

Inteligencia artificial, vampirismo y la caja negra: aproximación especulativa en el marco del aprendizaje automático (AA)¹

David Serra-Navarro

UdG, Universidad de Girona  

<https://dx.doi.org/10.5209/aris.95515>

Recibido: 2 de abril de 2024 • Aceptado: 14 de junio de 2024

ES Resumen: Las relaciones creativas entre humano y máquina a partir de los modelos de inteligencia artificial (IA) generativa están experimentando un crecimiento inédito entre los usuarios, así como las nuevas modalidades artísticas que encuentran en esta tecnología un medio expresivo para explorar e investigar. Este artículo toma como caso de estudio el proyecto *vAlmpir*, una producción e investigación artística donde la IA se convierte en un medio tecnológico de trabajo, pero que también deviene un espacio de reflexión para abordar nuevas narrativas en un trasfondo poético que especula sobre la propia naturaleza de la IA. Se pretende identificar un marco conceptual que pueda contextualizar *vAlmpir* en su inmersión en los procesos de aprendizaje automático (AA) como motor de creación de nuevas interrelaciones, entre el substrato experiencial de la producción artística y la reinterpretación en clave contemporánea de conceptos como *ready-made* y deconstrucción. Una aproximación posthumanista que permita especular sobre futuros posibles mediante la práctica artística, con y sobre IA, para aportar una visión particular a un complejo escenario en el que la colaboración entre algoritmos y humanos supone una necesidad cultural y antropológica.

Palabras clave: Inteligencia artificial, AI-Art, Vampirismo, Creatividad, Ready-made

ENG Artificial intelligence, vampirism and the black box: speculative approach within the machine learning (ML) framework

Abstract: The creative relationships between human and machine based on generative artificial intelligence (AI) models are experiencing unprecedented growth among users, as well as new artistic modalities that find in this technology an expressive means to explore and investigate. This article takes as a case study the *vAlmpir* project, an artistic production and research where AI becomes a technological medium, but which also becomes a space for reflection to address new narratives in a poetic background that speculates on the nature of AI. The aim is to identify a conceptual framework that can contextualize *vAlmpir* in its immersion in machine learning (ML) processes as an engine for creating new interrelationships, between the experiential substrate of artistic production and the contemporary reinterpretation of concepts such as ready-made and deconstruction. A posthumanist approach that allows us to speculate on possible futures through artistic practice, with and about AI, to provide a particular vision to a complex scenario in which collaboration between algorithms and humans is a cultural and anthropological necessity.

Keywords: Artificial intelligence, AI-Art, Vampirism, Creativity, Ready-made

Sumario: 1. Introducción. 2. Datos, transparencia y control. 3. Vampirismo y deconstrucción. 4. Error. 5. Consideraciones finales. Referencias.

Cómo citar: Serra-Navarro, D. (2024). Inteligencia artificial, vampirismo y la caja negra: aproximación especulativa en el marco del aprendizaje automático (AA). *Arte, Individuo y Sociedad* 36(4), 967-977. <https://dx.doi.org/10.5209/aris.95515>

¹ Esta investigación nace del proyecto de entornos digitales: “vAlmpir: la IA como herramienta vampírica”, propuesta seleccionada y financiada por BCN Producció 23/24, LaCapella, ICUB.

1. Introducción

Con la intención de situar esta investigación basada en la práctica artística, y centrada en una forma alternativa de interpretar la relación creativa entre humano-máquina, desde los modelos entrenados mediante inteligencia artificial (IA) para generar imágenes, a continuación, se describe una breve perifrasi contextual y cronológica para acercarnos al término aprendizaje automático (AA, *machine learning*) que acuñó Arthur Samuel en 1959 como un campo de estudio que otorga a las computadoras la capacidad de aprender sin ser programadas explícitamente. En otras palabras, esta rama de la IA que permite a las máquinas llevar a cabo tareas sin instrucciones explícitas se define como:

(...) una serie de métodos que aprenden² de los datos para crear o construir un modelo que pueda tomar decisiones informadas basadas en lo aprendido. Esta definición implica directamente que el modelo/algoritmo necesita aprender. El procedimiento de aprendizaje hace que el algoritmo sea lo suficientemente confiable como para predecir cualquier propiedad en datos nuevos que no se hayan utilizado para el aprendizaje. Aquí es donde se deben utilizar las metodologías supervisadas (Amigo, 2012, p. 54).

El uso de la IA para crear contenido coherente podemos iniciarlo a finales de los años 50 con la aparición de ELIZA (Weizenbaum, 1976), un programa desarrollado por el MIT basado en la coincidencia de patrones y procesamiento del lenguaje natural (Dehouche y Dehouche, 2023), posiblemente uno de los primeros chatbots³ que conocemos con la capacidad de mantener una conversación escrita con un humano y responder con palabras clave. El programa llevaba el nombre del personaje ficticio Eliza Doolittle de la obra *Pígalión*, que a su vez se inspiró en el legendario escultor de la mitología griega⁴ que esculpió una estatua tan realista que se enamoró de ella e hizo que los dioses la trajeran a la vida (Brean, 2023). Recordemos que la historia de Frankenstein (1818) de Mary Shelley, el "*moderno Prometeo*" (Marijuán, 2018, p.36), también es una referencia⁵ a Pígalión. En 1968 Harold Cohen, pintor expresionista abstracto inglés que hizo del ordenador el principal vehículo de su planteamiento artístico desarrolló AARON, un programa dotado con IA capaz de realizar dibujos a mano alzada (Morbey, 1993). De esta forma Cohen da inicio a un trabajo arte algorítmico marcado por una relación en la que predominaba la colaboración hombre-máquina (Audry y Ippolito, 2019). Según remarca Cohen (2010):

La creatividad... no reside ni únicamente en el programador ni solo en el programa, sino en el diálogo entre programa y programador; un diálogo basado en la relación especial y peculiarmente íntima que había crecido entre nosotros a lo largo de los años (p. 9).

