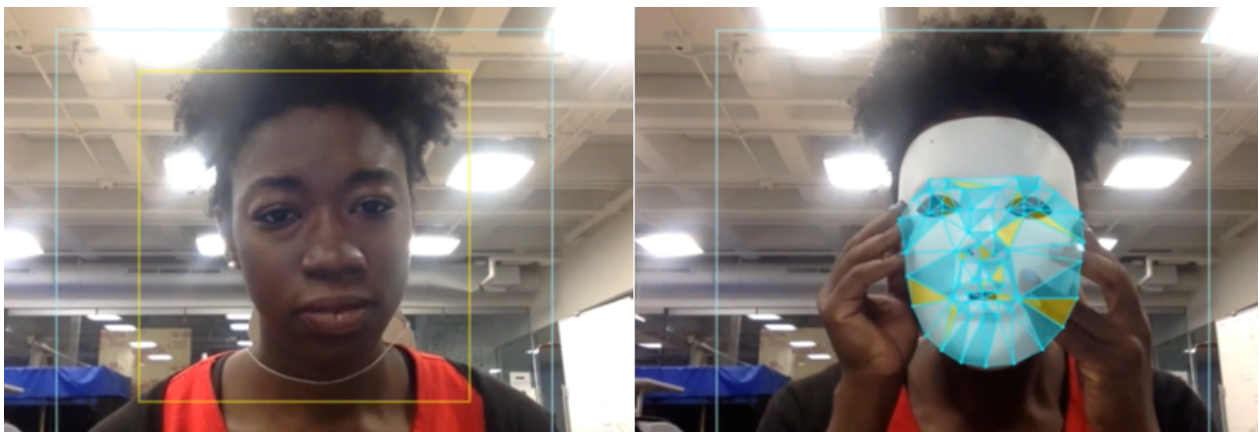
queda en entredicho a través de debates como este. Las prácticas que desde contextos culturales y artísticos surgen al respecto pueden ayudar a localizar y traer a primer plano situaciones de injusticia y desequilibrios con el fin de que futuras conceptualizaciones, implementaciones e interpretaciones alrededor de los modelos de IA sean compartidas, aceptables y contribuyan a la creación de significado para grupos con diferente género, raza o nivel económico, evitando hostilidades contra determinados grupos sociales y/o minorías.

3. Creación, algoritmos y datos, datos, datos

La actividad artística y las personas que la desarrollan tienen tendencia a adoptar enfoques atrevidos en relación con parámetros y situaciones que nos afectan como sociedad. De hecho, el arte siempre ha existido en una relación compleja, simbiótica y en continua evolución con las capacidades tecnológicas de la cultura y la sociedad que le ha tocado vivir. Los algoritmos y el análisis predictivo permiten anticiparnos y poner remedio a posibles desgracias pero también son una potente herramienta de control social que puede minar nuestra voluntad de transformar y mejorar el mundo (Manovich, 2023). En ese sentido los enfoques intrépidos que desde el arte se hacen de las tecnologías de aprendizaje automático, de las IAs, resultan acercamientos audaces e interesantes para las posibles transformaciones de cara a un futuro mejor y más justo para las personas y para el planeta. Artistas de todo el mundo, a través de la colaboración y la confrontación con la tecnología y la sociedad, están trabajando en estos temas y abriendo paso a posibilidades de resignificación vinculadas a estos medios.

En los últimos tiempos los algoritmos de ML se han utilizado entre otras muchas cosas para insertar digitalmente la identidad de diferentes personas y personajes públicos en escenas de video en las que no aparecían inicialmente, en ámbitos tan dispares como el porno, el arte, la publicidad o la política. A esta técnica se le denomina *deepfake*. Este tipo de algoritmos entrenados con grandes conjuntos de datos de imágenes reales, usan redes neuronales (redes generativas de confrontación, del inglés *Generative Adversarial Networks* (GAN), para fabricar

Imagen 1. How I'm fighting bias in algorithms, 2016.
Fuente: Fotogramas del video de la charla TED de la investigadora (Buolamwini, 2016)



nuevas imágenes. La tecnología ha alcanzado un nivel de fotorrealismo sin precedentes de tal modo que distinguir un video real de uno sintético es ya en muchas ocasiones imposible para el ojo humano.

