Diseña 19

Crear arte con IA *de* forma responsable: *una* guía *de* campo para artistas

Cómo citar este artículo: Leibowicz, C. R.; Saltz, E., & Coleman L. (2021). Crear arte con IA de forma responsable: una guía de campo para artistas. *Diseña*, (19), Article.5. https://doi.org/10.7764/disena.19.Article.5

DISEÑA | 19
AGOSTO 2021
ISSN 0718-8447 (impreso)
2452-4298 (electrónico)

Proyecto
Recepción ☑
02 NOV 2020
Aceptación ☑
23 JUN 2021

Original English Version here

Claire R. Leibowicz
Partnership on AI

Emily Saltz
The New York Times

Lia Coleman
Rhode Island School of Design

Las herramientas de aprendizaje automático utilizadas para generar medios de comunicación sintéticos permiten la expresión creativa, pero también pueden generar contenidos que inducen a error y causan daño. La *Guía de campo para el arte con IA responsable* (*Responsible AI Art Field Guide*) ofrece un punto de partida para que diseñadores, artistas y otros creadores puedan utilizar las técnicas de inteligencia artificial (IA) de forma responsable y cuidadosa. Sugerimos a los artistas y diseñadores que utilizan la IA que sitúen su trabajo en el contexto más amplio de la IA responsable, prestando atención a las consecuencias perjudiciales que potencial e involuntariamente podría conllevar su trabajo, tal como se entiende en los ámbitos de la seguridad de la información, la desinformación, el medio ambiente, los derechos de autor y los sesgos en los medios sintéticos apropiativos. En primer lugar, describimos las dinámicas más generales de los medios generativos para subrayar que los artistas y diseñadores que utilizan la IA operan al interior de un campo con características sociales complejas. A continuación, describimos nuestro proyecto, una guía centrada en cuatro puntos clave para controlar el ciclo de vida de la creación con IA: (1) el conjunto de datos, (2) el código del modelo, (3) los recursos de entrenamiento y (4) la publicación y la atribución. Por último, destacamos la importancia que tienen estas instancias de control para los artistas y diseñadores que utilizan la IA, ya que ofrecen puntos de partida o provocaciones para construir un campo de IA creativo y a la vez atento a las repercusiones sociales de sus trabajos.

Palabras clave

Medios sintéticos

Arte con IA

IA responsable

Ética de la IA

Medios generativos

Claire R. Leibowicz—Licenciada en Psicología e Informática, Harvard University. Máster en Ciencias Sociales de Internet, University of Oxford (como becaria Clarendon). Es directora del programa de Inteligencia Artificial e Integridad de los Medios de Partnership on AI, una organización global sin fines de lucro dedicada a la IA responsable. Bajo su dirección, el equipo de IA e Integridad de los Medios investiga el impacto de las tecnologías emergentes de IA en los medios digitales y la información en línea. Es becaria 2021 de periodismo en *Tablet Magazine*, donde explora asuntos que se sitúan en la intersección de la tecnología, la sociedad y la cultura digital, y candidata entrante al doctorado en el Oxford Internet Institute. Algunas de sus últimas publicaciones son "Encounters with Visual Misinformation and Labels Across Platforms: An Interview and Diary Study to Inform Ecosystem Approaches to Misinformation Interventions" (con E. Saltz y C. Wardle; *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, nº 340) y "The Deepfake Detection Dilemma: A Multistakeholder Exploration of Adversarial Dynamics in Synthetic Media" (con A. Ovadya y S. McGregor; *Proceedings of the 2021 ACM Conference on Artificial Intelligence, Ethics, and Society*).

Emily Saltz—Máster en Human-computer Interaction, Carnegie Mellon University. Investigadora de UX, estudia los medios de comunicación y la desinformación, trabajando con organizaciones como Partnership on AI y First Draft. Dirigió la UX para The News Provenance Project en The New York Times, donde trabaja como investigadora de UX. Entre sus trabajos se cuentan una colaboración con Oobah Butler en un artículo de opinión generada por IA sobre ser capturado por la IA (*The Independent*, 2021); exploraciones con *softwares* de predicción de texto como "Human-Human Autocompletion" (presentada en WordHack en Babycastles, 2020) y "Super Sad Googles" (presentada en Eyeo 2019); y "Filter Bubble Roulette", una experiencia de realidad virtual móvil para habitar los *feeds* de redes sociales específicos de usuario (presentado en The Tech Interactive, San Jose, 2018).

Lia Coleman—Licenciada en Informática, Massachusetts Institute of Technology. Artista, investigadora de IA y educadora, es profesora adjunta en la Rhode Island School of Design, donde enseña a crear obras de arte con aprendizaje automático. Es autora de "Machines Have Eyes" (con A. Raina, M. Binnette, Y. Hu, D. Huang, Z. Davey y Q. Li; en *Big Data. Big Design: Why Designers Should Care About Machine Learning*; Princeton Architectural Press, 2021), "'Art'ificial" (con E. Lee; *Neocha Magazine*, 2020), y "Flesh & Machine" (con E. Lee; *Neocha Magazine*, 2020). Entre sus talleres y charlas recientes destacan "How to Play Nice with Artificial Intelligence: Artist and AI Co-creation" (presentada en la Universidad de Arte y Diseño Burg Giebichenstein, 2021); "A Field Guide to Making AI Art Responsibly" (presentada en Art Machines: International Symposium on ML and Art), y "How to Use AI for Your Art Responsibly" (presentada en Mozilla Festival, 2020 y Gray Area, 2020).