A finales del siglo XX y principios del XXI, el arte generativo se convirtió en una práctica común con ejemplos como los algoritmos evolutivos de Karl Sims⁶, la música generativa⁷ de Brian Eno en 1996, o las infinitas iteraciones de un modelo 3D de una ciudad a través del programa Argenia⁸ de Celestino Soddu. Años más tarde, con el desarrollo de las *Generative Adversarial Nets* (GAN) propuestas por Goodfellow (2014), la producción artística de arte generativo experimenta un cambio de paradigma, así como un aumento de su visibilidad en la sociedad (Ren y Du, 2024). Con el desarrollo de esta tecnología la idea que la IA es simplemente una herramienta empieza a diluirse para acercarnos a una máquina que deviene un ente que colabora con usuarios humanos en tareas creativas (Davis et. al., 2015). Fijémonos como el algoritmo (GAN), responsable de generar *Portrait of Edmond de Belamy*⁹ (2018) del colectivo artístico Obvious aparece como firmante de la propia obra; o como la pieza *Memories of Passers-by I* (2019), del artista Mario Klingemann, "no muestra un producto final curado por humanos, ¡un trabajo autónomo de IA ejecutado en tiempo real; constituyendo una obra infinita, y en cierta forma efímera, ya que los retratos que se visualizan nunca se repiten" (Marinero, 2020, p. 2).

Recientemente, el uso de GAN para la generación de imágenes ha sido superado por modelos de difusión, modelos que funcionan aplicando repetidamente ruido gaussiano en una imagen y luego eliminando el ruido del resultado en la misma cantidad de pasos (Song et al., 2020). En 2022, Rombach et al. lanzaron el modelo Stable Diffusion, plataforma open-source que centrará nuestra atención, mientras que OpenAI lanzaba Dall-E 2 con una arquitectura de modelo similar pero con un conjunto de datos de entrenamiento opaco al público. Actualmente encontramos en la red una larga lista plataformas de IA generativa (Midjourney, StarryAI, NightCafe, o Firefly de Adobe), que utilizando AA nos presentan modelos cuyos espacios de salida

² "Aprender" en este contexto significa identificar patrones complejos entre millones de datos.

³ Un chatbot es un programa informático que utiliza inteligencia artificial y procesamiento del lenguaje natural (NLP) para comprender las preguntas de los usuarios y automatizar las respuestas a dichas preguntas simulando la conversación humana.

⁴ En la leyenda griega, rey y escultor se enamora de una estatua, una mujer que él mismo hizo. A petición suya, Afrodita da vida a la estatua para convertirla en humana, su esposa Galatea. Ovidio, en *Las metamorfosis* (8 d.C.), describió el paso de marfil a vida con la expresión: *corpus erat: salient temptatae pollice venae*, una alusión al latido de las venas que conecta directamente con el elemento simbólico de la sangre.

⁵ La historia de Pígalión ha sido ampliamente representada en películas como *Die Puppe* (Ernst Lubitsch, 1919); *Metrópolis* (Fritz Lang, 1927) o *Los cuentos de Hoffmann* (M. Powell y E. Pressburger, 1951).

⁶ <https://www.karlsims.com/> (Consultado 5 de febrero de 2023).

⁷ <https://proyectoidis.org/generative-music-1/> (Consultado 10 de febrero de 2023).

⁸ http://www.soddu.it/edinburgh/soddu_edinb.htm (Consultado 26 de marzo de 2024).

⁹ En octubre de 2018 la obra generada por un algoritmo de IA fue subastada en Christie's. Véase: Christie's. Is artificial intelligence set to become art's next medium? (12 diciembre de 2018). Disponible en: <https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx> (Consultado 3 de agosto de 2022).

se superponen parcial o totalmente con sus espacios de entrada durante el entrenamiento (Jiang et al., 2023). A través de un conjunto de datos de entrenamiento estas herramientas IA generativas memorizan y pueden llegar a replicar con precisión el estilo de artistas específicos. Concretamente Rombach et al. (Stable Diffusion) utilizaron un subconjunto de LAION-5B¹⁰ (*Large-scale Artificial Intelligence Open Network*) en el entrenamiento de *latent diffusion models* (LDM) logrando resultados de última generación en imágenes (Schuhmann et al., 2022). En este punto empieza la investigación anclada sobre el particular proyecto *vAImpir* (2024), una producción/investigación artística que bascula sobre la IA y en las representaciones creadas por los generadores de imágenes que utilizan un espacio latente creado durante el proceso de entrenamiento de un modelo, y que se vincula con la idea de aprendizaje automático con la que hemos iniciado esta introducción. Los siguientes apartados tienen como objetivo complementar y ampliar los resultados del propio proyecto con un contexto referencial que permita matizar conceptualmente las actuales manifestaciones artísticas vinculadas a la noción *AI-Art*, más allá de entenderse simplemente como una producción mediante IA, y acercar el proceso de la práctica artística en un punto de convergencia entre algoritmos y humanos. Adentrarse e identificar un uso experimental de la IA como un medio que despierte la imaginación, la creación y la crítica.

2. Datos, transparencia y control

El concepto “aprender” en IA no podemos vincularlo con una cognición real propia de los humanos, cuando nos referimos a AA estamos delante un proceso estadístico en el que un algoritmo tiene la capacidad de encontrar patrones y crear sus propios modelos que se relacionan con los datos. Esta idea es clave. Por un lado, las predicciones no se basan en una teoría, sino en los propios datos, y por otro lado los datos determinan la predicción. Si tenemos en cuenta que muchos de los procesos actuales de AA provienen del *big data*, una cantidad masiva de datos y bases de datos complejas, podemos ver como la disciplina de la estadística en este contexto se convierte específicamente en ciencia de datos¹¹. La promesa es que los algoritmos de AA pueden extraer patrones no obvios de grandes conjuntos de datos y que los patrones revelados proporcionarán información (novedosa) sobre los problemas que, en última instancia, resultarán en mejores decisiones (Ker et al. 2020; Kelleher y Tierney, 2018). ¿Pero podemos confiar en estas decisiones? ¿Sabemos qué hay detrás del AA?

