

# **UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ**

**Colegio de Ciencias e Ingenierías**

**“Art challenges technology, technology inspires art”: Una discusión sobre las problemáticas que rodean a la inteligencia artificial desde el desarrollo de una obra artística implementada a base de redes generativas antagónicas (GANs)**

**Pedro Antonio Guerrero Negrete**

**Ingeniería en ciencias de la computación**

Trabajo de fin de carrera presentado como requisito  
para la obtención del título de  
Ingeniero en ciencias de la computación

Quito, 20 de diciembre de 2022

# **UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ**

**Colegio de Ciencias e Ingenierías**

## **HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA**

**“Art challenges technology, technology inspires art”: Una discusión sobre las problemáticas que rodean a la inteligencia artificial desde el desarrollo de una obra artística implementada a base de redes generativas antagónicas (GANs)**

**Pedro Antonio Guerrero Negrete**

**Nombre del profesor, Título académico**

**Felipe Grijalva, Ph.D.**

Quito, 20 de diciembre de 2022

## © DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Pedro Antonio Guerrero Negrete

Código: 00200204

Cédula de identidad: 1722379300

Lugar y fecha: Quito, 20 de diciembre de 2022

## **ACLARACIÓN PARA PUBLICACIÓN**

**Nota:** El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

## **UNPUBLISHED DOCUMENT**

**Note:** The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

## RESUMEN

Recientes avances en el campo de la inteligencia artificial han vuelto a poner preguntas sobre autonomía y consciencia en la boca del público. Esta discusión -pobre y malintencionada- coincide irónicamente con el punto de inflexión histórico en el que la sociedad se encuentra respecto a su dependencia tecnológica; haciendo del enriquecimiento de la discusión un deber.

Con esta premisa, se indaga en la arquitectura de una red neuronal artificial compleja o “inteligente” (específicamente una VQ-GAN) con el fin de explicar su funcionamiento simple y transparentemente. Una vez explicada, se usa esta misma arquitectura -a través del código abierto de PyTTI-Tools- para realizar una obra de arte generativo, y valerse del dichoso paralelismo entre esta rama del arte y la inteligencia artificial para su investigación.

Por último -y en base al análisis de los prototipos realizados y la experiencia adquirida en el proceso- la discusión y las conclusiones se centran en el inmenso potencial del arte generativo para fundar una nueva etapa artística, en desmentir el discurso de autonomía sensacionalista y en enfatizar la inminente necesidad de democratización del conocimiento sobre procesos inteligentes.

**Palabras clave:** Inteligencia artificial, autonomía, GAN, VQ-GAN, arte generativo, explainable AI, XAI.

## ABSTRACT

Recent progress in the field of artificial intelligence have brought questions about autonomy and conscience back on the table. The discussion regarding these questions -poor and misguided- coincides with an historic inflex point regarding society's dependence towards technology, making it an obligation to enrich it.

With this premise, the architecture of a complex or "intelligent" artificial neural network (specifically a VQ-GAN) is investigated in order to explain its operation simply and transparently. Once explained, this same architecture is used -through the PyTTI-Tools open-source code- to create a generative artwork, making use of the blissful parallelism between this branch of art and artificial intelligence for its investigation.

Finally -and based on the analysis of the prototypes made and the experience gained in the process- the discussion and conclusions focus on the immense potential of generative art to found a new artistic stage, on refuting the sensationalist discourse regarding autonomy and emphasizing the imminent need for democratization of knowledge regarding intelligent processes.

**Key words:** artificial intelligence, autonomy, GAN, VQ-GAN, generative art, explainable AI, XAI.

**TABLA DE CONTENIDO**

Introducción.....	10
Desarrollo de una obra artística .....	27
Conclusiones.....	40
Referencias bibliográficas .....	42
Anexo A: Portrait of edmond belamy .....	46
Anexo B: Documentación PyTTI-Tools y código usado .....	47
Anexo C: Propotipos y Obra final .....	48
Anexo D: artstation art by Azat Nurgaleev .....	49
Anexo E: Link descarga modelo adabins .....	50
Anexo F: Obras de Steve Reich, Erwin Driessens y Maria Verstappen, Jackson Pollock .....	51

**ÍNDICE DE ECUACIONES Y TABLAS**

Ecuación 1: Ecuación Minimax loss.....	17
Ecuación 2: Loss function del Crítico.....	18
Ecuación 3: Loss function del Generador.....	19
Tabla 1. Otros parámetros configurados para la realización del prototipo 1.....	30

## ÍNDICE DE FIGURAS

Figura 1. Modelos discriminativo y generativo de números escritos a mano.....	14
Figura 2. Diagrama de flujo de una GAN.....	16
Figura 3. Diferencia al variar la dimensionalidad del espacio latente (2x16 a la izquierda, 20x60 a la derecha) en reconstrucción imágenes con AutoEncoders.....	22
Figura 4. Un espacio latente arbitrariamente distribuido (izquierda) vs uno normalmente distribuido (derecha).....	23
Figura 5. Arquitectura de la VQ-GAN.....	26
Figura 6. Cuatro imágenes que comprenden el prototipo 1 tomadas a saltos de 84 unidades. Abajo de cada una su id secuencial. Nombre oficial: Anima.....	33
Figura 7. Cuatro imágenes que comprenden el prototipo 2 tomadas a saltos de 20 unidades. Abajo de cada una su id secuencial. Nombre oficial: Canicas sobre el mantel.....	34

## INTRODUCCIÓN

En la década de los setenta, un grupo liderado por Edwin Catmull y Alvy Ray Smith se empeñó en la extravagante idea de hacer una película animada totalmente por computadora [1]. El plan fue el de juntar a científicos computacionales y a artistas de la animación en un mismo cuarto con el fin de que ambos conjuntos se retasen mutuamente: los científicos se rompían la cabeza creando tecnologías que digitalicen el trabajo de los artistas, y los artistas devolvían el favor empujando estas tecnologías al límite y explotando sus capacidades. ¿Su mantra?: “Art challenges technology, technology inspires art” [1]. El resultado no fue únicamente la fundación del grupo que en 1986 pasaría a llamarse Pixar, sino también el nacimiento de su propia forma de arte: la animación digital [1].

Medio siglo después, una nueva tendencia digital derivada de la ciencia computacional (específicamente de la IA o inteligencia artificial) abre de nuevo la posibilidad de interrelacionar al artista con el científico: el arte generativo. Además de guardar el potencial para explorar una nueva etapa artística; el arte generativo presenta una conveniente oportunidad para indagar en las preguntas sobre autonomía que rodean a la IA, siendo el arte un valor históricamente considerado como inherente a la condición humana.

Así -aprovechándose de estas características y considerando los fundamentos del diseño de ingeniería descritos por Yousef Haik [2]- la siguiente investigación pretende, en primer lugar, introducir las problemáticas derivadas del arte generativo en el contexto artístico y la relación que estas mantienen con las aquellas derivadas de la IA en el contexto general.

Adicionalmente, estudiar la arquitectura y los procesos que dan vida a las obras y describirlos para -en la segunda parte- realizar una obra artística y detallar el procedimiento. Finalmente, una vez explorado tanto el universo teórico como el práctico, se llegarán a conclusiones derivadas de las preguntas iniciales.