Esta indeterminación es el punto de partida del corto publicitario *La Compil des Bleues* (Guerassimov, 2023) de la agencia francesa Marcel. Mediante el uso de *deepfake* se presentan una serie de jugadas épicas de la selección francesa de fútbol masculina para a continuación revelar que la base de esas tomas de video son partidos de la selección femenina a las que se ha añadido, mediante técnica de ML, la cara de jugadores famosos masculinos. Se utiliza aquí la 'veracidad' de la imagen procesada para evidenciar determinados sesgos de género en nuestra sociedad.

El gran volumen de informaciones digitales que generamos ha abierto la puerta a la reflexión sobre los procesos de interpretación sobre esta datificación. Los modos en los que se producen esos datos, la manera en la que se ponen en circulación y los procesos de reconocimiento de los datos, está ayudando a desnaturalizar la idea de que el dato es neutral y compacto. La ciencia de datos es sin duda alguna una forma de poder. Se utiliza para denunciar injusticias, para mejorar los resultados sanitarios o para derrocar gobiernos. Pero también para discriminar, vigilar y controlar. Este doble espacio de posibilidades hace evidente la necesidad de tener que cuestionarnos qué perfiles y que personas están a los mandos y que intereses mayoritarios se están persiguiendo. En términos generales, y así lo recogen las autoras Catherine D'Ignazio y Lauren F. Klein (2020), las narrativas en torno a los macrodatos y la ciencia de datos son abrumadoramente blancas, masculinas y tecno-heroicas.

La creación de sistemas de IA requiere de estos datos. Los sistemas supervisados de aprendizaje automático diseñados para el reconocimiento facial o de objetos se entrenan con grandes cantidades de datos contenidos en *datasets* formados por muchas imágenes distintas. Para crear un sistema de visión por ordenador capaz de reconocer la

diferencia entre imágenes de dos objetos distintos, las personas que lo desarrollan tienen que recopilar, etiquetar y entrenar una red neuronal con miles de imágenes etiquetadas con esos dos objetos. Los algoritmos realizan un estudio estadístico de las imágenes y desarrollan un modelo para reconocer la diferencia entre las dos clases de objetos. Si todo va según lo previsto, el modelo entrenado será capaz de distinguir la diferencia entre imágenes de ambas categorías que nunca ha encontrado antes (Crawford y Paglen, 2019). Hasta aquí todo neutro e imparcial.

Pero detengámonos en la epistemología de los conjuntos de datos de entrenamiento. Los *datasets* no son simples materias primas para alimentar algoritmos, sino intervenciones políticas. Como tal, gran parte del debate sobre la parcialidad de los sistemas de IA no da en el clavo: no existe un punto de vista neutral o apolítico sobre el que puedan construirse los datos de entrenamiento. La propiedad de los datos, su extracción, su resguardo, los mecanismos de su almacenamiento y todos los procesos intermedios humanos o algoritmos que califican y combinan los datos por su valor y los sentidos que los atraviesan están condicionados por significados y tienen implicaciones sociales y políticas (Gindin, Cingolani y Rodríguez-Amat, 2021). En demasiadas ocasiones está siendo desatendida la conexión entre los procedimientos tecnológicos para producir datos y los discursos que los interpretan, cayendo estas relaciones en contextos de indeterminación que habitualmente son asumidas de manera acrítica.

4. Prácticas artísticas críticas sobre diferentes fases de los procesos integrales de la IA

Las áreas clave de enfoque dentro de esta investigación incluyen los riesgos y las potencialidades de la inteligencia artificial en sectores sociales específicos, con un interés particular en las arquitecturas de la información, las estructuras de poder y los

Imagen 2. *La compil des Bleues*, 2023.
Fuente: Fotograma del video comercial (Guerassimov, 2023)



modelos políticos. Nos interesa reconocer diferentes acercamientos artísticos al uso y el discurso sobre las IAs, vinculados a las fases integrales de desarrollo de la IA y en concreto a lo que hemos señalado como puntos críticos de intervención.