Crear arte con IA de forma responsable: una guía de campo para artistas

Claire R. Leibowicz
Partnership on AI
Nueva York, EE. UU.
claire@partnershiponai.org

Emily Saltz
The New York Times
Nueva York, EE. UU.
essaltz@gmail.com

Lia Coleman
Rhode Island School of Design
Providence, EE. UU.
liailiad@gmail.com

CREAR ARTE CON IA DE FORMA RESPONSABLE: UNA GUÍA DE CAMPO PARA ARTISTAS

Antecedentes del problema

Las herramientas de Inteligencia Artificial (IA) utilizadas para generar medios de comunicación son cada vez más accesibles (Lomas, 2020; Nicolaou, 2020), ofreciendo un potencial para que los medios que han sido generados sintéticamente engañen y causen daño. Si bien los *deepfakes* o las imágenes y vídeos generados por la IA han captado la atención del público, lo cierto es que hasta las técnicas *low-tech*, como las *cheap fakes* —utilizadas con frecuencia, por ejemplo, en vídeos de políticos— pueden ser utilizadas para alterar la percepción de las figuras públicas y los acontecimientos de interés general (Chesney & Citron, 2018; Paris & Donovan, 2019).

Sin embargo, las mismas herramientas de IA proporcionan a los artistas y diseñadores un nuevo campo creativo con posibilidades únicas. Naoko Hara, por ejemplo, utiliza imágenes extraídas de su propio trabajo de animación como datos para generar arte (Hara, 2020). A su vez, Derrick Schultz aprovecha imágenes de ilustraciones florales para generar arte con IA (Figura 1). Las técnicas de medios sintéticos también se han aprovechado en el cine para proteger la privacidad (Li & Lyu, 2019). En la película "Welcome to Chechnya", estrenada en 2020, los cineastas protegieron las identidades de los sujetos que hablaban de experiencias LGBTQ+, permitiendo que las personas contaran sus historias de forma segura (Rothkopf, 2020). Recientemente, los artistas que trabajan con IA han estado utilizando la tecnología incluso para cuestionar sus efectos en la sociedad, incluyendo

Figura 1: Obra generada con CycleGAN a través del uso de imágenes de un gato e ilustraciones de flores, parte del proyecto "faces2flowers", de Derrick Schultz (2019). El posteo incluye un enlace para que los lectores puedan experimentar por sí mismos con el modelo fuente en la plataforma RunwayML. Fuente: Schultz, 2019. Imagen recuperada de https://artificial-images.com/project/faces-to-flowers-machine-learning-portraits/. Cortesía de Derrick Schultz.

los que ofrece el propio medio sintético. En 2019, la instalación "Spectre", de Bill Posters y Daniel Howe, presentó un vídeo *deepfake* de Mark Zuckerberg para ilustrar la influencia de Facebook en el comportamiento de los usuarios.

A medida que artistas, diseñadores y otros creadores aprovechan las tecnologías de IA, específicamente los métodos de aprendizaje automático, es crucial que comprendan, a través de la mirada más amplia de las comunidades de investigación de medios sintéticos e IA, el contexto y los posibles daños que podría ocasionar su trabajo, así como también es importante que esas comunidades de investigación atiendan a las necesidades y percepciones particulares de los creadores independientes. Si bien hay abundantes directrices de IA responsable para

los investigadores y las empresas tecnológicas (Gebru et al., 2018; Mitchell et al., 2019; Raji & Yang, 2019), estas suelen dirigirse a los ingenieros y más ampliamente al personal técnico que trabaja en las grandes plataformas de medios sociales y en las compañías de tecnologías de la información y la comunicación, en lugar de enfocarse en los objetivos y las necesidades únicas de los artistas que equilibran la expresión creativa y el impacto social positivo con los posibles daños de sus creaciones.

Objetivos

En este contexto, hemos creado un fanzine digital, la *Guía de campo para el arte con IA responsable* (Figura 2), cuyo fin es equipar a artistas y diseñadores que trabajan con IA con conocimientos sobre cómo crear arte con IA de forma responsable. Dicha guía surgió del Programa de IA e Integridad de los Medios de Partnership on AI, que investiga el impacto de la tecnología emergente de IA en los medios digitales y la información en línea (Saltz et al., 2020). La guía está estructurada en torno a preguntas que los artistas y los diseñadores que utilizan la IA deben plantearse a lo largo de todo el ciclo de vida de su práctica creativa, para así situar de mejor manera su trabajo dentro de una práctica responsable con la IA, combinando ideas de prácticas responsables de IA con tácticas de diseño destinadas específicamente a los creadores que utilizan IA. También ofrece buenas prácticas emergentes, las que son extraídas de las experiencias multidisciplinares de artistas, profesionales, investigadores y miembros de la comunidad académica de campos como la comunicación, la informática, la ciencia forense de medios, la sociología y los estudios de medios.