En clave ética es responsabilidad humana conocer la razón de cualquier decisión particular que genere la máquina mediante IA. La explicabilidad del funcionamiento de un algoritmo, poder llegar a explicar el por qué se ha emitido una decisión (o en nuestro caso un resultado visual concreto), tendría que ser posible. Por este motivo la supervisión humana en cada una de las fases previas a los resultados es crucial. Definir, seleccionar, interpretar, obtener, testar o agrupar datos son funciones específicas realizadas por humanos (equipos) para dar forma a un algoritmo con una finalidad concreta; en síntesis, el tratamiento de los datos que va desde el análisis inicial hasta la representatividad de los *outputs* conllevará un carácter ético inherente en su validación. Evaluar estas herramientas en eficiencia, ética y sostenibilidad es un requerimiento necesario para garantizar su impacto y las sinergias globales que surgirán en las inevitables interacciones entre sociedad y IA (Vinuesa et al., 2020).

No obstante, pese a esta conciencia ética, debemos señalar que los algoritmos de AA pueden discriminar según clases como raza y género (Buolamwini y Gebru, 2018) y que se asocian al *problema de la caja negra*. Mark Coeckelbergh en *AI ethics* (2020) nos subraya que mientras la IA de los años 80 seguía el paradigma de investigación del *conexionismo* y se basaba en las representaciones simbólicas, es decir, un proceso determinista diseñado por un flujo de decisiones, condicionales y consecuencias que desembocaban en decisiones complejas, la tecnología IA actual incorpora redes neuronales. Esto significa que la máquina realiza interacciones simples (neurona) de forma masiva (red), proceso que permite el llamado AA pero que contrasta con la transparencia e interpretación de los árboles de decisiones que pueden ser revisados y evaluados por seres humanos. De esta manera el AA puede convertirse en una *caja negra* ya que, aunque los programadores conozcan la arquitectura de la red no está claro cómo se alcanza una decisión (Coeckelbergh, 2020). Además, otro factor que añade complejidad a la trazabilidad de la lógica del algoritmo original, es que en la historia de un algoritmo de IA posiblemente se hayan visto involucradas numerosas personas con visiones parciales en relación al primer desarrollo. El control absoluto en todo el proceso, especialmente comprendido en el tramo entre las entradas y salidas de datos, demuestra que su explicabilidad y transparencia puede resultar inalcanzable¹². En contraposición, como nos apunta Han Byung-Chul (2022):

La palabra latina *datum*, que viene de *dare*, significa literalmente, lo *dado*. El dataísmo no imagina otra realidad detrás de lo dado, detrás de los datos, porque es un *totalitarismo sin ideología*. (p. 21) [...] los datos por sí solos no iluminan al mundo. Su esencia es la transparencia (p. 84).

De algún modo, se crea un marco paradójico en el que se visibiliza la tensión entre dos fuerzas aparentemente opuestas: la rigidez de los datos que nutren al algoritmo para su entrenamiento y la opacidad en el proceso de

¹⁰ <https://laion.ai/about/> (Consultado 5 de noviembre de 2023).

¹¹ En 1997 el profesor Jeff Wu, dada la especificidad del tratamiento y gestión de los datos, aparte de los cálculos numéricos, renombró la estadística en ciencia de datos. Véase, Statistics= data science? Recuperado de la URL: <https://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf>

¹² En este sentido, el autor del algoritmo AICAN, Ahmed Elgammal, afirmaba en una entrevista publicada en el portal theconversation.com: “Como científico, creé el algoritmo, pero no tengo control sobre lo que generará la máquina” (2018). Recuperado de la URL: <https://theconversation.com/meet-aican-a-machine-that-operates-as-anautonomous-artist-104381>

AA. En este intento donde el usuario intentará “domesticar” los algoritmos IA a partir de los datos seleccionados con los que alimentará sus AA aparece un diálogo creativo entre humano y máquina, una lucha entre control y descubrimiento. Dentro de esta intersección, zona de acción híbrida, se inscribe la investigación. Un espacio de retroalimentación social que despliega un nuevo imaginario por encima del espectro artístico:

Si bien los artistas han utilizado algoritmos en la creación artística desde la década de 1960, hoy en día la ‘IA cultural’ a escala industrial está integrada en dispositivos y servicios utilizados por miles de millones de personas. En lugar de ser un instrumento de una única imaginación artística, la IA se ha convertido en un mecanismo para influir en la imaginación de miles de millones. Los datos recopilados y agregados sobre los comportamientos culturales de multitudes se utilizan para modelar nuestro ‘yo estético’, prediciendo nuestras decisiones y gustos estéticos futuros y potencialmente guiándonos hacia las elecciones preferidas por la mayoría (Manovich, 2018, p. 2).

La IA de texto a imagen ha experimentado un gran avance con el lanzamiento de Dall-E y su contraparte de código abierto, Stable Diffusion. Un nuevo medio de creación artística que demuestra además un potencial para enseñar la historia del arte, la estética y la técnica (Dehouche y Dehouche, 2023). Precisamente, Stable Diffusion deviene uno de los sistemas de código abierto más extendidos, hecho que supone una democratización de estas herramientas, ofreciendo a cualquier usuario la total libertad para producir y/o contribuir en la comunidad. Esta plataforma se nutre de las imágenes de LAION-5B, que proceden de múltiples dominios, entre los que se encuentran las plataformas DeviantArt, Flickr, Pinterest, Wikimedia Commons y WordPress. Pero lo más atractivo para la comunidad de código abierto es que puede crear y distribuir sus propios modelos a través de entrenamientos AA realizados desde los datos (generalmente imágenes) que cada usuario haya seleccionado y etiquetado. Estas prácticas podemos conectarlas con la noción de *AI-Art* que “evoca tanto las técnicas involucradas en la producción de una obra de arte (IA) como el propósito del campo (Arte)” (Caramiaux y Fdili Alaoui, 2022, p. 477).