4.1. Recopilación y preparación de datos

Vinculados a la primera fase de desarrollo de las IA, su fase de diseño, surgen proyectos que desde las prácticas artísticas ponen en tela de juicio los modos en los que se hace la recolección, limpieza y pre-procesado de los datos. *Cleaning Emotional Data* (Giardina Papa, 2020) es una videoinstalación de tres canales que aborda las nuevas formas de trabajo precario que surgen en las economías de la inteligencia artificial. En concreto, se centra en la infraestructura global de micro-trabajadores que 'limpian' datos para entrenar algoritmos en reconocimiento de emociones. En el invierno de 2019, la artista Elisa Giardina Papa trabajó a distancia para varias empresas norteamericanas en procesos human-in-the-loop HITL que proporcionan conjuntos de datos limpios para entrenar algoritmos de IA en la detección de emociones. Entre las tareas que realizó estaba la taxonomización de emociones, la anotación de expresiones faciales y la grabación de su propia imagen para animar emociones de personajes tridimensionales. Esta demanda de legibilidad emocional es utilizada cada vez más para identificar el estado de ánimo de las personas consumidoras o para detectar a personas potencialmente peligrosas que suponen una amenaza para el Estado. La pieza reflexiona sobre cómo las sensibilidades emocionales superan las categorizaciones reductoras con las que están siendo alimentados los algoritmos.

Al hilo de este tipo de nuevas tareas que se realizan dentro de infraestructuras globales con microtrabajadores precarios que limpian datos es interesante revisar la aplicación desarrollada por la profesora de la UC San Diego Lilly Irani. *Turkopticon* (Irani y Silberman, 2019) permite que los trabajadores de Amazon Mechanical Turk puedan valorar tanto el tipo de tarea a realizar, como a las entidades que se lo encargan. Esta mejora pone fin a un problema real al que se enfrentan las personas trabajadoras en estos entornos, que a menudo no saben cómo es la entidad que encarga la tarea y no tienen manera de compartir información con otras personas que trabajan como Turkers.

Con la preocupación de visibilizar las condiciones a partir de las cuales los conjuntos de datos son provistos de sentido, como constructos discursivos situados social y culturalmente, el proyecto *Excavating AI: The Politics of Training Sets for Machine Learning* (Crawford y Paglen, 2019) revisa durante más de dos años las lógicas subyacentes de cómo se utilizan las imágenes para entrenar a los sistemas de IA a 'ver' el mundo. En ese periodo temporal examinan cientos de conjuntos de datos (colecciones de imágenes) utilizados en inteligencia artificial, desde los primeros experimentos con reconocimiento facial a principios de los años sesenta, hasta conjuntos de entrenamiento contemporáneos que contienen millones de imágenes, en una suerte de arqueología de los conjuntos de datos. El trabajo, en forma de ensayo, explora cómo y por qué la interpretación automatizada de imágenes es un proyecto inherentemente social y político, más que puramente técnico. Crawford y Paglen (2019) pretenden hacernos conscientes de que entender la política de los sistemas de IA es fundamental dado que estos modelos han incorporado a la arquitectura de las instituciones sociales y toman decisiones del tipo

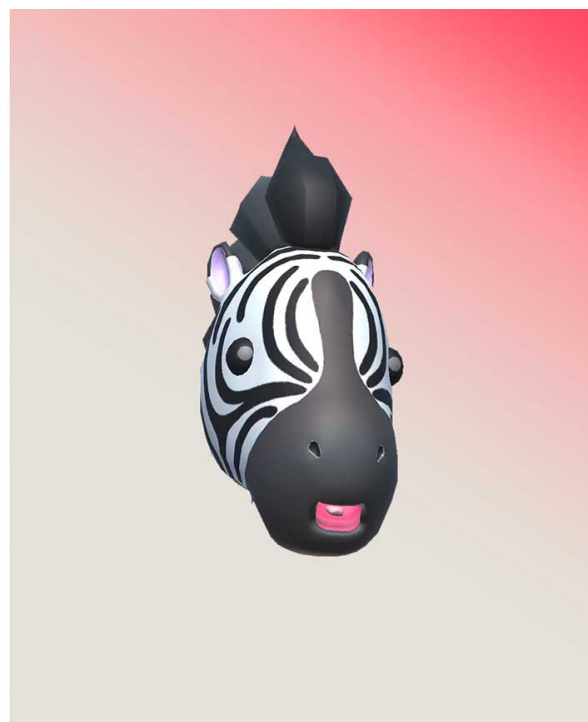
Imagen 3. *Cleaning Emotional Data*, 2020.
Fuente: Website de la artista, Elisa Giardina Papa (Giardina Papa, 2020)