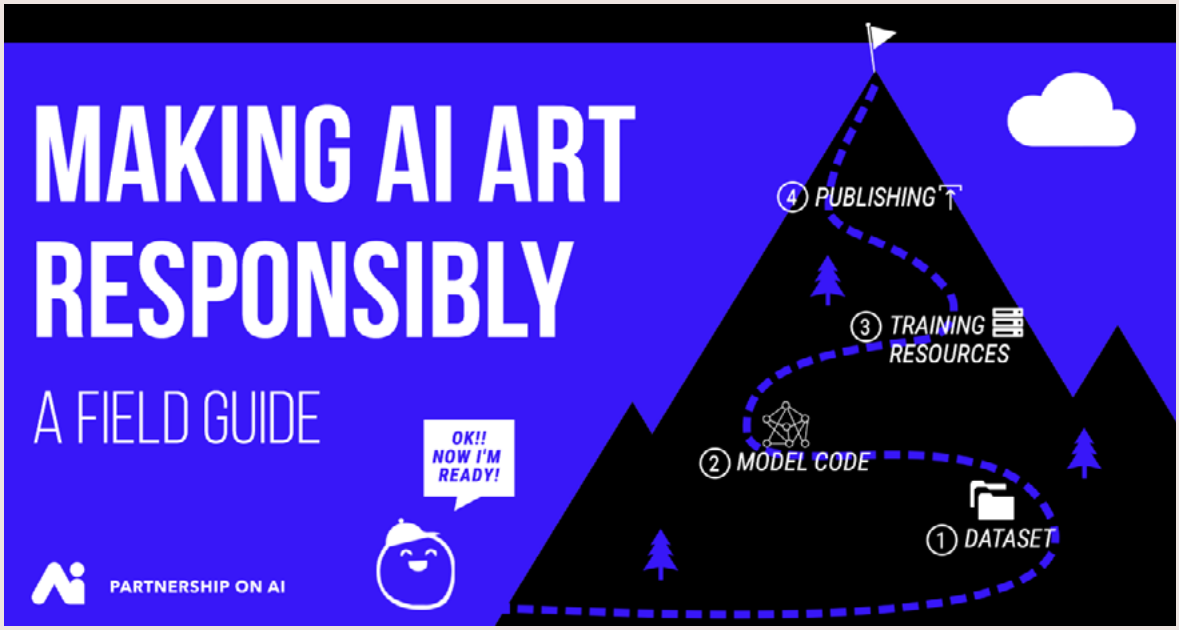

Figura 2: Imagen de la portada de la *Guía de campo para el arte con IA responsable*, de Emily Saltz, Lia Coleman y Claire R. Leibowicz (Partnership on AI, 2020). Captura de pantalla: Claire R. Leibowicz, 2020. Obtenido de Medium: https://medium.com/partnership-on-ai/a-field-guide-to-making-ai-art-responsibly-f7f4a5066ee

Marco conceptual

Nuestra definición de "IA responsable" considera tanto las críticas a la forma en que las corporaciones crean y lanzan productos de IA en la actualidad, como las críticas más amplias a las estructuras corporativas que otorgan poder a los sistemas de IA, incluyendo los derechos laborales y la equidad de los trabajadores (Diehm & Sinders, 2020; Rakova et al., 2020). En el campo de los medios generados por IA, garantizar una práctica responsable con IA requiere prestar especial atención a los casos de uso malicioso de la tecnología. Basándonos en nuestro quehacer y en la colaboración con muchas organizaciones que trabajan en —y en torno a— la IA responsable y los medios sintéticos en Partnership on AI, entendemos que las consecuencias perjudiciales de los medios sintéticos implican, en gran medida, los ámbitos de la desinformación, la seguridad de la información, los contenidos sesgados/apropiados, los derechos de autor y el medio ambiente. Por supuesto, en el arte, el diseño y el entretenimiento también hay casos de uso no malicioso de los medios generados con IA. Definimos el arte con IA como aquel que incluye a todas las obras realizadas con intención creativa mediante técnicas en las cuales los programas informáticos acceden a los datos y aprenden automáticamente de ellos con una mínima intervención humana. La *Guía de campo para el arte con IA responsable* busca recopilar las ideas emergentes y las mejores prácticas, tanto de los artistas que trabajan con IA como de las comunidades de investigación de IA responsable y medios sintéticos, para ayudar a guiar los procesos de reflexión que generan obras creativas o diseñan productos con IA (mitigando, al mismo tiempo, las consecuencias no deseadas).

Antecedentes

La mayor parte de la atención suscitada por los impactos negativos de los medios sintéticos ha procedido de plataformas tecnológicas enfocadas en mitigar la desinformación. Plataformas como Facebook han realizado grandes inversiones para prepararse ante el uso de medios manipulados que podrían afectar a la opinión pública y difundir desinformación, centrándose específicamente en políticas e innovaciones de aprendizaje automático para combatir la desinformación visual (Bickert, 2020; Dolhansky et al., 2020; Leibowicz, 2020; Roth & Achuthan, 2020). Si bien actualmente la mayor parte de la desinformación visual no está siendo generada por la IA, esta se utiliza con mayor frecuencia para otros casos de uso malicioso que afectan a la seguridad de la información: crear imágenes para la explotación sexual no consentida. Según un informe divulgado en 2019 por Sensity, una empresa de inteligencia de amenazas visuales, el 96 por ciento de los *deepfakes* en línea son de naturaleza pornográfica (Patrini, 2019).

Durante los últimos años, los investigadores han intentado realizar compendios de las mejores prácticas para el uso de conjuntos de datos de aprendizaje automático, como "Datasheets for datasets" (Gebru et al., 2018), "Model

Cards for Model Reporting" (Mitchell et al, 2019) y "ABOUT ML" (Raji & Yang, 2019), con el fin de ayudar a las personas a abordar de mejor manera el sesgo inherente a los modelos de aprendizaje automático, identificando además de qué manera los conjuntos de datos pueden estar sesgados hacia ciertos atributos. Junto con ello, organizaciones como la Algorithmic Justice League han surgido para comunicar los posibles daños y sesgos de las tecnologías de IA (Buolamwini, 2016). Gran parte del trabajo de la Liga de Justicia Algorítmica se centra en el despliegue de sistemas de alto impacto como el reconocimiento facial. Sus cuatro principios fundamentales incluyen el consentimiento afirmativo; la transparencia significativa; la supervisión y la responsabilidad continua; y la crítica accionable (Buolamwini, 2016). Aunque todos estos elementos pueden aplicarse a los sistemas de IA de doble riesgo para hacerlos más responsables, incluido el arte creado con IA, será de vital importancia contextualizar estos principios de IA responsable a partir de los objetivos y motivaciones de los creadores, diseñadores y artistas que utilizan técnicas de IA. A diferencia del impacto medioambiental negativo que implica generar medios sintéticos utilizando técnicas de IA, los que son independientes de las motivaciones del creador, el uso responsable de la IA para mitigar el sesgo y el contenido apropiativo requiere la sensibilidad de las motivaciones del creador.