## Estudio preliminar

### ¿Qué es el arte generativo?

Según el Tate Museum, el arte generativo se hace a través de un sistema predeterminado que usualmente incluye el elemento de aleatoriedad [3]. Sobre esta definición, Christiane Paul añade que este es el producto de un sistema aparentemente autónomo, lo que lo dota de un rasgo muy particular, diferente a cualquier otro en la historia del arte [4].

Quizás su ejemplo más notable es el de *Portrait of Edmond Bellamy*, una obra subastada en 2018 por \$432,500 (Anexo A) [5]. En ella se muestra el retrato -inconcluso y aparentemente antiguo- de un hombre y, debajo de este, una fórmula matemática con el nombre “min max”. Peculiar de la obra es que esta fue concebida a través de una red generativa antagónica (GAN por sus siglas en inglés) entrenada a base de pintores renacentistas [5], pero más peculiar aún es el hecho de que ni el inventor de la arquitectura GAN (Ian Goodfellow) ni la persona que escribió el código que la generó (Robbie Barrat) recibieron un centavo de la suma [5].

Más allá del dinero, los titulares que hablan de la subasta de la obra tampoco mencionan a Goodfellow o a Barrat, sino a “una obra generada por inteligencia artificial”. En la página oficial de *Christie’s* (empresa que subastó el retrato) por ejemplo, se lee: “El retrato, en todo caso, no es producto de la mente humana. Fue creado por inteligencia artificial, un algoritmo definido por una fórmula algebraica con muchos paréntesis” [6]. Estas pobres palabras, además del cuestionamiento de otros medios sobre el mérito artístico del arte generativo, serán las tesis puestas a prueba a lo largo del presente estudio.

### Problemáticas de la IA general relacionadas al arte generativo

En junio del presente año, Blake Lemoine -un ingeniero de Google de 41 años- declaró públicamente que el programa inteligente de Google, LaMDA (Language Model for Dialogue Applications) era sensible, capaz de sentimientos [7]. El mundo -como era de esperarse-

explotó: los medios más grandes dieron cobertura a Lemoine, los titulares variaron entre lo apocalíptico y lo irreal, y las pregunta -por tiempos descartadas, pero siempre presentes- “¿pueden los algoritmos más inteligentes volverse sensibles? ¿Conscientes? ¿Autónomos?” volvieron a la boca del público.

Ahora bien, habiendo estudiado los comentarios que giran en torno al arte generativo, se destapa un claro paralelismo entre estos y los que giran en torno a la IA en general: En primer lugar, de ambos se desprende una suerte de misticismo que guía al público a relacionar con ellos conceptos como los de autonomía y consciencia, incluso la antropomorfización [5]. En segundo, este misticismo se esconde detrás de imponentes términos como el de la “inteligencia artificial”, pero como esta funciona y lo que implica suele omitirse, incluso malintencionadamente, como en el caso de la página web de *Christie's*.

Si algo es cierto es que la IA solo se va a volver más relevante al pasar de los años. El desconocimiento sobre los procesos que ocurren detrás de bambalinas -cuando ya no se trate de temas hipotéticos sino de responsables de inversiones mal destinadas, diagnósticos equivocados y accidentes de tránsito- será un error muy costoso. Preguntarse sobre el funcionamiento de la IA es entonces primordial, y preguntarse sobre el arte generativo es preguntarse sobre la IA en general (dichosamente). Empecemos por estudiar el funcionamiento de una GAN, arquitectura con la que se realizarán las obras.

## **¿Qué es una red generativa antagónica (GAN)?**

### ***Modelos generativos y discriminativos***

#### ***Modelo generativo***

“Los modelos generativos son aquellos que codifican distribuciones de probabilidad y especifican cómo generar datos que se ajusten a tales distribuciones.” [8]. En términos simples, un modelo generativo se encarga de crear nuevas instancias de datos semejantes a las instancias

de los datos con los que fue entrando. El modelo recibe solo las instancias de datos ( $X$ ) y extrae de ellos patrones de interés para modelar sus propias instancias, cayendo de esta forma en el campo del entrenamiento no supervisado [9].

Así, los modelos que predicen la siguiente palabra en una oración, por ejemplo, son de naturaleza generativa, pues modelan la distribución de probabilidad de las palabras en la oración ( $X$ ) y se preguntan qué tan lejana está la nueva palabra  $x$  en el espacio de datos.

### *Modelo discriminativo*

Estos modelos deciden si una cierta instancia de datos  $X$  pertenecen a la categoría  $Y$  al capturar la probabilidad condicional  $p(Y/X)$  [10] y calcular la función discriminante  $f(X)$ , que mapea cada  $x$  con una etiqueta del conjunto  $Y$  [9].

Mientras un modelo generativo crea fotos inéditas de animales, uno discriminativo se limita a declarar si la foto en cuestión se trata de un perro o de un gato. Estos ignoran si la nueva instancia  $x$  es probable o cercana a la distribución de  $X$  para encargarse de especificar qué tan aplicable es una etiqueta  $Y$  a dicha nueva instancia [10].

### *Modelo generativo vs discriminativo*

Los modelos generativos tienen una tarea mucho más complicada que los discriminativos por el simple hecho de que tienen que modelar más: mientras el primero saca conclusiones de tipo: “Las cosas que parecen botes son probables a estar cerca de cosas que parecen agua”; el segundo debe aprender a diferenciar entre un bote y un no-bote basado en patrones evidentes, ignorando así todo lo demás [10].

El siguiente ejemplo captura bien sus diferencias esenciales:

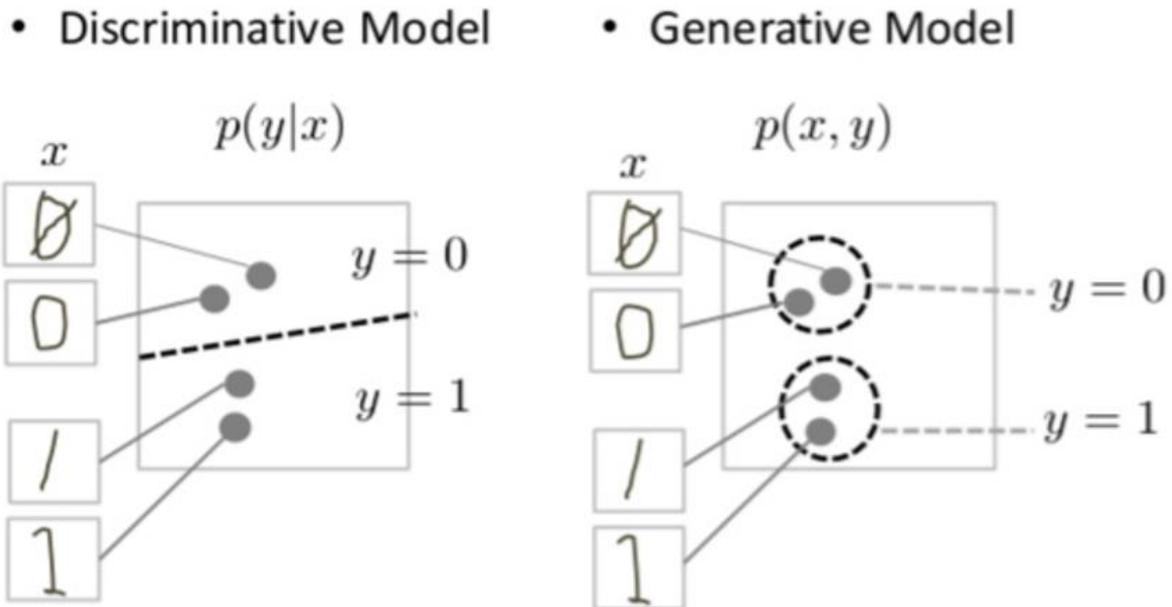


Figura 1. Modelos discriminativo y generativo de números escritos a mano [10]

De ser exitoso, el modelo discriminativo distingue las diferencias entre un cero y un uno al dibujar una línea que corta al espacio de los datos en dos: la parte donde están los ceros y la parte donde están los unos.