TASK 2: ANIMATE THESE AVATARS WITH YOUR OWN EMOTIONS

Please animate the zebra by expressing with your face the following emotion for at least 10 seconds:

SURPRISE

**PAYMENT: \$0.35
each video**



a quién se entrevista para un puesto determinado de trabajo, qué estudiantes prestan atención en clase o a qué sospechosos detener.

4.2. Entrenamiento del modelo

Otro de los puntos críticos de intervención en el proceso integral de las IAs, en la fase de desarrollo, es el entrenamiento del modelo, mediante conjuntos de datos que se realiza en el contexto de las ciencias de aprendizaje automático ML. En ese aspecto la colaboración entre Adam Harvey y Jules LaPlace (2021), *Exposing.ai* es una investigación artística que indaga en la ética, los orígenes y las implicaciones de privacidad individual de los conjuntos de datos creados 'a las bravas' y utilizados con posterioridad para reconocimiento facial. La investigación deriva de un proyecto técnico-político anterior de Adam Harvey titulado *MegaPixels* (2017-2020), en la que el artista rastreaba el uso opaco de enormes conjuntos de datos de rostros humanos, por parte de la academia, de corporaciones y de algunos gobiernos. En este nuevo proyecto, *Exposing.ai*, los autores rastrean y analizan cientos de estos conjuntos de datos y encuentran un patrón: millones de imágenes han sido descargadas de Flickr.com, donde las licencias de contenido de la plataforma son permisivas y los datos biométricos abundantes. Como señalan los autores:

Si fuiste usuario de Flickr.com y has subido fotos que contienen rostros u otra información biométrica entre 2004 y 2020, tus fotos pueden haber sido utilizadas para entrenar, probar o mejorar tecnologías de vigilancia de inteligencia artificial para su uso en aplicaciones académicas, comerciales o relacionadas con la defensa (Harvey y LaPlace, 2021).

Parte del objetivo de este proyecto es relatar la compleja historia de cómo las fotografías de ayer se han convertido en los datos de entrenamiento de hoy forma.

Desde una perspectiva interseccional, el artista de origen africano afincado en Londres, Lex Fefegha

trabaja con tecnologías de aprendizaje automático y con temas de raza, justicia social, identidad e interacción social. Su obra *Hip Hop Poetry Bot* (Fefegha, 2021), es una pieza de arte experimental creada para generar respuestas poéticas a preguntas cotidianas en forma de rap y hip hop. En ella Fefegha (2021) investiga en el desarrollo de un algoritmo entrenado en un conjunto de datos personales de letras de canciones de autores negros y explora la generación de voz utilizando GPT-2. La obra de arte existe como prueba de concepto, ya que para construir una pieza completa se necesita un gran conjunto de datos públicos de letras de rap y hip hop en el que se pueda entrenar el algoritmo, y actualmente no existe un archivo público de este tipo. El autor ha lanzado una invitación a artistas, poetas y escritores de rap y hip hop para que se conviertan en colaboradores creativos y contribuyan con sus letras a crear un nuevo conjunto de datos público de música negra.

A partir de estos proyectos podemos ser algo más conscientes de la necesidad de tratar críticamente tanto los procesos de interpretación mecánica hecha por los algoritmos, como las operaciones de reducción e imposición de interpretaciones en manos de autoridades interpretativas.

4.3. Procesos integrales

Acometiendo en sus proyectos el proceso integral de creación e implementación de las IAs queremos destacar los proyectos de dos creadoras, la artista Anna Ridler y la diseñadora Caroline Sindere.