Aunque muchos recursos de IA responsable ofrecen detalles técnicos y bases lógicas relevantes para los creadores independientes, se centran principalmente en las prácticas de aprendizaje automático a gran escala en la industria y el mundo académico. En consecuencia, es poco probable que estos conocimientos lleguen a los artistas y diseñadores que utilizan la IA en contextos creativos. Aunque existen marcos de diseño crítico que son relevantes para el diseño de IA, como los Principios de Justicia en el Diseño (Costanza-Chock, 2018), es posible que para un creador no esté claro cómo debe aplicarse en la práctica un principio como «priorizamos el impacto del diseño en la comunidad por sobre las intenciones del diseñador», específicamente en lo que respecta a los pasos involucrados en la IA creativa, como la creación de conjuntos de datos, el código del modelo, los recursos de entrenamiento y la publicación.

Existen indicios de una creciente literatura que explora algunas consideraciones para el uso responsable de la IA en las prácticas creativas; sin embargo, esta debería ampliarse para incorporar recomendaciones prácticas que permitan navegar por el proceso creativo y los intercambios inherentes a tales actividades. Esto es lo que pretendemos llevar a cabo con nuestra *Guía de campo para el arte con IA responsable*. Las complejidades de esta tarea son evidentes en trabajos como el de Lyons (2020), quien ofrece una evaluación crítica de la exposición "Training Humans", de Kate Crawford y Trevor Paglen. Dicha exposición, que pretendía criticar las prácticas corporativas empleadas para entrenar sistemas de visión por ordenador, se presentó junto al valioso trabajo escrito "Excavating AI: The Politics

of Images in Machine Learning Training Sets" (Crawford & Paglen, 2019). Sin embargo, Lyons, uno de los coautores del conjunto de datos JAFFE que critican los autores, señala que, al criticar prácticas corporativas como el uso de imágenes y vídeos faciales sin consentimiento, los mismos Paglen y Crawford también reprodujeron y exhibieron esas mismas imágenes sin consentimiento. Lyons se refiere a este descuido, en este caso para uso artístico, como un "doble estándar ético". Este debate pone de manifiesto la necesidad de seguir evaluando de forma crítica el uso de datos humanos en los sistemas de IA con fines artísticos y de diseño, así como la consideración de los derechos de autor y la seguridad de la información, incluido el uso de imágenes personales sin el consentimiento informado y los términos de uso de los conjuntos de datos utilizados.

DESARROLLO Y RESULTADOS PARCIALES

Marco metodológico

Desarrollada para abordar esta carencia, la *Guía de campo para el arte con IA responsable* surgió de los aportes de múltiples *stakeholders*, artistas que usan la IA, investigadores de la desinformación visual, ingenieros de aprendizaje automático y creadores de políticas públicas. La información inicial se obtuvo en una charla realizada en julio de 2020 con Gray Area, un centro cultural de arte y tecnología ubicado en San Francisco, así como a través de conversaciones con los miembros de Partnership on AI, una organización global sin fines de lucro dedicada a la IA responsable que cuenta con más de 100 organizaciones asociadas de la sociedad civil, la industria, los medios de comunicación y el mundo académico. En la charla organizada con Gray Area participaron más de 50 personas de diversos ámbitos, entre las cuales se contaban artistas que trabajan con IA, investigadores de la IA y personas interesadas en la política tecnológica, entre otras. Facilitamos que estos participantes ofrecieran sus comentarios sobre la primera fase de la guía en un documento de trabajo, lo que nos permitió perfeccionar las recomendaciones y los puntos de control. Lia Coleman, artista y diseñadora de IA, diseñó la guía como un fanzine en línea que conduce a un hipotético artista que trabaja con IA a través de los puntos de control para que pueda crear arte con IA de forma responsable, centrándose en cuatro elementos: (1) el conjunto de datos, (2) el código del modelo, (3) los recursos de entrenamiento y (4) la publicación y la atribución. Así, nuestro enfoque metodológico incorpora un proceso de consulta a profesionales y académicos de las comunidades de Partnership on AI y Gray Area, una revisión de proyectos de diseño y arte relacionados con la IA, una revisión de literatura que apunta a la falta de directrices para crear arte responsable con IA que favorezcan las necesidades únicas de los artistas, y la experiencia de dos artistas que trabajan con IA, Lia Coleman y Emily Saltz, quienes se basan en sus experiencias personales para perfeccionar la guía. Mientras gran parte de los escritos sobre la IA responsable se centran en el dis-

curso académico o en los profesionales que se desempeñan en gran medida dentro de la tecnología y la industria, nosotros nos centramos en la comunidad artística y de diseño que trabaja con IA, una audiencia distinta para la IA responsable, actualmente desatendida en la literatura (Rakova et al., 2020). Además de las preguntas, incluimos ejemplos de casos para subrayar las implicaciones de la creación con IA en diferentes aspectos sociales como propiedad, impacto medioambiental, atribución, explicabilidad y privacidad.

Estrategia metodológica del proyecto

Estratégicamente, la guía evita ser prescriptiva. Dado que se trata de un campo naciente y en evolución, el de la IA creativa, muchos temas están sujetos a debate y deben construirse a través de prácticas de prueba y error, y en conversación con el campo de la desinformación visual, que cambia rápidamente tanto en la industria como en el mundo académico. Otros actores en el campo de los medios sintéticos han adoptado este enfoque; por ejemplo, Twitter ha establecido recientemente que es necesario que sus respuestas políticas a la desinformación visual sean "documentos vivos", subrayando que están «dispuestos a actualizarlos y ajustarlos cuando [ellos] se encuentren con nuevos escenarios» (Twitter Safety, 2020). Lo mismo puede decirse de nuestra guía de arte y diseño con IA para navegar responsablemente por la creación. Si bien debemos aspirar a marcos más seguros o herméticos, en esta etapa del desarrollo de la IA debemos seguir siendo adaptables.