En contraparte, el modelo generativo intenta entender la distribución de las instancias para generar sus propios ceros y unos que caerán cerca de sus contrapartes reales [10].

### ***Definición y arquitectura de una GAN***

#### *Definición*

Las “GAN” (Generative Adversarial Networks) fueron introducidas por Ian Goodfellow, et al. en 2014 [11]. En síntesis, una GAN se puede describir como una arquitectura para entrenar un modelo generativo en el que comúnmente se emplean técnicas de deep learning [12]. Estas se comprenden de dos redes neuronales: El Generador, que se encarga de generar nuevas instancias a través del modelo generativo; y el Discriminador, que clasifica las instancias como

reales o generadas a través del modelo discriminativo. Estos interactúan entre sí en un escenario basado en la teoría de juegos donde el Generador compite contra el Discriminador hasta engañarlo (esto es, cuando el Discriminador ya no puede diferenciar entre instancias reales o generadas) [13] de ahí su nombre.

### *Generador*

El generador recibe un vector aleatorio de tamaño fijo sacado normalmente de una distribución gaussiana y devuelve una instancia de datos. El Generador puede aprender características estadísticas de imágenes, de música, de historias y en base a esa abstracción del espacio con el que fue entrenado, generar nuevas obras artísticas [14].

### *Discriminador*

Este es simplemente un clasificador que toma una instancia del dominio (el set de instancias del entrenamiento más el set de instancias generado) y predice una clase binaria (real o generada). Para ello puede utilizar cualquier arquitectura bien conocida y efectiva. Este sub-modelo es descartado una vez la fase de entrenamiento haya culminado, a menos de que parte de él sea reciclado en procesos de transferencia de aprendizaje [13].

### *Entrenamiento*

El entrenamiento es en donde recae la brillantez de la arquitectura. Este ocurre simultáneamente entre ambos sub-modelos previamente discutidos, en el marco de lo que se conoce en la teoría de juegos como “Zero-sum game”: una representación matemática de una situación en la que dos lados compiten y cuyo resultado es la ventaja de un lado, y la equivalente pérdida para el otro [11]. Así, mientras el Generador cree instancias poco convincentes será penalizado y sus parámetros serán actualizados a través de backpropagation,

y mientras el Discriminador no sepa diferenciar entre instancias generadas por el Generador y las reales, será penalizado también y deberá actualizar sus parámetros de la misma forma [11]. El flujo continuo de feedback entre los sub-modelos está bien descrito por el siguiente diagrama:

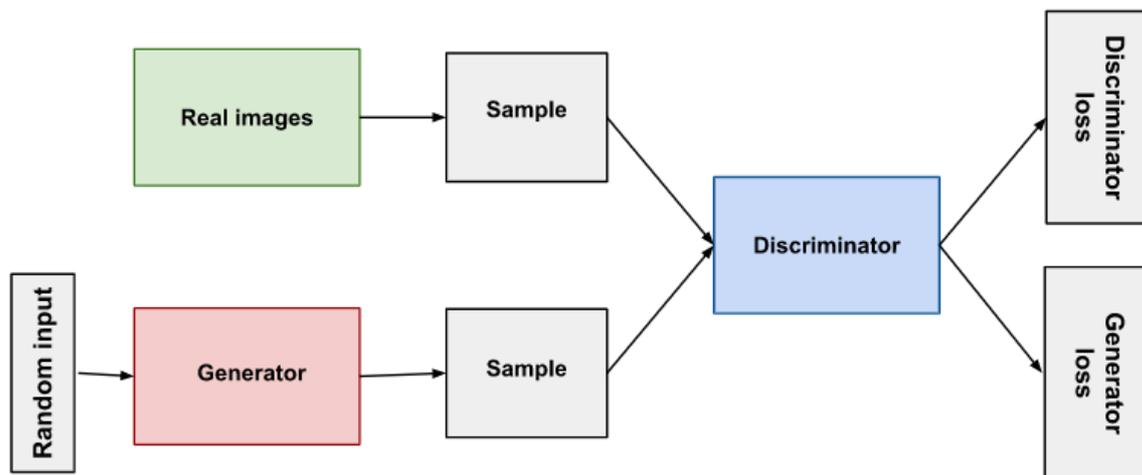


Figura 2. Diagrama de flujo de una GAN [10]

El output del Generador está conectado directamente al input del Discriminador, quien recibe la instancia generada, pero también una instancia real del set de datos de entrenamiento. Este etiqueta la imagen generada como real o generada y, dependiendo de si su decisión fue acertada, los parámetros de la parte perdedora se actualizan para esperar mejores resultados [10].

Este proceso ocurre hasta la convergencia, esto es, cuando la efectividad del Discriminador es del 50% (en el caso ideal, valores cercanos a este en casos reales) lo que significa que el Discriminador está básicamente adivinando y, por ende, que ya no puede diferenciar entre instancias reales y generadas [10]. La convergencia no es un estado estable, sino más bien uno transitorio, por lo que debe ser tomada con cuidado: si la GAN continúa entrenando cuando el

Discriminador solo produce output aleatorio, el Generador se entrenará con feedback basura, decayendo así su calidad [10].

### *Loss functions*

El propósito de entrenar una GAN es el de acercarnos a la distribución de probabilidad del set de datos  $X$  con la mayor precisión posible. Ahora ¿Cómo sabemos que tan cerca (o lejos) nos encontramos? Para ello se han definido funciones que, tomando el valor de output obtenido y el valor de output esperado tanto del Generador como del Discriminador, regresen un valor que sirva para actualizar los parámetros con los que toman decisiones en la dirección correcta [10]. Existen, por el momento, dos funciones principales que cumplen con este propósito.

### *Minimax loss*

Esta función, derivada de la métrica “cross-entropy”, fue sugerida en el artículo de Ian Goodfellow, et al. en el 2014. Se define como [10]:

$$E_x[\log \log D(x)] + E_z[\log \log (1 - D(G(z)))]$$

Ecuación 1: Ecuación Minimax loss [10]

Donde  $E_x$  es el valor esperado de los datos reales,  $D(x)$  es el valor estimado por el Discriminador de la probabilidad de que la muestra real  $x$  sea real,  $E_z$  es el valor esperado de las instancias generadas,  $D(G(z))$  es el valor calculado por el Discriminador de la probabilidad de que la muestra generada  $G(z)$  sea real, y  $G(z)$  es la muestra generada por el Generador dado el vector aleatorio  $z$  [10].

Nótese que el Generador y el Discriminador pretenden minimizar y maximizar el resultado de esta ecuación correspondientemente. En el caso del Generador, ya que este no tiene ninguna influencia en el lado izquierdo de la ecuación, basta con minimizar el término  $\log(1 - D(G(z)))$  [10].