Myriad (Tulips) (Ridler, 2018) es una instalación con miles de fotografías de tulipanes etiquetadas a mano. Estas fotografías se utilizaron posteriormente como conjunto de datos que alimentaban una red generativa antagonista (GAN) para crear tulipanes falsos en sus piezas *Mosaic Virus 2018* y *Mosaic Virus* (Ridler, 2019). El proyecto pone de relieve el proceso integral de la inteligencia artificial (IA), desde el trabajo humano y la recopilación de datos hasta la arquitectura del modelo de aprendizaje automático y la infraestructura material.

Imagen 4. Portada de la aplicación web *Hip Hop Poetry Bot* del artista Lex Fefegha. Fuente: Página web (Fefegha, 2021)

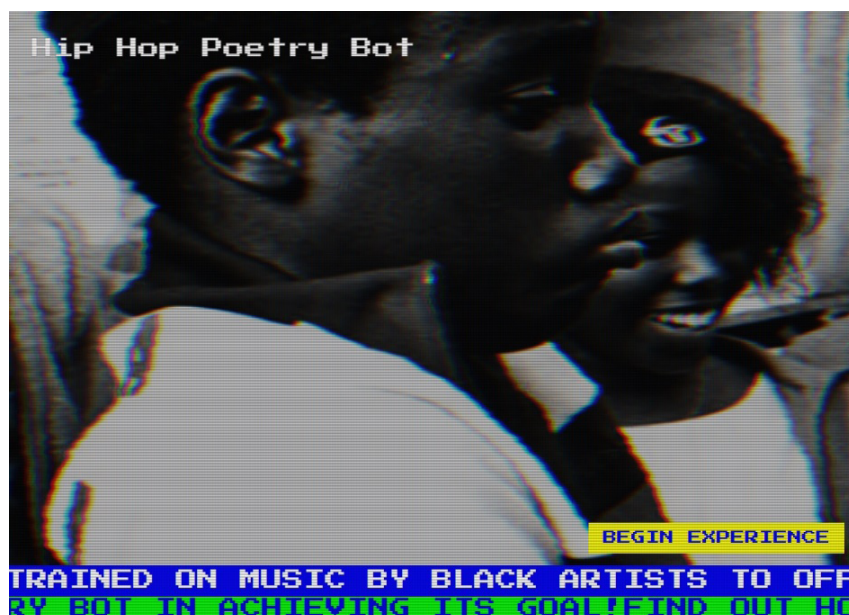


Imagen 5. Mosaic Virus, 2019, de Anna Ridler. 3 pantallas, video instalación GAN.
Fuente: Pantallazo del site de la artista (Ridler, 2019)



Con una propuesta de carácter marcadamente político en el discurso, la obra *Feminist Data Set* (Sinders, 2017) es un proyecto que cuestiona cada paso del proceso integral de la IA. Desde la recopilación de datos, el etiquetado, el entrenamiento con los conjuntos de datos, la selección de un algoritmo a utilizar, el modelo algorítmico. La pieza culmina con el diseño de cómo se coloca ese modelo en un chatbot, planteándose también cómo es el chatbot. Cada paso sirve para cuestionar y analizar el proceso de creación de modelos de aprendizaje automático. En cada fase la autora interpela: ¿es este paso feminista, interseccional? ¿Tiene esta fase un sesgo? ¿Cómo se puede eliminar ese sesgo? En palabras de la propia diseñadora se trata de una práctica artística de justicia social con intención de trascender de cara a lo público y lo social.

Todas estas propuestas artísticas, en las que la IA es herramienta y/o temática, tienen en común la capacidad para facilitar la comprensión de las implicaciones de los sistemas de IA. Las posibilidades de estos proyectos para trasladar a lo social las debilidades e injusticias actuales de estas tecnologías hacen que funcionen casi como herramientas pedagógicas que fomentan la alfabetización de la población no experta en materia de IA. El trabajo cultural y artístico de estas personas previene contra credulidades y servidumbres, y aumenta el valor del uso de nuestra inteligencia reflexiva, y la práctica artística como un «campo de batalla en el que se dirime el sentido y el valor de la experiencia humana» (Garcés, 2017, p. 58).