Con estas herramientas se exhibe un cambio de paradigma en el que las imágenes ya no se capturan, se generan, “las imágenes generadas por IA no son meras representaciones; son creaciones en sí mismas, y su propósito no es reflejar la realidad, sino desafiarla y transformarla” (Domenech et al., 2023). Si para Susan Sontag fotografiar las personas era ya un acto de violación, una transformación de las personas en objetos que podían ser poseídas simbólicamente, “un asesinato blando.” (2001, p. 31), con los modelos IA, que habilitan también la pornografía sintética entre malabarismos de DeepFake (Fig. 1), la posesión se convierte en un color más de la paleta del artista IA. La idea de que la “fotografía es una ficción que se presenta como verdadera” (Fontcuberta, 1997, p. 25) adquiere una especial relevancia cuando lo trasladamos a este escenario de transformación digital que está sufriendo el mundo. De nuevo las potencialidades y peligros de la tecnología IA representan un desafío ético, ya sea en el proceso de entrenar un algoritmo con los datos correspondientes o por el uso dirigido que puedan hacer gobiernos, grandes corporaciones o organizaciones criminales (Hochreiter, 2020), talmente como el *Vampyroteuthis* en la fábula de Vilém Flusser donde esta figura impone su información a otros, “un vampirismo ‘inverso’ que bombea una sustancia a los demás en lugar de succionarla” (Jue, 2014, p. 91). La IA encarna la singularidad tecnológica que nos plantea “interrogantes sobre el futuro ontológico del hombre y la profundidad epistemológica de una tecnología mediadora en la composición de un posible nuevo orden” (García-Sedano, 2019, p. 656).



Figura 1. Fotograma de la obra *FakeBeauty* (2024). Imagen generada a través del modelo entrenado Juggernaut XL. La representación de cada persona está subordinada a los datos como se ha entrenado el algoritmo. Captura de imagen del autor.

3. Vampirismo y deconstrucción

Con este extenso preámbulo se introduce el laboratorio conceptual de la indagación/producción artística *vAlmpir: la AI como herramienta vampírica*, un punto estratégico que cuenta con múltiples convergencias: la

reflexión sobre el abismo creativo que supone la tecnología IA, el AA como espacio inestable para la estética del error, y la IA como una posible interpretación de un expansionismo vampírico contemporáneo.

VAImpir (2024), un proyecto artístico de entornos digitales desarrollado en el Centro de Arte La Capella¹³, toma como referente y pretexto la obra *Vampir-Cuadecuc* (1970), una investigación del lenguaje cinematográfico realizada por Pere Portabella durante el período¹⁴ del tardofranquismo. Tal como apuntan Deltell y García: “*Vampir-Cuadecuc* no es un *making of* sino una obra completamente distinta. En ella no se explica cómo se rodó una película sino que se cuenta de otra manera esa obra filmada” (2015, p. 71), un juego de reescritura y deconstrucción que se elaboró sobre la película *Count Dracula* de Jesús Franco. “Cuadecuc”, en catalán, hacía referencia a los restos de película que sobraban de un rodaje, una denominación utilizada por el poeta Joan Brossa, colaborador habitual¹⁵ de Portabella. Precisamente, la estrategia de Portabella y del Brossa guionista se basó en crear una meta-historia a partir de otro material cinematográfico ya existente y exhibido en “un equilibrio entre lo hermético y lo anárquico [...] en una narrativa quebradiza” (Alcoz, 2013, p. 24).

En cierta manera esta pieza icónica de Portabella ejerce una vampirización sobre el film original y abre un proceso creativo que aparentemente parasita la imagen original. Pese a que la “vampirización” puede caracterizarse por “una marcada infertilidad que se cierra sobre sí misma donde lo único que ofrece es la negatividad del mecanismo entrópico que la define.” (Espinosa, 2023, p. 20), aquí el vampiro nos devela otra realidad, y se refuerza como un símbolo del “otro” (Onishi, 2013). La primera analogía superficial que se establece entre *Cuadecuc* y la IA es una correlación metafórica entre la sangre y los datos, pero de forma más explícita, si tenemos en cuenta que la generación de imágenes IA puede darse desde la semilla de otra imagen (fotograma), encontramos que los mecanismos de exploración de Portabella para encontrar nuevas películas dentro de otras películas guardan una cierta similitud con la traducción y síntesis IA de *image-to-image*¹⁶, ya que no habrá *output* sin antes succionar la “imagen guía”.

Fuera de la poética que pueda suscitar esta interpretación, lo destacable de esta producción es la colaboración entre humano y máquina, en un tono posthumanista en el que la dualidad no significa competir, sino aprender mutuamente, a diferencia de una visión transhumanista defendida por autores como Nick Bostrom (2002), en la que se considera “que es un deber moral trabajar en la línea de intentar trascender la especie humana con todas sus imperfecciones para crear esos seres perfectos que compondrían la singularidad” (Cortina, 2019, p. 384). Es en la convivencia y en el juego dialéctico donde se desarrolla una acción que no toma al humano como centralidad. En esta perspectiva, el planteamiento que “la creatividad no es un proceso mágico que debe ser admirado ya que la IA puede exhibir comportamientos creativos parecidos a los humanos” (Boden y Edmonds, 2019) abre un debate sobre como se puede establecer un diálogo creativo, ¿Cuál es el espacio para unir autonomías creativas entre IA e inteligencia humana? Tal como nos señalan Marín-Viadel et al. (2024), en relación al uso de los entrenamientos IA en su vertiente educativa, lo decisivo es incorporar al usuario dentro de un proceso colaborativo y convertirlo en un usuario activo que le permita descubrir y aprender las posibilidades del medio.