### *Wasserstein loss*

Esta, en cambio, deriva de la métrica sobre distancias entre distribuciones de probabilidad “earth mover” y su uso implica un nuevo esquema derivado del tradicional llamado “WGAN”, en el que el Discriminador no etiqueta instancias de forma binaria, sino que asigna un valor que intenta que sea mayor para instancias reales [10]. A este Discriminador alternativo se lo llama el Crítico. Este nuevo esquema es útil en tanto las WGANs son menos vulnerables al estancamiento relacionado al desvanecimiento de gradiente o al “mode collapse” (estado en el que el Generador genera instancias demasiado similares entre sí, ya que estas engañan bien al Discriminador) [10].

Aquí, el Generador y el Crítico tienen diferentes loss functions:

Loss function del Crítico:

$$D(x) - D(G(z))$$

Ecuación 2: Loss function del Crítico [10]

Donde  $D(x)$  es el valor asignado por el Crítico a la instancia real  $x$ , y  $D(G(z))$  el valor asignado por el Crítico a la instancia generada por el Generador dado el vector aleatorio  $z$ .

Se debe notar que el Crítico, al querer maximizar esta ecuación, lo que busca en verdad es maximizar la diferencia entre el valor que asigna a una instancia real y a una generada [10].

Loss function del Generador:

$$D(G(z))$$

Ecuación 3: Loss function del Generador [10]

El Generador, al intentar maximizar este término, lo que busca es maximizar el output del Crítico para las instancias generadas.

## **Estado del arte**

En el presente apartado se busca describir los mecanismos específicos de la arquitectura a utilizar en el marco de la transparencia y explicabilidad definidos por la ISO/IEC.

Sobre el primero, el estándar “Overview of trustworthiness in artificial intelligence” (24028) menciona que la búsqueda de transparencia de procesos inteligentes a través de la accesibilidad de sus características, algoritmos, modelos y métodos de entrenamiento son necesarios en un escenario inherentemente opaco (esto es, difícil de interpretar para un humano) como este [15]. Adicionalmente, sobre la explicabilidad, se expone que la explicación es siempre un intento de comunicar entendimiento. La efectividad del intento recae en el acoplamiento al contexto en el que la explicación ocurre, considerando la audiencia a la que está dirigida y el nivel de entendimiento que se pretende transmitir [15]. Sobre este apartado cabe decir que -dado que esta investigación pretende llegar tanto a personajes de trasfondos técnicos como artísticos- un mínimo de entendimiento técnico puede llegar a ser necesario, pero parte de su motivación es la de explicar procesos complejos con conceptos generales.

## **VQ-GAN**

En la obra a realizar se usará una arquitectura derivada de la GAN, la “vector quantised GAN” (o VQ-GAN). La idea fue presentada por vez primera por Esser, Romachar y Ommer en el 2021, en su ya famoso artículo: “Taming Transformers for High-Resolution Image Synthesis” [16]. Las VQ-GANs -en síntesis- son una arquitectura de Deep Learning capaces de extraer características del set de imágenes de entrenamiento con admirable fidelidad, además de generar imágenes inéditas y de alta resolución que bien podrían pasar como reales [16].

A continuación, se introducirán -con el fin de explicar el mecanismo con mayor profundidad y así enriquecer la discusión derivada de nuestro tema- los componentes que hacen de este un procedimiento posible.

### *AutoEncoder*

Los “AutoEncoders” describen una técnica de compresión construida a partir de dos redes neuronales artificiales: “Encoder” y “Decoder” [17].

El input es alimentado al Encoder y comprimido (o “encoded”) a un espacio dimensional reducido (espacio latente) para luego ser reconstruido proyectándolo al espacio dimensional original a través del Decoder, siempre intentado reducir el error de reconstrucción basándose en el “loss” del Decoder [17].

Así, es preciso que el modelo sea entrenado de tal forma que el Encoder aprenda a comprimir el tipo de dato específico a un espacio latente que guarde el suficiente potencial para que un Decoder entrenado pueda reconstruir el input [17].

### *Espacio latente*

En esencia, el espacio latente debe ser entendido como la representación comprimida de datos, pero ¿para qué comprimir datos? Resulta que la compresión de datos (y el concepto de espacio latente en particular) está en el corazón de Machine Learning, pues el proceso de reducción de dimensionalidad conlleva la abstracción de los datos en sus características y patrones más esenciales [18].

Ahora ¿de dónde sale el nombre? Por un lado, la palabra “latente” significa oculto [18] y es que -como se puede observar del funcionamiento de los AutoEncoders- la representación reducida de las imágenes (es decir, el espacio latente) no nos es visible de la misma forma en

la que los son los inputs o los outputs del modelo. Este, más bien es un agente intermediario, digno de la abstracción computacional.

Por otro, la palabra “espacio” se deriva del concepto de dimensión en el contexto del álgebra lineal [18]. Nuestro espacio material es tridimensional, pues lo podemos describir en función de tres características: largo, ancho y profundidad; ahora, pensemos en un punto con cuatro dimensiones, donde su dimensión adicional es el color: Este nuevo punto traído del espacio cuatro-dimensional no sólo es comparable con los demás puntos del espacio por su posición, sino también por qué tanto los colores entre ellos se parecen, abriendo campo para derivar nuevas conclusiones. Los espacios latentes de hecho pueden tener muchísimas dimensiones y este es un parámetro de peso:

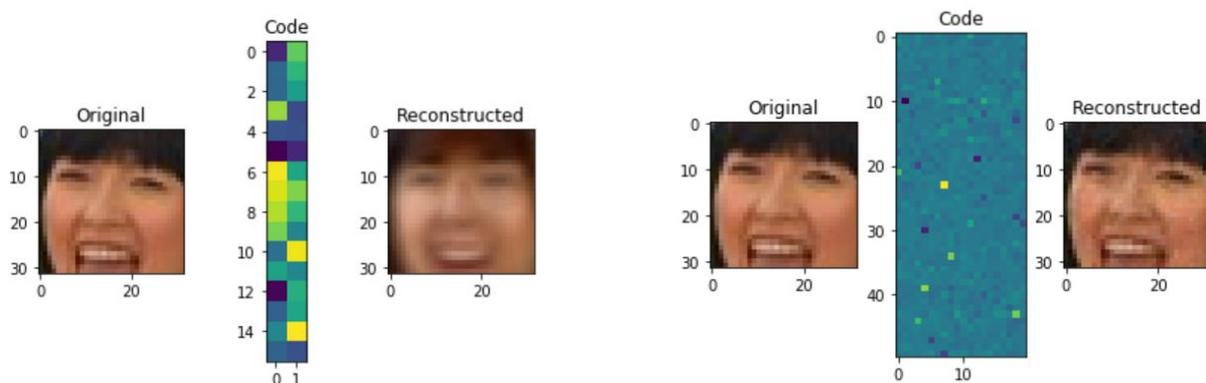


Figura 3. Diferencia al variar la dimensionalidad del espacio latente (2x16 a la izquierda, 20x60 a la derecha) en reconstrucción imágenes con AutoEncoders [19]

### *Variational AutoEncoders (VAE)*

En el escenario del AutoEncoder original, no habría razón para creer que los vectores latentes provenientes de dos imágenes lógicamente relacionadas (dos fotos de un perro, por ejemplo)

se encuentren cerca, sino arbitrariamente alejados, si decidimos comparar sus vectores [20]. Al AutoEncoder original le basta con que el Decoder regenere la imagen, pero el hecho de que los vectores sean comparables entre sí juega un rol importante durante el proceso de cuantificación vectorial, que es el siguiente proceso a tratar.