5. Conclusiones y perspectivas

Asumir acríticamente la noción de interdependencia que como sociedad tenemos ahora mismo con los modelos de aprendizaje automático nos puede llevar a olvidar que sobre esos vínculos y enredos se consolidan relaciones de poder y estructuras que definen y determinan la vida social. En el presente texto se han revisado las prácticas artísticas contemporáneas realizadas sobre tecnologías de inteligencia artificial con el fin de discernir las posibilidades que la creación artística y el diseño crítico tienen de poder

enriquecer e informar sobre nuevos paradigmas y acercarnos a unos modelos tecnológicos de inteligencia artificial más justos. El texto trata de ofrecer unas bases teóricas y prácticas que funcionen como punto de partida a la cuestión inicial ¿Pueden el arte y el diseño crítico enriquecer la reflexión social sobre las diferentes fases de los procesos integrales de la IA? ¿Son capaces las prácticas artísticas que trabajan con o sobre procesos de inteligencia artificial funcionar como agentes de cambio y acercarnos a modos más conscientes y consecuentes sobre la creación y los usos de estas tecnologías?

La comprensión crítica que muestran las piezas descritas permite identificar y analizar las dinámicas de poder que configuran nuestro mundo a nivel intelectual, político, económico y social, dentro de los diferentes discursos en torno a la IA. Se trata de iniciativas que, de manera central en unas ocasiones y transversal en otras, mejoran las nociones básicas y el conocimiento sobre la prevalencia que estas tecnologías tienen en nuestras vidas cotidianas. El planteamiento de la investigación defiende que este tipo de proyectos hacen tangibles y explícitos los sistemas de inteligencia artificial y pueden crear una suerte de alfabetización social crítica en públicos no expertos, aportando luz sobre cuestiones técnicas y sociales clave relacionadas con la IA. La perspectiva es que tanto el arte como el diseño críticos que utilizan IA como herramienta o como temática pueden mejorar la comprensión pública de esta tecnología e iluminar cuestiones técnicas y sociales, fomentando conversaciones transdisciplinares.

Ante la realidad de tener que mejorar la transparencia y la gobernanza algorítmicas en relación a muchos contextos sociales que están atravesados por los procesos de IA en el día a día, estas prácticas pueden ser un mecanismo válido para explicitar los vínculos entre los sistemas de IA y determinadas cuestiones estructurales de poder. En ese aspecto valorar el enfoque experimental que desde la cultura y las artes se puede dar de las IAs involucra a las personas de una forma concreta, sensible y cognitiva que permite que

agentes no expertos interactúen con los sistemas de IA para profundizar en su entendimiento y compromiso social con estas nuevas tecnologías. La investigación pretende demostrar cómo el arte y las experiencias tangibles que de él derivan pueden mediar entre el código informático y la comprensión humana para superar las limitaciones que la legibilidad de los sistemas de IA tiene a día de hoy a nivel social. La puesta en marcha de este tipo de experiencias artísticas puede activar, en personas no expertas, nuevas formas de concebir y comprender estas tecnologías y también nuevos planteamientos sobre para qué utilizarla.

Al mismo tiempo, es necesario resaltar las limitaciones de este artículo, las cuales radican en la ausencia de material empírico de trasvase de lo artístico al tejido social y por ello deja abiertas futuras líneas de investigación que aborden las múltiples dimensiones de esta confluencia. En cualquier caso, la voluntad de esta investigación es articular preguntas y debates abiertos sobre el presente con casos de estudio, que no buscan cerrar las preguntas, sino sostenerlas materialmente. En ningún caso se presentan respuestas cerradas o conclusiones, sino más bien una discusión abierta sobre las posibilidades y las capacidades del arte y el diseño críticos para extender modos más conscientes y consecuentes con las tecnologías de nuestra época.

6. Declaración sobre el uso de LLM

Este artículo no ha utilizado ningún texto generado por un LLM (ChatGPT u otro) para su redacción.