Como hemos descrito anteriormente, la producción IA está subordinada a los algoritmos y principalmente a los datos con los que se han entrenado los modelos, y aunque estos sistemas son capaces de emular comportamientos creativos tenemos que señalar, como apunta Fabio Morandín-Ahuerma que “no tienen experiencia subjetiva de su propio comportamiento, y no poseen el tipo de agencia e intencionalidad que se asocia con el tipo de decisiones humanas” (2023, p. 204), no obstante, el mismo autor recuerda que en la filosofía de Martin Heidegger, “la tecnología y la ingeniería pueden cambiar la relación del hombre consigo mismo y con el mundo” (2023, p. 194). Es esta premisa la que posiblemente defina un encuentro relacional, no tan alejada del semionauta de los años 90 que describía Bourriaud, y que ahora en un marco mutante de postdigitalidad se transforman en *AI-Art* (un primer impacto de las tecnologías de inteligencia artificial sobre las prácticas creativas). La idea de que una superinteligencia no humana, superior a la humana nos puede acabar vampirizando, o, dicho de otra manera, gobernando (Tegmark, 2017; Harari, 2015) es un relato que encierra un doble sentido, como un espejo¹⁷ que nos refleja nuestra fragilidad y a la vez se enlaza con al mito de Prometeo.

Siguiendo la trama hematófaga, en la que cada mordedura provoca una réplica del estado vampírico, los modelos IA que circulan entre los grandes repositorios como Hugging Face devienen frecuentemente la base otros modelos (*merge models*) que incorporan entre sus datos restos de su primer entrenamiento; una cadena de producción acumulativa que se distribuye a través de la comunidad *open-source*, y en la que los posibles sesgos provenientes de los datos como se han entrenado los modelos (Chui et al., 2018) continuará presente. ¿Pero es la IA una herramienta vampírica?

La ya cotidiana imitación de estilo invasiva por parte de las plataformas generadoras de imágenes supone una amenaza que plana sobre los derechos de autor, otro punto relevante a tratar y que concuerda con

¹³ La Capella es el espacio que el Institut de Cultura de Barcelona dedica a la creación emergente. Barcelona Producció, que se inició en 2006, es ahora el proyecto central de la programación de La Capella. *VAImpir* es uno de los proyectos seleccionados en la convocatoria pública 23/24. <https://www.lacapella.barcelona/es/presentacion-de-la-publicacion-vaimpir-lai-com-eina-vampierica-vaimpiro-la-ai-como-herramienta>

¹⁴ En este mismo período de cine experimental cabe destacar títulos como *Contactos* (1970) de Paulino Viota, o *La celosía* (1972) de Isidoro Valcárcel Medina.

¹⁵ El artista y compositor Carles Santos fue el encargado de crear la banda sonora de *Vampir-Cuadecuc*.

¹⁶ Véase estudio de Tumanyan et al. (2023) en <https://doi.org/10.48550/arXiv.2211.12572>

¹⁷ Según las leyendas los vampiros no se reflejan en el espejo, una ausencia de realidad que identificamos como experiencia subjetiva.

esta idea parasitaria, no obstante, vislumbrando en el horizonte la lógica duchampiana de la apropiación (Fig. 2), debemos recordar que “los algoritmos (o, en términos kantianos: ‘imperativos de habilidad’ producidos por medios tecnológicos) ni se inventan ni comienzan por sí mismos; ni se fijan sus propios objetivos fruto de la reflexión. [...] Son un producto humano” (Nagl, 2022, p. 70), y reafirmamos la presencia humana activa en todo el proceso, ya que los generadores de imágenes “no son artistas: requieren objetivos y propósitos humanos para dirigir su producción o reproducción, y son estos objetivos y propósitos humanos los que dan forma a las direcciones en las que se producen sus productos” (Jiang et al, 2023). En la misma dirección, Aaron Hertzmann nos reafirma que el arte es una actividad social y que el software sólo sigue mecánicamente las instrucciones que le damos, “llamar artista a un sistema de software es irresponsable, porque induce a error: podría hacer pensar a la gente que el software tiene inteligencia, autonomía y emociones humanas” (2020, p. 48). Según matiza Juan Martín Prada (2024), en la aproximación de esta cierta actualización tecnológica de las prácticas apropiacionistas del posmodernismo de los años ochenta:

Se trata de prácticas artísticas que, lejos de conformarse con lo ofrecido por ese apropiacionismo tecnológico basado en la efectista combinación y transformación derivativa propias de la IA, optan por referirse críticamente a los discursos mismos de la IA, muchas veces apropiándose de ellos, haciendo de este cuestionamiento ético su centro temático (p. 12).



Figura 2. Extracto de 12 salidas-imagen aplicando modelo entrenado sobre la referencia de *Las Meninas* (Velázquez, 1656) tomada como pantalla representacional. Elaboración del autor + IA (2024, 2 de febrero)

4. Error

Las producciones y reflexiones de Luigi Russolo y John Cage en la música, de Marcel Duchamp, Nam June Paik y Wolf Vostell, entre otros, no hacen más que continuar la propuesta de inclusión del error en el ámbito artístico y confrontar los usos habituales de los aparatos tecnológicos que nos rodean (Schianci, 2012, p. 31).

La búsqueda de lo impredecible¹⁸ representa una forma de humanización de la tecnología y de expresión artística a través del accidente que se nos aparece recurrentemente desde las vanguardias artísticas hasta las actuales herramientas digitales. Un tipo de arte que se articula por procesos imprevistos, procesos en los que el error resulta inmanente a la máquina (Barker, 2011, p. 49) y que podemos identificar en un primer momento con la noción de *glitch*¹⁹, término que ha ido adoptando diferentes connotaciones en la evolución de la tecnología, y que consecuentemente, como nos señala Emilio Vavarella (2015), ha incidido sobre aquellas reflexiones ligadas al poder tecnológico contemporáneo:

Existen numerosas prácticas que hacen uso del *glitch* de forma más o menos sistemática, no siempre con fines artísticos y sin caer necesariamente en la categoría de *glitch-art: new media art, media activism, digital art, postmedia art, net art, e-poetry, y counter-gaming* (p. 11).