Para que la distribución de los vectores en el espacio latente del *variational AutoEncoder* (VAE) estén distribuidos uniformemente, el Encoder no solo retorna el input codificado, sino también la media y la distribución estándar de la distribución latente generada [21]. Con estos nuevos elementos se puede calcular un nuevo loss: la divergencia entre la distribución gaussiana y la verdadera distribución del espacio latente generado. Así, durante el entrenamiento, las redes neuronales no solo aprenderán a recrear la imagen alimentada, sino a generar un espacio latente normalmente distribuido [21]. El impacto de la variación es ilustrado a través de la siguiente figura.

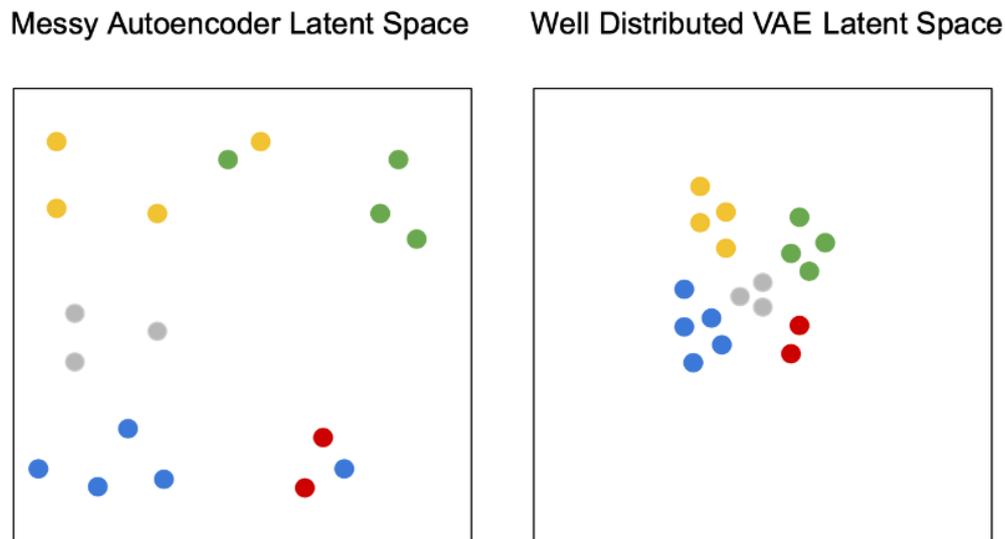


Figura 4. Un espacio latente arbitrariamente distribuido (izquierda) vs uno normalmente distribuido (derecha) [20].

### *VQ-VAE (Vector Quantised Variational AutoEncoder)*

Esta variación de VAE -según el artículo que la introdujo en el 2017: “Neural Discrete Representation Learning”- difiere del VAE en dos puntos fundamentales [22]. En el primero, la arquitectura VQ-VAE discretiza el espacio latente generado por el Encoder a partir de la cuantificación vectorial: estos es, los grupos de vectores impresos en la Figura 4 (derecha) pasan a ser representados por un solo vector, de la misma forma en la que lo es el medoide de un clustering [21].

La segunda diferencia es la generación del “codebook”, representación latente en el que los vectores representantes de cada característica abstraída en el espacio latente (o “codewords”) son enlistados [21].

### *Transformers*

Introducidos en el 2017 por Vaswani et al. en el artículo “Attention is all you need”, los Transformers pueden ser entendidos como una arquitectura de tipo “sequence-to-sequence” particularmente talentosa en el procesamiento de lenguaje natural [23].

Por una arquitectura tipo sequence-to-sequence (o Sq2Sq) se entiende una red neuronal que - usando la lógica Encoder-Decoder- toma una secuencia de elementos y los transforma en otra secuencia [24]. Un buen ejemplo de aplicación de este tipo de procesamiento es el de la traducción: se toma una secuencia de palabras en el idioma *A*, se las procesa, y como output tenemos otra secuencia, esta vez en el idioma *B*.

Para esto se valen de lo que se conoce como “mecanismo de atención”. En este escenario se quiere averiguar qué tan importante es la *i*-ésima palabra para el significado de la oración. Para ello se calcula el “vector de atención” por cada palabra procesada, que captura relaciones contextuales entre las palabras en una oración y cuya información facilita enormemente la tarea

de generar una nueva secuencia prediciendo elementos en vez de traducirlos secuencialmente [24].

### *VQ-GAN*

Ahora que hemos introducido a todos los elementos de la arquitectura de la VQ-GAN, miremos como estos trabajan (Figura 5) entre sí para realizar la obra, tomando siempre en cuenta que para usarla en el contexto del arte generativo son necesarias tres fases.

En la primera, el Encoder y el Decoder (ambos CNNs o redes neuronales convolucionales) se entrenan al estilo VQ-VAE hasta que el Decoder sepa reconstruir la imagen que codifica el Encoder, generando el codebook ( $Z$ ) y el espacio vectorial cuantificado ( $Zq$ ) [21]. La diferencia con la VQ-VAE radica en que -y aquí es donde le juego adversarial entra en juego- durante la primera etapa del entrenamiento no solo se toman en cuenta el *reconstruction loss* del AutoEncoder original y el *loss* del VAE, sino también el *loss* adversarial generado a partir de la interacción entre el Discriminador “D” y el Generador “G” (que también hace las veces de Decoder) [21].

En la segunda fase, los pesos del Decoder y el Encoder son congelados para entrenar solo el Transformer (particularmente, el GPT-2 [Generative Pre-trained Transformer] de open AI [16]) quien aprende a predecir secuencias usando el espacio vectorial cuantificado ( $Zq$ ) como target. Así, este aprende a construir un espacio como el de  $Zq$  a partir de las características que extrae del codebook en forma de vectores [21].

Finalmente, el input (que es ingresado como texto por el usuario) es procesado por el Transformer quien, en base al mecanismo de atención y dada su naturaleza Sq2Sq, abstrae características del input, las extrae del codebook y predice un Zq inédito, que servirá al Decoder ya entrenado para construir una imagen inédita también [16].

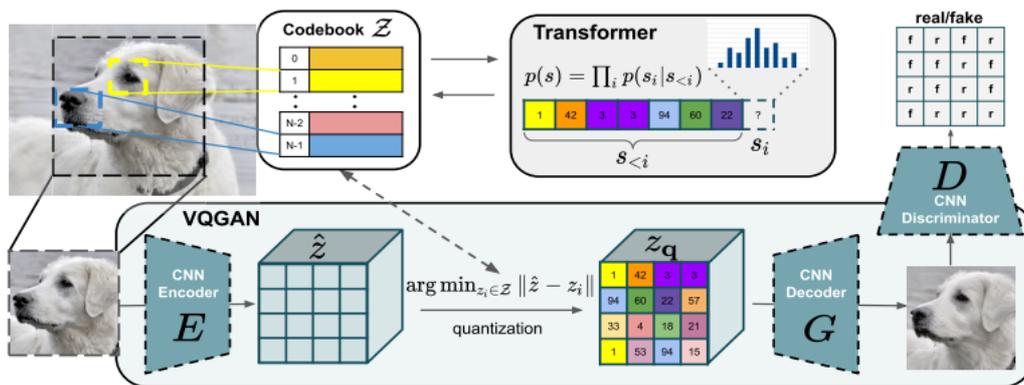


Figura 5. Arquitectura de la VQ-GAN [16]

## DESARROLLO DE UNA OBRA ARTÍSTICA

Una vez que hemos descrito lo necesario para entender el comportamiento básico de una VQ-GAN y cómo convierte esta texto en una imagen, es momento de indagar en la implementación de una obra con ella.