7. Referencias

- Alsina, Pau (2022). Focus: Inteligencia artificial y creación. En *Anuario AC/E de cultura digital 2022* (pp. 136-179). <https://www.accioncultural.es/es/anuario2022>
- Buolamwini, Joy (2016). *How I'm fighting bias in algorithms*. https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms
- Crawford, Kate y Paglen, Trevor (2019). *Excavating AI: The politics of images in machine learning training sets*. <https://excavating.ai/>
- Crawford, Kate y Schultz, Jason (2019). AI systems as state actors. *Columbia Law Review*, 119(7). <https://columbialawreview.org/content/ai-systems-as-state-actors/>
- De Silva, Daswin y Alahakoon, Daminda (2022). An artificial intelligence life cycle: From conception to production. *Patterns*, 3(6) 100489. <https://doi.org/10.1016/j.patter.2022.100489>
- D'Ignazio, Catherine y Klein, Lauren F (2020). *Data feminism*. The MIT Press.
- Fefegha, Lex (2021). *Hip Hop Poetry Bot*. <https://artsexperiments.withgoogle.com/hip-hop-poetry-bot>.
- Garcés, Marina (2017). *Nueva ilustración radical* (8ª). Anagrama.
- García Angulo, Ana, Lucas, Gabriel, Laporta, Javier, Conde, Luis, y Sánchez Coterón, Lara (2019). *Deefakes: Ficción, política y algoritmos*. #edcd19_03. Encuentros de Diseño y Cultura Digital en Madrid. <https://edcd.es/edcd19/>
- Giardina Papa, Elisa (2020). *Cleaning emotional data*. <http://www.elisagiardinapapa.org/>
- Gindin, Irene Lis, Cingolani, Gastón y Rodríguez-Amat, Joan Ramón (2021). *Autoridades interpretativas: Una perspectiva teórica sobre datificación y producción de sentido*. *Palabra Clave*, 24(3), Article 3. <https://doi.org/10.5294/pacla.2021.24.3.6>
- Guerassimov, Youri (Director). (2023). *La Compil des Bleues*. Marcel Agency. <https://www.youtube.com/watch?v=QVNZRHIZVL8>
- Harvey, Adam y LaPlace, Jules (2021). *Exposing AI*. <https://exposing.ai/about/>
- Irani, Lilly y Silberman, Six (2019). *Turkopticon*. <https://turkopticon.net/>
- Ishikawa Sutton, Mai y Ryan, John (2024). Decentralized web principles [A global network of builders and dreamers working to create a better, decentralized web]. *Decentralized Web*. <https://getdweb.net/principles/>
- Kantayya, Shalini (2020). *Coded bias*. <https://www.codedbias.com/>
- Larson, Erik J. (2021). *The myth of artificial intelligence: Why computers can't think the way we do* (Belknap Press). Harvard University Press.
- Manovich, Lev (2023). AI image media through the lens of art and media history. *IMAGE*, 34-41. <https://image-journal.de/ai-image-media-through-the-lens-of-art-and-media-history/>
- Meadows, Donatella H. (2009). *Thinking in systems: A primer*. Earthscan.
- Oliver, Julian, Savičić, Gordan y Vasiliev, Danja (2021). *The critical engineering manifesto*. <https://critical-engineering.org/es>
- Ridler, Anna (2018). *Myriad (Tulips)*. <http://annaridler.com/myriad-tulips>
- Ridler, Anna (2019). *Mosaic Virus*. <https://annaridler.com/mosaic-virus>
- Rouvroy, Antoinette (2020). Algorithmic governmentality and the death of politics. *Green European Journal*. <https://www.greeneuropeanjournal.eu/algorithmic-governmentality-and-the-death-of-politics/>
- Sánchez Coterón, Lara (2020). Diálogos críticos con sistemas complejos. *Revista Científica Arte e Investigación*. <https://doi.org/10.24215/24691488e060>
- Sinders, Caroline (2017). *Feminist data set*. <https://carolinesinders.com/feminist-data-set/>
- Vicente, Lucía y Matute, Helena (2023). Humans inherit artificial intelligence biases. *Scientific Reports*, 13(1), Article 1. <https://doi.org/10.1038/s41598-023-42384-8>
- Vora, Kalindi y Atanasoski, Neda (2019). *Surrogate humanity: Race, robots, and the politics of technological futures*. Duke University Press.
- Wilson, Stephen (2005). Artificial Intelligence research as art. En *Mechanical bodies, computational minds: Artificial intelligence from automata to cyborgs* (MIT Press).