Recapitulando, la significación de la “materia sucia” (Parikka, 2012) cobra una especial importancia en todo el recorrido expuesto, ya que si bien la *el problema de la caja negra* en los AA suponen un reto ético, en el escenario de esta modalidad “AI-glitch” se dan las condiciones idóneas para una metafísica de nuevas relaciones entre humano y máquina, relación simbiótica dentro de las experiencias estéticas. En la investigación *vAlmpir*, una forma experiencial de relacionarnos con la tecnología, el error viene dado por defecto en cada *output*, por lo tanto, el espacio subjetivo transcurre por moldear la sorpresa en un diálogo de significaciones. Otro aspecto concreto de *vAlmpir* es que el entrenamiento de los modelos está basado en sutilezas subjetivas, en *assets* amorfos (Fig. 3 y 4) que excita la aparición del *glitch*. Lo destacable no es la relación etiqueta-forma, sino el entrenamiento de “ausencias”, de imagen pixelada, de inconcreción contaminadora de datos. Béla Lugosi (Fig. 5) se vuelve omnipresente cada vez que el algoritmo identifica un patrón de “face”, pero el celuloide coloreado sin la calidad pornográfica del 4K busca filtrarse en el ataúd (*caja negra*, Fig.6) para contaminar la hemoglobina (los datos), igual que la plastilina residual que no quiere pegarse a las etiquetas de “cuerpo”. Una metodología que valora los llamados ataques antagónicos, aquellos que “explotan puntos ciegos y regiones débiles en el modelo estadístico de una red neuronal, generalmente para engañar a un clasificador y hacerlo percibir algo que no está allí” (Pasquinelli y Joler, 2021, p. 12). Se trata de una mordedura recíproca para acercarse al medio, como una forma especulativa de sumergirse en los automatismos para generar expresión.

La IA de forma autónoma no produce significados, se asemeja a una fantasmagoría libre de lecturas históricas, pero es en la colisión entre humano y el entorno tecnológico donde cristaliza el espectro digital del *objet trouvé*, un significado que como apunta Robert Pepperell “solo ocurre como parte de un contexto más amplio que, en lo que respecta a un organismo, es el proceso continuo de existencia llevado a cabo en un entorno que es a la vez hostil y de apoyo” (2003, p. 140).



Figura 3. Secuencia de 6 fotogramas de la película *Grave of the Vampire* (John Hayes, 1972) en los que se ha aplicado un modelo entrenado (sin etiquetas) a partir restos de plastilina desechados en una producción audiovisual de *stop motion*. <https://vimeo.com/922459650/718e079d96?share=copy>

¹⁸ Para Hertzmann (2020), investigador de la corporación Adobe, el software IA da paso a una evolución y transformación de nuevas formas de arte, emocionantes e impredecibles.

¹⁹ Literalmente *glitch*, hace referencia a “un pico o cambio de voltaje en un circuito eléctrico”. Véase: Vespe. J. (2019, octubre). It Was John Glenn Who Popularized the Word “Glitch”. Recuperado de la URL: <https://www.smithsonianmag.com/air-space-magazine/just-right-word-180973113/>



Figura 4. Imágenes generadas a partir de modelo entrenado con las sobras de aluminio de los bocadillos. Elaboración del autor + IA (2023, 10 de octubre)



Figura 5. Imagen generada a partir modelo entrenado *Béla*. La cara del actor austro-húngaro Béla Lugosi que encarnó *Drácula* queda en un plano superficial, ya que el modelo se basa en el detalle del pixel coloreado que se aplicó sobre el celuloide original de Tod Browning (1931). Elaboración del autor + IA a través del modelo XL de descarga libre en <https://civitai.com/models/371773/bela> (2024, 29 de marzo)



Figura 6. Publicación/Artefacto del proyecto *vAlmpir* que alude a la “caja negra” y a la metáfora del espejo. El contenido de esta publicación ha nacido mediante la técnica del método socrático con chatGPT 3.5. <https://vaimpir.kennethrusso.net/> (véase apartado +researchcatalogue)

5. Consideraciones finales

Al igual que la fotografía revolucionó la forma de relacionarnos con las imágenes, liberando la pintura de su función mimética, abriendo nuevas visiones oníricas, documentando la realidad, o dando paso al cine, la IA se constituye como una herramienta altamente compleja que puede entenderse como un medio de expresión artística. La IA se desarrolla en la evolución de las prácticas basadas en el apropiacionismo, siguiendo la estela de *Vampir-Cuadecuc*, en una producción de nuevas auras que se sitúan en clave contemporánea sobre el marco relacional de Nicolas Bourriaud (2009):

(...) abolir la distinción tradicional entre producción y consumo, creación y copia, *ready-made* y obra original. La materia que manipulan ya no es materia prima. Para ellos no se trata ya de elaborar una forma a partir de un material en bruto, sino de trabajar con objetos que ya están circulando en el mercado cultural, es decir, ya firmados por otros (p.26).

vAlmpir es un trabajo creativo que responde a la idea de “arte acerca de la IA” (Martín, 2024, p. 11) y que es tangente a las prácticas emergentes de *AI-Art*, en el sentido de la IA como medio, pero también da paso a diferentes reflexiones críticas sobre la naturaleza humana y el tipo de relación que ésta establece con la máquina, una alternativa para abordar los imaginarios colectivos que pueden parecer agotados desde la “deconstrucción y resignificación conceptual de la imagen” (Ocaña, 2023, p. 2). Los valores poéticos y técnicos en los procesos de *AI-Art* ofrecen un profundo marco de investigación de corte estético, no obstante, abordar esta tecnología significa afrontar una serie de preguntas que abarcan desde el ámbito cultural, a los aspectos antropológicos, a su vertiente sociopolítica, y especialmente su entidad ética. Concretamente, al tratar el AA como foco de atención en un proceso especulativo en el que confrontar constructivamente, datos y automatismos con experiencia y subjetividades, se han obtenido en una primera fase una serie de resultados audiovisuales que carecen de significado sin la comprensión previa del proceso. La recepción final de la propuesta *vAlmpir* pasa por desenredar el conocimiento que permite generar imágenes, “movilizar las herramientas de la imaginación, la visualización, narrativa, metáfora, parábola e ironía” (Zylinska, 2020, pp. 31), y permite proyectar los límites de una tecnología que parece integrarnos en su flujo de datos. Entre los posicionamientos posthumanistas que fusionan hombre y máquina, y los transhumanistas, donde la tecnología quiere cambiar nuestra forma de ser para hacernos mejores en mente y cuerpo, la idea de la singularidad tecnológica orbita sobre un *AI-Art* en el que planea un espacio crítico para repensar la experiencia en un mundo hibridado de ciencia y tecnología.