Para ello vamos a hacer uso de PyTTI-Tools (Python Text To Image) una colección de herramientas para facilitar la generación, animación y manipulación de imágenes que implementa una VQ-GAN pre-entrenada [25].

Desarrollada por Katherine Crowson, Henry Rachootin y David Marxs [25] PyTTI-Tools es una versión gratuita del código original Rachootin entendida para usuarios de alto nivel, entre quienes este tipo de herramientas se han vuelto muy populares debido a su impresionante poder generativo y la relativa facilidad con la que este es accesible.

A continuación, se describirán los pasos que se tomaron para realizarla, para después discutir los resultados, las impresiones de un artista generativo y, por último, abordar la discusión de la problemática inicial habiendo mojado los pies tanto en el campo teórico como práctico.

Los links, tanto al notebook de PyTTI-Tools y al código en Github (Anexo B), como a los diferentes prototipos realizados con él y la obra final (Anexo C) serán adjuntados a los anexos.

## **Metodología**

### **Google Colab**

El código se corrió en la plataforma de Google Colab debido a la necesidad (propia de la gran mayoría de arquitecturas de Deep Learning) de GPUs. También se puede usar una máquina local que los tenga, pero este no era el caso y probablemente tampoco el de gran parte del público objetivo.

Google Colab permite conectarse a GPUs de forma remota para correr el código de manera gratuita, pero debido a las limitaciones que la versión gratuita trae consigo -GPUs inferiores y frecuentes interrupciones durante el proceso- es recomendable suscribirse a la membresía “pro”. Antes de conectarse a la máquina remota, se debe cambiar el apartado de “Hardware accelerator” de CPU a GPU.

### **Paso 1: Conectar ejecución a Google Drive**

Paso opcional, pero recomendado: El video se armó a partir de la secuencia cientos de imágenes individuales, la carpeta donde se guardan y los modelos pre-entrenados que se descargan automáticamente pueden llegar a pesar mucho. Correr la celda 1.1 y hacer click en la casilla “mount\_gdrive” para que las carpetas necesarias sean creadas con los nombres predeterminados y la ejecución sea fluida.

### **Paso 2: Configurar experimento**

Tengamos presente que se trabaja -en su vasta mayoría- con modelos pre-entrenados (VQ-GAN con imagenet.ckpt, Autoencoder convgg.pth) y por ende con entrenamientos no parametrizables (recordemos que el notebook pretende ser usado por usuarios de alto nivel) por lo que en este apartado, las variables llamadas “Prompt Settings” son en realidad donde el usuario puede explotar su imaginación y creatividad.

### *Prompt Settings*

Aquí se describen las escenas a generar en forma de texto. Los elementos a introducir deben separarse por un “|” y además pueden llevar un peso (especificado después de dos puntos) que, dependiendo de su valor (positivo o negativo) estará más o menos presente en la obra.

Los autores recomiendan prácticas como tomar en cuenta cuando fue entrenado el modelo para no tener errores de tipo “introducir la estética de una película que se hizo pública en el 2022 a un modelo que se entrenó en 2020”. También se recomienda el uso de fuentes de imágenes en la web. En este caso, por ejemplo, se especificó el sitio de un artista para que el estilo de este sea imitado: artstation art by Azat Nurgaleev (Anexo D).

Los valores que se introdujeron en este apartado para el prototipo 1 fueron:

**scenes:** eternal feminine | big translucent women eyes | rivers of wine

**scene\_suffix:** | satellite image:-1:-.95 | text:-1:-.95 | watermark:-1:-.95 | backyard telescope:-1:-.95 | map:-1:-.95

**scene\_prefix:** astrophotography #pixelart | artstation art by Azat Nurgaleev | #conceptart eternal feminine |

### *Otros parámetros configurados*

A continuación, la descripción de los demás parámetros que se configuraron para la realización de la obra. Cabe mencionar que -si bien estos no tienen voz sobre los elementos protagonistas que aparecerán en la obra ni relación con el campo de la ciencia de datos- sí pueden determinar rasgos estéticos de ella.

Ejecutar las 19 celdas de este apartado una vez configurados los parámetros descritos para guardarlos.

Parámetro	Descripción	Valor introducido
steps_per_scene:	Total number of steps to spend rendering each scene	60100
image_model:	Select how your image will be represented	vqgan
pre_animation_steps	Number of steps to run before animation starts, to begin with a stable image.	100
sampling_mode	How pixels are sampled during animation	bicubic
animation_mode	Select animation mode or disable animation	3D
translate_x	Horizontal image motion as a function of time in seconds.	$-1700 * \sin(\text{radians}(1.5))$
translate_z_3d	Forward image motion as a function of time in seconds	$(50 + 10 * t) * \sin(t / 10 * \pi) ** 2$
rotate_3d	Image rotation as a quaternion as a function of time in seconds.	$[\cos(\text{radians}(1.5)), 0, -\sin(\text{radians}(1.5)) / \sqrt{2}, \sin(\text{radians}(1.5)) / \sqrt{2}]$

Tabla 1. Otros parámetros configurados para la realización del prototipo 1.

### Paso 2.3: Ejecución

Posteriormente se debe correr la celda 2.3 para generar las imágenes. Fue a esta altura donde se topó con el primer problema: Al correr la celda saltaba el error “pytti\_test\AdaBins\pretrained\archive file not found” lo cual fue resuelto -después de algo de indagación- descargando el modelo pre-entrenado de Adaptive Bins y descomprimiéndolo en la carpeta con esa dirección en Drive. Así mismo, el link para descargar el modelo estará en los anexos (Anexo E) si alguien se topa con el mismo inconveniente.

Una vez resuelto el problema, las imágenes son generadas e impresas cada 50 unidades en pantalla, guardándose en la carpeta “images\_out” del directorio creado por el código.

### **Paso 3: Renderización del video**

Correr la celda 3 debería renderizar el video a generar en función de las imágenes en la carpeta “images\_out”, pero durante la experimentación esto no ocurrió. Habiendo ya generado las imágenes el problema no era crítico, por lo que se utilizó la parte del código de Henry Rachootin “pytti 5 beta” que renderiza los videos para hacerlo. Este código no será compartido, siendo este una fuente de ingreso del programador y exclusivo para usuarios que han pagado por él. En todo caso, se debe recalcar que este -teniendo ya las imágenes en la carpeta- es un problema relativamente trivial y su solución puede venir de la mano de un código simple generado por el usuario o de un software auxiliar.

## **Resultados**

Con los parámetros descritos, el tiempo de ejecución es de aproximadamente 14 horas, dependiendo de la disponibilidad de equipos remotos de Google Colab (incluso la versión pro tiene limitantes) pero se decidió parar con la ejecución una vez los elementos se mostraran repetitivos y no se viera una clara evolución de la obra. En total se realizaron cinco prototipos, cuyas configuraciones y versiones finales se encuentran en el Anexo C.

Para el análisis de los presentes resultados es pertinente indagar brevemente en la motivación del agente que está tomando el rol del “artista”: las siguientes imágenes se hicieron en base a letras de canciones, versos e imágenes impresas a través de música previas al presente proyecto. Estas ejemplifican virtuosamente el puente ciencia-arte; identificamos con ellas conclusiones de la misma naturaleza.