Descripción de la propuesta

Antes de que los artistas y diseñadores de IA empiecen a crear y a lidiar con los cuatro puntos de control, hacemos hincapié en que el punto de partida debe ser examinar por qué están utilizando técnicas de IA en su trabajo. Los futuros usuarios de la IA deben considerar qué objetivos tienen para aplicar las técnicas de IA, cómo entienden el papel de las tecnologías de IA en la sociedad y si están utilizando la IA para comentar cuestiones sociales o políticas.

Punto de control 1: Conjunto de datos. El primer punto de control de la guía de campo considera el conjunto de datos, la base del trabajo personal con IA (Figura 3). Los artistas y diseñadores deben considerar la selección de datos de entrenamiento como un acto curatorial inherentemente subjetivo, evitando explotar el trabajo de otros creadores o causar daño a través de lo que está o no está representado en el conjunto de datos. Por ejemplo, un artista de IA llamado Arfa experimentó en carne propia este tipo de problemas de derechos de autor después de generar imágenes de personajes *furry* a partir de un modelo StyleGAN2 entrenado con más de 55.000 obras de arte *fandom furry*, extraídas sin permiso de un foro de arte *furry* (Mix, 2020). Los creadores originales del arte *furry* protestaron porque el proyecto de Arfa, titulado "This Fursona Does not Exist", no respetó su trabajo,

CHECKPOINT 1:

DATASET

El conjunto de datos es la base de tu arte generado con IA. La selección de un subconjunto de medios como ***datos de entrenamiento*** es un acto curatorial inherentemente subjetivo: piensa cuidadosamente cómo seleccionas tus datos en bruto para evitar explotar el trabajo de otros creadores o causar daño a través de quién y qué está representado (y quién o qué no lo está). Algunas preguntas que debes hacerte:

¿DE DÓNDE PROCEDEN LOS DATOS DE ENTRENAMIENTO? ¿CUÁL ES MI RELACIÓN CON ELLOS?

- ☐ ¿Cuál es el contexto histórico y social de los medios de comunicación que estoy utilizando como datos de entrenamiento?
- ☐ ¿Estoy ***extrayendo*** datos de un foro público o de una plataforma social? Si es así, ¿cómo me relaciono con estas comunidades? ¿De qué maneras tengo más o menos poder que otros miembros de la comunidad?
- ☐ ¿Hay contenidos en mi conjunto de datos que puedan infringir un derecho de autor válido, son de dominio público o están etiquetados como *creative commons* para uso no comercial?
- ☐ ¿Estoy utilizando un conjunto de datos existente? Si es así, ¿comprendo cómo y por qué fue creado?

CONJUNTOS DE DATOS DE OBRAS DE ARTE PERSONALES

Esteban Salgado es un artista que trabaja con IA y *collage* que crea sus propios conjuntos de datos. Salgado genera algorítmicamente miles de formas vectoriales abstractas en Adobe Illustrator y entrena los ***modelos de StyleGAN2*** en ellas para crear manchas animadas meditativas.

Figura 3: Imagen de la *Guía de campo para el arte con IA responsable* (página 8,), de Emily Saltz, Lia Coleman y Claire R. Leibowicz (Partnership on AI, 2020). Captura de pantalla: Claire R. Leibowicz, 2020. Obtenido de Medium: https://medium.com/partnership-on-ai/a-field-guide-to-making-ai-art-responsibly-f7f4a5066ee

por cuanto Arfa se benefició de la utilización del arte de los creadores originales, quienes no otorgaron su permiso ni tuvieron la opción de no participar. Las similitudes entre las obras originales y los resultados de los modelos también dieron lugar a quejas por violación de los derechos de autor. Más allá de los derechos de autor, los artistas que trabajan con IA también deben tener en cuenta la diversidad del conjunto de datos y si respetan o no a los creadores y sujetos de los datos. Por el contrario, el artista Esteban Salgado crea sus propios conjuntos de datos generando algorítmicamente miles de formas abstractas en Adobe Illustrator y entrenando los modelos StyleGAN2 con las formas (Salgado, 2020). Tanto si se crean datos propios como si no, es necesario tener en cuenta el contexto histórico y social de los medios de comunicación utilizados como datos de entrenamiento, si se recogen datos de foros públicos o plataformas sociales o no, y las restricciones de derechos de autor del conjunto de datos.

Punto de control 2: Código del modelo. Una vez que los creadores de IA se deciden por un conjunto de datos, deben entrenar sus modelos con esos datos. Alentamos a todos los que estén pensando en utilizar la IA para el arte o el

diseño a que conozcan la historia y la cadena de suministro de las arquitecturas de IA que están utilizando, ya que esto les permitirá respetar mejor a las personas que han contribuido al código de sus modelos, reconociendo así a las personas y el trabajo dedicado al código utilizado para producir las obras, y evaluando de forma crítica cómo se ha desarrollado y etiquetado el código base. También hay complicadas cuestiones de propiedad entre los marcos, las herramientas, los modelos y los resultados de la IA. Por ejemplo, el artista de IA Robbie Barrat puso a disposición del público un modelo GAN que generaba imágenes falsas basadas en imágenes de pinturas al óleo. En 2018, luego de replicar el método de la red neuronal de Barrat para producir una pieza llamada "Edmond de Belamy, from La Famille de Belamy", el colectivo de artistas Obvious vendió una pieza enmarcada por 432.500 dólares estadounidenses. Barrat no recibió nada de este dinero, lo que lleva a plantear cuestionamientos sobre la propiedad y el crédito en el mundo del arte generado con IA (Simonite, 2018). Las preguntas sobre la propiedad se complican aún más debido a las percepciones antropomorfizadas que conciben el arte generado con IA como obras creadas por la IA como agente, en lugar de obras creadas por personas que utilizan la IA como herramienta, lo que fue recientemente explorado por Epstein y sus colegas (2020).