Como estrategia creativa, *vAlmpir* invita a los usuarios a descargarse los modelos entrenados para que experimenten en primera persona, y explora los procesos de AA tomando como referencia el cine experimental de los años 70, conectando con un juego de deconstrucción que avanza a través de autoría e identidad; pero en este caso en la integración de los datos como materia, y cuestionando el peso específico que la opacidad (*la caja negra*) de los procesos AA ejercen sobre la construcción de nuevas narrativas, de futuros posibles. La IA, disruptiva en todos los campos, demuestra un enorme potencial dentro de la producción artística, y representa un espacio de experimentación en el que desde el pensamiento y el lenguaje tendremos que aprender a convivir; un volumen de información descomunal y algoritmos silenciosos que no deben eclipsar la interpretación y subjetividad humana.

Referencias

Alcoz, A. (2013). Reverberaciones en celuloide: Una aproximación al sonido en el cine experimental español (1955-1979). *Arte y Políticas de Identidad*, 8, 17-31.

- Amigo, J. M. (2021). Data mining, machine learning, deep learning, chemometrics: Definitions, common points and trends (Spoiler Alert: VALIDATE your models!). *Brazilian Journal of Analytical Chemistry*, 8(32), 45-61. doi: 10.30744/brjac.2179-3425.AR-38-2021
- Audry, S. y Ippolito, J. (2019). Can Artificial Intelligence Make Art without Artists? Ask the Viewer. *Arts*, 8(1), 35. doi: <https://doi.org/10.3390/arts8010035>
- Barker, T. (2011). Aesthetics of the error: Media art, the machine, the unforeseen, and the errant. *Error: Glitch, noise, and jam in new media cultures*, 42-58.
- Boden, M. A. y Edmonds, E. A. (2019). *From fingers to digits: an artificial aesthetic*. MIT Press.
- Bostrom, N. (2002). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge.
- Bourriaud, N. (2009). *Postproducción*. Adriana Hidalgo.
- Brean, A. (2023). From ELIZA to ChatGPT. *Tidsskr Nor Laegeforen*, 143(6). doi: 10.4045/tidsskr.23.0279
- Buolamwini, J. y Gebru, T. (2018). Gender shades: intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in Proceedings of Machine Learning Research 81*, 77-91.
- Caramiaux, B. y Fdili Alaoui, S. (2022). "Explorers of Unknown Planets": Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-24. doi: <https://doi.org/10.1145/3555578>
- Chui, M., Harrysson, M., Manika, J., Roberts, R., Chung, R., Nel, P. y Van Heteren, A. (2018). McKinsey Global Institute. Notes from the ai frontier applying ai for social good. Recuperado 20 de marzo de 2024 de la URL: www.mckinsey.com/mgi
- Coeckelbergh, M. (2020). *AI ethics*. MIT Press. doi: <https://doi.org/10.7551/mitpress/12549.001.0001>
- Cohen, H. (2010). Driving the creative machine. *Orcas Center, Crossroads Lecture Series, Orcas, WA*, 9.
- Cortina, A. (2019). Ética de la inteligencia artificial. *Anales de la Real Academia de Ciencias Morales y Políticas* (pp. 379-394). Ministerio de Justicia.
- Davis, N., Hsiao, C.P., Popova, Y. y Magerko, B. (2015). An Enactive Model of Creativity for Computational Collaboration and Co-Creation. N. Zagalo & P. Branco (eds.), *Creativity in the Digital Age*. Springer.
- Dehouche, N. y Dehouche, K. (2023). What's in a text-to-image prompt? The potential of stable diffusion in visual arts education. *Heliyon*, 9. doi: <https://doi.org/10.1016/j.heliyon.2023.e16757>
- Deltell Escolar L. y García Crego J. (2015). Límites del lenguaje cinematográfico en el cine experimental durante el tardofranquismo. *Escritura e Imagen*, 11, 65-76. https://doi.org/10.5209/rev_ESIM.2015.v11.50966
- Domenech, L., Torres, N., y López D. (2023). Ατλαντίς: Arqueología Artificial. Atlantis, Atlántida, Numenor, Akalabêth, Tihuanaco... *Umática. Revista sobre Creación y Análisis de la Imagen*, 6. doi: <https://doi.org/10.24310/Umatica.2023.v5i6.17890>
- Elgammal A. (17 de octubre de 2018). *Meet AICAN, a machine that operates as an autonomous artist*. Recuperado el 4 de marzo de 2024 de <https://theconversation.com/meet-aican-a-machine-that-operates-as-an-autonomous-artist-104381>
- Espinosa, M. (2023). Un deleite nocivo: arte e improductividad en la figura vampírica. *La otredad textualizada: Estudios desde la crítica literaria y la estética*, Universidad Nacional de San Luis, 15-33.
- Fontcuberta, J. (1997). *El beso de Judas. Fotografía y verdad*. Gustavo Gili.
- García-Sedano, M. (2019). La rebelión de las máquinas. Arte, singularidad tecnológica y ciencia ficción. *Arte, individuo y sociedad*, 31(3). doi: <https://doi.org/10.5209/aris.61598>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. y Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27. arXiv e-prints, page arXiv:1406.2661
- Grba, D. (2022). Deep Else: A Critical Framework for AI Art. *Digital*, 2(1), 1-32. doi: <https://doi.org/10.3390/digital2010001>
- Han, B. C. (2022). *Infocracia: La digitalización y la crisis de la democracia*. Taurus.
- Harari, Y. N. (2015). *Homo deus a brief history of tomorrow*. Hervill Secker.
- Hertzmann, A. (2020). Computers do not make art, people do. *Communications of the ACM*, 63(5), 45-48. doi: <https://doi.org/10.1145/3347092>
- Hochreiter, S. (2020). The Algorithm Can Learn Anything—Good Things as Well as Bad Things. *Digital Transformation and Ethics*, ed. by M. Hengstschläger. Ecowin Verlag, 362-375.
- Jiang, H.H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, F. y Gebru, T. (2023). AI Art and its Impact on Artists. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES '23), August 08--10, 2023, Montréal, QC, Canada*. ACM, New York, NY, USA 12 Pages. doi: <https://doi.org/10.1145/3600211.3604681>
- Jue, M. (2014). Vampire squid media. *Grey Room*, 57, 82-105. doi: https://doi.org/10.1162/GREY_a_00155
- Kelleher, J. D. y Tierney, B. (2018). *Data science*. MIT press.
- Kerr, A., Barry, M., y Kelleher, J. D. (2020). Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance. *Big Data & Society*, 7(1). doi: <https://doi.org/10.1177/2053951720915939>
- Manovich, L. (2018). *AI Aesthetics*. Moscow: Strelka Press. [2 de enero de 2024: http://manovich.net/content/04-projects/165-ai-aesthetics/manovich.ai_aesthetics_2018.pdf].
- Marijuán, A. A. (2018). La rebelión de Galatea: autómatas, ciborgs y otras construcciones femeninas subversivas del siglo XXI. *Umática: revista sobre creación y análisis de la imagen*, 1, 33-56. doi: <https://doi.org/10.24310/Umatica.2018.v0i1.4743>