### *Descripción de los resultados*

Lo que salta a primera vista (Figura 6) es el protagonismo del parámetro “big translucent women eyes”, no solo por la presencia del elemento “ojos”, pero las características que el modelo supo definir y de las que se adueñó para hacer de estos “ojos de mujer”: pestañas largas, contornos sombreados, cejas delgadas.

El concepto del “eterno femenino”, por otro lado, se deja atestiguar por partes del cuerpo femenino que, sin tanto protagonismo como los ojos, componen la obra desde un segundo plano. Aquellas partes que parecen diluirse sobre un blanco manto bien pueden imitar el flujo del río de vino, pero la ausencia del licor parece indicar que estos imitan las figuras que dan forma al cuerpo de una mujer, lo que a su vez deja formular la pregunta ¿por qué están relacionados el concepto de feminidad y lo que hasta ahora se ha descrito como mantos blancos? ¿Hay una forma de averiguar si lo son, si están relacionados? Estas preguntas serán abordadas en las conclusiones. Por ahora, cabe decir que estas no tienen una respuesta clara

debido a la complejidad de cálculos latentes y procesos implícitos, y que toda un área de investigación se ha formado alrededor de esta problemática.

Por último, es necesario mencionar que la estética misma de la obra (trazos gruesos, figuras que no pretenden ser realistas) tiene una clara relación con la estética del artista citado; y que los elementos de tipo “cósmicos” (presentes a lo largo de toda la secuencia, pero más evidentes en la imagen 169 y en la 253) se deben al parámetro “astrophotography”.

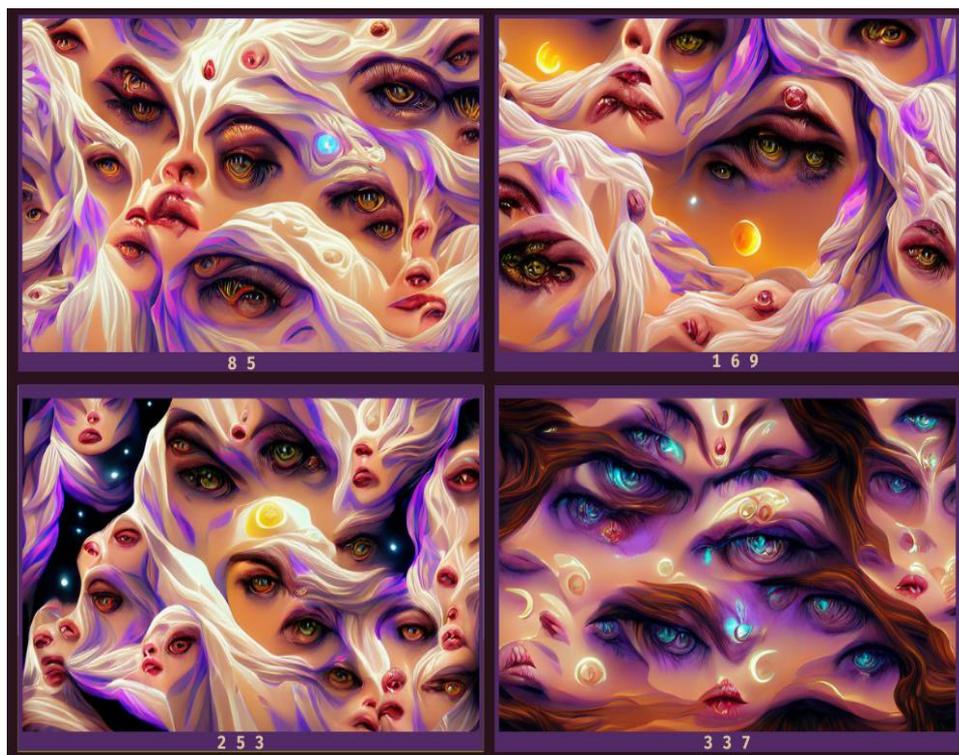


Figura 6. Cuatro imágenes que comprenden el prototipo 1 tomadas a saltos de 84 unidades. Abajo de cada una su id secuencial. Nombre oficial: Anima.

Las imágenes que configuran el prototipo 2 (Figura 7) son de una composición más simple que las que acabamos de observar. El texto introducido para este prototipo fue: “eternal feminine | dress made out of stars | rings made out of the moon | space full of cybernetic neon:3\_galactic nebula”

El concepto de eterno femenino fue usado de nuevo con el propósito de contrastar e identificar su efecto en las obras, mas este parece sucumbir ante las telas negras que responden al componente que de los vestidos que -más que hechos de estrellas- aparentan un mantel sobre el que las nebulosas se riegan como canicas.



Figura 7. Cuatro imágenes que comprenden el prototipo 2 tomadas a saltos de 20 unidades. Abajo de cada una su id secuencial. Nombre oficial: Canicas sobre el mantel.

Los elementos espaciales responden -de nuevo- al input del usuario, por lo que no merecen mayor discusión. Los que sí la merecen son “los anillos hechos de luna” que, además de no parecer en nada a la idea inicial que sirvió como motivación para su introducción (la luna - esférica- sirviendo como piedra preciosa sobre un anillo) trajeron consigo también una figura no deseada: las manos.

En base a lo que se ha discutido en el presente documento, la presencia de estas figuras no debería parecer arbitraria, ni lo es: las imágenes reales de anillos son acompañadas con mucha

probabilidad por manos, el modelo fue entrenado con estas imágenes y abstraídas como características de ellos en forma de vectores latentes. Cuando el Transformer extrajo características relacionadas a un anillo del codebook, extrajo consigo lo que nosotros percibimos como manos, pero lo que este considera como una característica esencial de anillos. Por último, el Decoder simplemente generó con aquello que se le fue dado.

Es evidente así, al analizar el prototipo 2 en particular, el problema con el que un artista puede toparse: la discrepancia entre lo que el artista cree que describe y lo que el modelo termina generando.

## ***Discusión***

### *La IA y el arte*

A esta altura es necesario contexto: ¿qué hay para discutir entre la IA y el arte?

En el 2022, Eva Cetinic y James She se inclinaron por la pregunta ¿tienen los procesos inteligentes, autonomía? Después de describir el episodio de *Portrait of Edmond de Belamy* en el artículo “Understanding and Creating Art with AI: Review and Outlook” [26]. En él, argumentan que los procesos inteligentes detrás del funcionamiento de una GAN caen estrictamente en el campo de la imitación y no de la inteligencia normativa; las GANs extraen el espacio latente que llega desde el input, no lo eligen [26].

Sobre el mismo tema, Hertzmann (contribuidor directo de algoritmos generativos como “Image Analogies” [1]) va más allá al decir que resulta irresponsable el guiar al público general a pensar que la IA tiene atributos humanos y que, como describe Notaro, la motivación para hacerlo responde a jugadas de marketing de la misma forma en que ocurrió en el caso de *Portrait of Edmond de Belamy* [26].

Hertzmann hace además un punto muy válido en su trabajo “Can Computers Create Art?”, comparando lo que significó la cámara para el arte con lo que significa el modelo generativo para arte [1]. Al hacerlo relata cómo en un inicio, el mundo de la pintura tembló frente al poder imitativo de una cámara, poder muy superior al de cualquier pintor y razón suficiente para que estos se pregunten: “¿por qué no ha de reemplazarme?” [1]. Lo cierto es que ocurrió lo contrario: este (el arte pintado) en vez de morir, se reinventó. Para entender cómo, se cita a Edvard Munch, autor de “El grito”: “No tengo miedo de la fotografía en tanto no puede ser usada en el cielo ni en el infierno. [...] Voy a pintar a personas que respiren, sientan, amen y sufran” [1]. Munch profundizó en su tipo de abstracción, de la misma forma lo hizo Van Gogh y el resto de artistas que no pretendían tomar una fotografía, abriendo ramas nuevas y posibilidades en el proceso.