Punto de control 3: Recursos de entrenamiento. Después de decidir qué datos se van a entrenar y con qué código se entrenarán, los creadores de IA necesitan una o varias máquinas GPU y otros recursos de entrenamiento para realmente entrenar los modelos. Este proceso puede consumir muchos recursos y tener una considerable huella de carbono. El entrenamiento de un solo modelo de IA, como el popular modelo de aprendizaje profundo Transformer, puede emitir más de 626.000 libras de dióxido de carbono, lo que equivale a casi cinco veces lo que emite un coche estadounidense medio durante su vida útil (Hao, 2019). Dado que comúnmente los modelos son entrenados muchas veces, las emisiones para el entrenamiento a gran escala pueden resultar significativas. Para entrenar modelos de manera responsable, los artistas deben considerar que podrían reducir los costos ambientales a través de métodos como el aprendizaje de transferencia, evitando así entrenar modelos desde cero. Para calcular las emisiones de carbono de la GPU que se esperan del entrenamiento se pueden utilizar herramientas como la Calculadora de Emisiones de Aprendizaje Automático (Lacoste et al., 2019).

Punto de control 4: Publicación y atribución. Alentamos a los artistas, una vez que han entrenado a los modelos y están listos para compartir su trabajo, a ser tan transparentes sobre su proceso como sea posible para que otros puedan aprender de su experiencia. Sin embargo, también es importante recordar que otros pueden encontrar y utilizar indebidamente el trabajo y los productos de IA con intenciones lucrativas o por motivos políticos (Moisejevs, 2019). Por lo tanto, los artistas y diseñadores de productos de IA deben considerar las amenazas y las consecuencias

no deseadas asociadas a la publicación del trabajo, sopesando los costos y beneficios asociados a la liberación de modelos, códigos y conjuntos de datos. Los creadores de IA pueden fijarse en el campo del aprendizaje automático explicable para considerar cómo hacer que su código sea accesible a otros y, al mismo tiempo, limitar la probabilidad de que actores maliciosos intenten armarse con las técnicas de un creador de IA (Bhatt et al., 2020). Varios investigadores de plataformas tecnológicas han empezado a considerar formas de publicar el trabajo para garantizar que no sirva para armar a actores maliciosos que busquen generar medios sintéticos para engañar o dañar, y los artistas y diseñadores deberían hacer lo mismo (Leibowicz et al., 2020).

CONCLUSIÓN

Resultados

La *Guía de campo para el arte con IA responsable* equipa a los artistas con las mejores prácticas emergentes y ofrece puntos de control para explorar su trabajo. Aunque el grueso del proyecto pretende servir de provocación más que como prescripción, se incluyen varias buenas prácticas al final de la obra. Llegamos a la conclusión de que el camino menos arriesgado para los creadores de IA es generar su propio conjunto de datos a través de medios originales como la ilustración, la fotografía, el texto y el vídeo. Si no es así, los creadores de IA deben acreditar el trabajo de otros siempre que sea posible. Esto se aplica a los creadores que trabajaron en el conjunto de datos que se usa, así como a las personas que han compartido su código. Además, si los artistas y diseñadores extraen trabajos de Internet, lo más responsable es dar prioridad a los trabajos de dominio público o directamente pedir permiso a aquellos cuya identidad y/o trabajo aparezca en el conjunto de datos. Hoy, crear con IA de forma responsable también implica prestar atención al impacto medioambiental de las propias contribuciones: los creadores de IA deben planificar formas de minimizar los recursos ambientales de entrenamiento utilizando el aprendizaje por transferencia a partir de un modelo previamente entrenado. Por último, los creadores deben documentar su trabajo en detalle para permitir que otros aprendan de su proceso y lo critiquen, sin dejar de ser sensibles a que eventualmente otros actores con intenciones maliciosas puedan armarse con sus métodos artísticos para crear medios sintéticos. Si bien la documentación técnica exhaustiva no es una práctica comúnmente asociada a la producción artística, así como se ha convertido en algo más habitual a pesar de la falta de precedentes en los campos de investigación y desarrollo de la IA, los artistas que utilizan código también deberían incorporar dichas prácticas en sus flujos de trabajo (Gebru et al., 2018; Mitchell et al., 2019; Raji & Yang, 2019).

Evaluación

El proyecto *Guía de campo para el arte con IA responsable* se esfuerza por situar el emergente campo de la IA creativa entre las comunidades más amplias de la IA responsable y los medios sintéticos. Esta guía de campo es la primera de su tipo que se centra en incorporar a los artistas y diseñadores de IA en las conversaciones sobre la IA responsable, considerando las formas en que pueden atender a los asuntos que surgen en el espacio de la IA responsable mediante la referencia a estudios de casos concretos. Aunque actualmente la guía de campo ofrece un valioso punto de partida para que los creadores de IA sigan un «sinuoso camino de cuestionamientos», la consideramos un paso inicial y un documento vivo que los creadores de IA podrán reconsiderar para continuar dándole forma (Saltz et al., 2020, p. 15). Esperamos que los creadores de IA, así como quienes se han formado en las clases de arte creado con IA, como los cursos de Imágenes Artificiales, puedan aprovechar esta guía y aportar comentarios a medida que cultivan sus habilidades (Schultz, 2020). Del mismo modo en que muchos han reclamado una formación ética en las clases de informática, también las aulas de arte y diseño que introducen y enseñan metodologías y herramientas de IA deberían considerar formas de pensar en el desarrollo y el despliegue responsable de la IA (Grosz et al., 2018).