- Marinero, A. E. (2020). Art and artificial intelligence, a window into the future of the evolution of contemporary society. *EAI Endorsed Transactions on Creative Technologies*, 7(22). doi: <http://dx.doi.org/10.4108/eai.13-7-2018.163834>
- Marín-Viadet, R., Campos, R. & Roldán, J. (2024). Inteligencia Artificial, Goya y el aprendizaje del dibujo. Primeros pasos. *Arte, Individuo y Sociedad* 36(2), 329-343. <https://dx.doi.org/10.5209/aris.92891>
- Martín Prada, J. (2024). La creación artística visual frente a los retos de la inteligencia artificial. Automatización creativa y cuestionamientos éticos. *Las fronteras de la historia del arte y los estudios visuales. Reflexiones en torno a su objeto de estudio*, editado por Gorka López de Munain. Monográfico temático, *Eikón Imago*, 13, e90081. <https://dx.doi.org/10.5209/eiko.90081>
- Morandín-Ahuerma, F. (2023). Conciencia e inteligencia artificial Heidegger, Searle y Bostron. *Stoa*, 14(28), 189-209. doi: <https://doi.org/10.25009/st.2023.28.2765>
- Morbey, M. L. (1993). AARON: Portrait of the Young Machine as a Male Artist. *RACAR: Revue d'art canadienne*, 20(1), 130-139. doi: <https://doi.org/10.7202/1072764ar>
- Nagl, L. (2022). Digital Technology: Reflections on the Difference between Instrumental Rationality and Practical Reason. *Kantian Journal*, 41(1). doi: 10.5922/0207-6918-2022-1-3
- Ocaña, M. A. B. (2023). AICA. *El Ornitorrinco Tachado. Revista de artes visuales*, 17, 1-20. doi: <https://doi.org/10.36677/eot.v0i17.21163>
- Onishi, B. H. (2013). Vampires, Technology, and Racism: The Vampiric Image in Twilight and Let Me In. *Race, Philosophy, and Film*, 197-210. Routledge.
- Pasquinelli, M. y Joler, V. (2021). El Nooscopio de manifiesto: La inteligencia artificial como instrumento de extractivismo del conocimiento. *LaFuga*, 25.
- Parikka, J. (2012). New Materialism as Media Theory: Medianatures and Dirty Matter. *Communication and Critical/Cultural Studies*, 9(1), 95-100. doi: <http://dx.doi.org/10.1080/14791420.2011.626252>
- Pepperell, R. (2003). *The Posthuman Condition: Consciousness Beyond the Brain*. Cromwell Press.
- Ren, L. y Du, M. (2024). From Canvas to Code: the Evolution of Generative Art in the AI Era. *ADDT 2023: Proceedings of the 2nd International Conference on Art Design and Digital Technology, ADDT 2023, September 15-17, 2023, Xi'an, China* (p. 479). European Alliance for Innovation. doi: <https://doi.org/10.4108/eai.15-9-2023.2340877>
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P. y Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684-10695. doi: <https://doi.org/10.48550/ARXIV.2112.10752>
- Samuel, A. L. (1959). Machine learning. *The Technology Review*, 62(1), 42-45.
- Schianchi, A. (2012). El error en los aparatos como posibilidad estética. En J. Allen (Coord.), *Artnodes*, 2, 30-33.
- Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk, R. y Jitsev, J. (2022). LAION-5B: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35, 25278-25294. doi: <https://doi.org/10.48550/ARXIV.2210.08402>
- Sims, K. (1992). Choreographed image flow. *J. Vis. Comput. Animat*, 3, 31-44. doi: <https://doi.org/10.1002/vis.4340030106>
- Song, J., Meng, C. y Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*. doi: <https://doi.org/10.48550/arXiv.2010.02502>
- Sontag, S. (2001). *Sobre la fotografía*. Alfaguara.
- Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Allen Lane/Penguin Books.
- Tumanyan, N., Geyer, M., Bagon, S. y Dekel, T. (2023). Plug-and-play diffusion features for text-driven image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1921-1930). doi: <https://doi.org/10.48550/arXiv.2211.12572>
- Vavarella, E. (2015). Art, error, and the interstices of power. *Journal of Science and Technology of the Arts*, 7(2), 7-17. doi: <https://doi.org/10.7559/citarj.v7i2.135>
- Vinuesa, R., Azizpour, H., Leite, I. et al. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat Commun*, 11, 233. <https://doi.org/10.1038/s41467-019-14108-y>
- Weizenbaum, J. (1976). *Computer Power and Human Reason: from Judgment to Calculation*. Freeman and Company.
- Zylinska, J. (2020). *AI Art: Machine Visions and Warped Dreams*. Open Humanities Press.