Hemos comenzado a probar la utilidad y la aplicabilidad de la guía con una cohorte de 12 estudiantes de posgrado que asisten al curso de primavera 2021 *Exhibiting Transdisciplinary Research* en la Escuela de Diseño de Rhode Island. Los estudiantes compilaron sus propios conjuntos de datos, entrenaron los modelos StyleGAN e incorporaron los resultados generados a sus exposiciones finales. Se les pidió que cada semana escribieran reflexiones sobre el punto de control correspondiente de la guía: conjunto de datos, código del modelo, recursos de entrenamiento y publicación. Además de recopilar sus reflexiones escritas, al final realizamos entrevistas verbales con los estudiantes, obteniendo así una valiosa información sobre su proceso utilizando la guía.

Más allá de las aulas, los artistas que utilizan IA han empezado a probar la guía y han documentado su uso, quizá como un elemento de transparencia para sus procesos creativos en sí mismos, lo que destaca su compromiso con la producción creativa responsable. El colectivo Moving Target, formado por Alexa Steinbrück, Natalie Sontopski y Amelie Golfuß, utilizó la guía al producir "Latent Riot", una serie de carteles artificiales de protesta producidos por una red neuronal generativa en 2021. Entrenaron un StyleGAN con imágenes de carteles de protesta de la Marcha de las Mujeres de Boston de 2017, generados a mano. Las artistas destacaron que el hecho de contar con una guía específica que sintetizaba los principios de la IA responsable para los desafíos particulares del dominio del arte producido con IA hizo posible crear material ético y responsable.

Esperamos que en el futuro otros artistas que trabajan con IA traten la guía como un documento vivo y se pongan en contacto con las creadoras para darnos su opinión y sus conocimientos sobre su experiencia con la guía. En particular, nos interesa conocer las opiniones de los artistas que generan obras con IA que no solo la están utilizando como una herramienta para crear de forma más amplia, sino también para cuestionar la IA como herramienta con profundas repercusiones sociales. Puede haber oportunidades para que los artistas aprovechen de formas artísticamente evocadoras las herramientas de la IA, contribuyendo así, explícitamente, al campo de la investigación responsable de la IA. El uso de estas herramientas, y el despliegue responsable de la IA en el arte creado con IA, permite a los artistas enfrentarse a preguntas complejas sobre el impacto de la IA en la sociedad. Comprender de qué manera ocurre esto representa un caso especialmente interesante de uso del arte con IA.

Conclusión

Los creadores de IA pueden aprovechar el potencial expresivo de la tecnología de manera responsable. Para ello será necesario reflexionar y prestar atención a la forma en que su trabajo encaja en el ámbito más amplio de la IA responsable y en las vastas y complejas dinámicas sociales. Derivada tanto de la experiencia práctica de artistas y diseñadores que utilizan la IA como de las ideas de investigadores interdisciplinarios de la IA y expertos en medios, la guía de campo ofrece una lista de preguntas y buenas prácticas emergentes que pretenden ser un punto de partida para que los creadores que utilizan la IA puedan reflexionar en forma crítica sobre el impacto social de su trabajo. De este modo, se puede reforzar el poder que las imágenes generadas tienen para contar historias, embellecer, arrojar luz e incluso simplemente ofrecer una salida creativa, al tiempo que se mitigan las consecuencias no deseadas sobre la integridad de la información, la atribución, los derechos e incluso el medio ambiente. ◘

REFERENCIAS

BHATT, U., ANDRUS, M., WELLER, A., & XIANG, A. (2020). Machine Learning Explainability for External Stakeholders. *Association for Computing Machinery ArXiv*, (arXiv:2007.05408). http://arxiv.org/abs/2007.05408

BICKERT, M. (2020, enero 6). Enforcing Against Manipulated Media. *Facebook Blog*. https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/

BUOLAMWINI, J. (2016). *Project Overview Algorithmic Justice League*. MIT Media Lab. https://www.media.mit.edu/projects/algorithmic-justice-league/overview/

CHESNEY, R., & CITRON, D. K. (2018). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review, 107*(1753). https://doi.org/10.15779/Z38RV0D15J

COSTANZA-CHOCK, S. (2018). Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice. *Proceedings of the Design Research Society 2018*. https://doi.org/10.21606/drs.2018.679

CRAWFORD, K., & PAGLEN, T. (2019). *Excavating AI: The Politics of Images in Machine Learning Training Sets*. Excavating AI. https://excavating.ai

DIEHM, C., & SINDERS, C. (2020, mayo 14). *"Technically" Responsible: The Essential, Precarious Workforce that Powers A.I.* The New Design Congress Essays. https://newdesigncongress.org/en/pub/trk

DOLHANSKY, B., BITTON, J., PFLAUM, B., LU, J., HOWES, R., WANG, M., & FERRER, C. C. (2020). The DeepFake Detection Challenge (DFDC) Dataset. *Association for Computing Machinery ArXiv*, (arXiv:2006.07397). https://arxiv.org/abs/2006.07397v4

EPSTEIN, Z., LEVINE, S., RAND, D. G., & RAHWAN, I. (2020). Who Gets Credit for AI-Generated Art? *IScience, 23*(9), 101515. https://doi.org/10.1016/j.isci.2020.101515

GEBRU, T., MORGENSTERN, J., VECCHIONE, B., VAUGHAN, J. W., WALLACH, H., DAUMÉ III, H., & CRAWFORD, K. (2018). Datasheets for Datasets. *Association for Computing Machinery ArXiv*, (arXiv:1803.09010). https://arxiv.org/abs/1803.09010v1

GROSZ, B. J., GRANT, D. G., VREDENBURGH, K., BEHRENDS, J., HU, L., SIMMONS, A., & WALDO, J. (2018). Embedded EthiCS: Integrating Ethics Broadly Across Computer Science Education. *Association for Computing Machinery ArXiv*, (arXiv:1808.05686). https://arxiv.org/abs/1808.05686

HAO, K. (2019, junio 6). *Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes*. MIT Technology Review. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

HARA, N. (2020). *Pause Fest* [Imagen generada con IA]. http://www.n-hara.com

LACOSTE, A., LUCCIONI, A., SCHMIDT, V., & DANDRES, T. (2019). Quantifying the Carbon Emissions of Machine Learning. *Association for Computing Machinery ArXiv*, (arXiv:1910.09700). https://arxiv.org/abs/1910.09700

LEIBOWICZ, C. R. (2020). *The Deepfake Detection Challenge: Insights and Recommendations for AI and Media Integrity*. Partnership on AI. https://www.partnershiponai.org/wp-content/uploads/2020/03/671004_Format-Report-for-PDF_031120-1.pdf

LEIBOWICZ, C. R., STRAY, J., & SALTZ, E. (2020, julio 13). Manipulated Media Detection Requires More Than Tools: Community Insights on What's Needed. *The Partnership on AI*. https://www.partnershiponai.org/manipulated-media-detection-requires-more-than-tools-community-insights-on-whats-needed/

LI, Y., & LYU, S. (2019). De-identification Without Losing Faces. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, 2019*, 83–88. https://doi.org/10.1145/3335203.3335719

LOMAS, N. (2020, agosto 17). Deepfake Video App Reface is just Getting Started on Shapeshifting Selfie Culture. *TechCrunch*. https://social.techcrunch.com/2020/08/17/deepfake-video-app-reface-is-just-getting-started-on-shapeshifting-selfie-culture/

LYONS, M. J. (2020). Excavating "Excavating AI": The Elephant in the Gallery. *Association for Computing Machinery ArXiv Preprint*, (arXiv:2009.01215). https://doi.org/10.5281/zenodo.4037538

MITCHELL, M., WU, S., ZALDIVAR, A., BARNES, P., VASSERMAN, L., HUTCHINSON, B., SPITZER, E., RAJI, I. D., & GEBRU, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

MIX. (2020, mayo 7). *This AI Spits Out an Infinite Feed of Fake Furry Portraits*. The Next Web. https://thenextweb.com/news/ai-generated-furry-portraits

MOISEJEVS, I. (2019, julio 14). Will My Machine Learning System Be Attacked? *Towards Data Science*. https://towardsdatascience.com/will-my-machine-learning-be-attacked-6295707625d8

NICOLAOU, E. (2020, agosto 27). Chrissy Teigen Swapped Her Face with John Legend's and We Can't Unsee It. *Oprah Daily*. https://www.oprahdaily.com/entertainment/a33821223/reface-app-how-to-use-deepfake/

PARIS, B., & DONOVAN, J. (2019). *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*. Data & Society. https://datasociety.net/library/deepfakes-and-cheap-fakes/

PATRINI, G. (2019, octubre 7). Mapping the Deepfake Landscape. *Sensity*. https://sensity.ai/mapping-the-deepfake-landscape/

POSTERS. (2019, mayo 29). *Gallery: "Spectre" Launches (Press Release)*. http://billposters.ch/spectre-launch/

RAJI, I. D., & YANG, J. (2019). ABOUT ML: Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles. *Association for Computing Machinery ArXiv Preprint*, (arXiv:1912.06166v1). http://arxiv.org/abs/1912.06166

RAKOVA, B., YANG, J., CRAMER, H., & CHOWDHURY, R. (2020). Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for shifting Organizational Practices. *Proceedings of the ACM on Human-Computer Interaction, CSCW1*. https://doi.org/10.1145/3449081

ROTH, Y., & ACHUTHAN, A. (2020, febrero 4). Building Rules in Public: Our Approach to Synthetic & Manipulated Media. *Twitter Blog*. https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media

ROTHKOPF, J. (2020, julio 1). Deepfake Technology Enters the Documentary World. *The New York Times*. https://www.nytimes.com/2020/07/01/movies/deepfakes-documentary-welcome-to-chechnya.html

SALGADO, E. (2020, agosto 5). *Yaku with Circular Loops* [Imagen generada con IA]. https://www.youtube.com/watch?v=ksQW8Q2WV9c

SALTZ, E., COLEMAN, L., & LEIBOWICZ, C. R. (2020). *Making AI Art Responsibly: A Field Guide* [Zine]. Partnership on AI. https://www.partnershiponai.org/wp-content/uploads/2020/09/Partnership-on-AI-AI-Art-Field-Guide.pdf

SCHULTZ, D. (2019). *Faces2flowers—Artificial Images*. https://artificial-images.com/project/faces-to-flowers-machine-learning-portraits/

SCHULTZ, D. (2020). *Artificial Images*. https://artificial-images.com/

SIMONITE, T. (2018, noviembre 28). How a Teenager's Code Spawned a $432,500 Piece of Art. *Wired*. https://www.wired.com/story/teenagers-code-spawned-dollar-432500-piece-of-art/

TWITTER SAFETY [@TWITTERSAFETY]. (2020, octubre 30). *Our policies are living documents. We're willing to update and adjust them when we encounter new scenarios or receive important...* [Tweet]. Twitter. https://twitter.com/TwitterSafety/status/1322298208236830720