#### *El artista generativo desde la experiencia*

En primer lugar, se deriva -de los problemas que se explican en la metodología- la primera recomendación para el futuro artista generativo: este deberá -antes de entender los procesos que ocurren a nivel conceptual- entender los procesos a nivel práctico, pues no hay forma de que fluya creativamente de otra forma: los obstáculos se abordaron con relativa facilidad, porque se abordaron desde la familiaridad con código, ambientes dinámicos de programación e incluso con conceptos de IA que el usuario medio no tiene.

En segundo lugar, y para pesar en la discusión que introdujeron Cetinic y She, se vale de la nota escrita al margen durante la experimentación: “Me es fácil hacer algo bello, me es difícil hacer lo que quiero”.

Realizar imágenes aparentemente estéticas no fue complicado de la misma forma en la que, para tomar una fotografía decente, no hace falta más que apuntar y enfocar la cámara de un celular. Es en el sentido en el que Cetinic y She describen la mano del artista generativo en la

que verdadera complejidad y propósito son reconocibles: “Desde la perspectiva del artista, el espacio latente no es ni un espacio real ni uno imaginario, sino un campo de sugerencias inagotables que emergen de la interacción entre lo conocido y lo desconocido. Cómo uno orquesta el diseño de este espacio y lo que encuentra en él eventualmente se vuelve la mayor tarea y la firma del artista” [1].

Por último, vale mencionar que es necesario que tanto el artista generativo, como el público participe de él funden, un nuevo tipo de discusión: dado su nivel de abstracción, el mérito de los procesos ocultos y su fácil acceso; el arte generativo debe y merece apuntar al nacimiento de una voz crítica que no tenga miedo de profundizar en la ciencia computacional (su abstracción, sus procesos, sus limitantes) pero tampoco en el discurso artístico que encarna. Que sea resistente a la romantización de la IA con fines lucrativos de la mano de una voz que sepa conciliar el arte y a la ciencia con madurez, y que se nutra de estos dos ejes dinámicamente.

### *El futuro del arte generativo*

Por un lado, un tema a señalar es el hecho de que el arquetipo de feminidad abstraído por el modelo en el primer prototipo responde a cánones de belleza puntuales y estáticos. Tomando en cuenta que los modelos aquí descritos están -y seguirán estando- íntimamente relacionados con el arte, y que este se caracteriza por hospedar discusiones derivadas del pensamiento crítico; esta es una discusión inminente y una buena oportunidad de entrelazar conscientemente el arte y la ciencia.

Por otro lado está el de considerar la introducción del estudio de procesos estocásticos en la discusión del arte. Que no se entienda por esto que estos procesos no han jugado un rol en el

arte hasta este punto (más bien, todo lo contrario) sino que sus implicaciones simplemente no son discutidas lo suficiente.

Hay una diferencia entre un sistema caótico y una verdaderamente aleatorio, siendo el primero uno en el que es muy complejo predecir el siguiente estado del sistema (y por ende muchas veces se nos presenta como aleatorio) y el segundo, uno -de forma simple- en el que no hay forma de predecir el siguiente estado [27]. A continuación, se tratarán episodios caóticos como aleatorios a pesar de esta definición, siendo aún verdad que -para la experiencia humana- lo caótico se manifiesta como incierto, y lo seguirá haciendo mientras no se encuentre la forma de cuantificar y considerar todas las variables que componen la realidad.

Así, Dorin señala como la aleatoriedad y diferentes técnicas generativas han trabajado en conjunto -de la mano de la tecnología- para potenciar la creatividad humana en varias ocasiones [28]. El “Makapansgat”, por ejemplo (una roca rojiza recogida por el Australopithecus y estudiado como una de las primeras abstracciones artísticas del humano [28]) fue esculpida por el flujo del río. Tres millones de años después, el compositor Steve Reich grabó la distorsión que surge del movimiento de micrófonos conectados al techo, moviéndose frente a un amplificador, en lo que llamó “pendulum music” [28]; Erwin Driessens y Maria Verstappen presentaron las estalagmitas que se terminan generando al dejar caer libremente cera desde el techo [28] y Jackson Pollock cambió el juego de la pintura introduciendo al expresionismo abstracto. Todos estos se encuentran en el Anexo F para su revisión.

Ahora, tomando la anterior idea en consideración, es realmente sustancial subrayar que, siendo la naturaleza estocástica no un valor agregado (como en los ejemplos anteriores) sino uno fundamental del arte generativo [28] este tiene la inmensa oportunidad -si no responsabilidad- de profundizar en el tema desde su interacción explícita y consciente.

Cómo se verían los resultados aquí presentados era una incógnita en nada alejada a las formas en las que se construyen los cuerpos de cera de Driessens y Verstappen (de hecho, entre ambos

resalta una nítida analogía: la cera y las imágenes de entrenamiento como como material conocido y primario, resultados derivados de ellos caóticos y abstractos). Lo cierto es que, hoy por hoy, el arte generativo se lleva a cabo a ciegas o bajo una suerte de luz de vela que -hasta ahora- parece ser una problemática, pero ¿lo es?

La introducción de la estocástica pretende diseñar suerte, considerar a los accidentes como como una fuerza creativa, abrazar el caos [28]. Su estudio y entendimiento como eje de la naturaleza misma permitirá la indagación de la creatividad a través de su imitación en procesos generativos. En resumen, el arte generativo precisa del renunciar de una parte del poder para ponerlo en las manos de una fuerza que esculpe, tanto la forma de la totalidad del universo, como la de cada ínfimo momento.

## CONCLUSIONES

Se empezó esta indagación en el arte generativo con preguntas sobre el nivel de autonomía de procesos inteligentes, sobre la validez de la rama artística y sobre el ilegible horizonte que los avances de la ciencia de datos dibujan para el futuro.

Respecto a la pregunta de la autonomía, el argumento de Herzmann sobre la romantización de la IA con fines lucrativos ganó peso a medida que se dilucidó un proceso inteligente que bien puede ser descrito como complejo, profundamente abstracto e incluso misterioso; pero nunca autónomo. Un computador no decide hacer arte, no reflexiona sobre su propósito ni se regocija de su valor. Las imágenes, texto e incluso código generado por IA pueden llegar a ser atemorizantes, pero todos ellos se pueden desglosar en la misma forma en la que ocurrió en el presente proyecto; exhibiendo no más que técnicas brillantes que, en vez de favorecer al argumento de la máquina consciente, hablan sobre el ingenio de la mente humana.

Sobre la validez de la rama, las razones para aceptarla abundan. Por un lado, los conocimientos necesarios para esculpir figuras con datos la convierten en una verdadera disciplina, y la crítica con la que el público objetivo deberá responder es una con un pie en la matemática y la otra en su filosofía; siendo el marco de la manos en el prototipo 2 una oportunidad para que futuros artículos concilien -de la mano del “feature visualization”- prácticas técnicas con conceptos como estética o cromática. Por otro lado, el estudio de procesos estocásticos en función del arte lo provee de suficiente novedad y potencial para una revolución; para encontrar nuevos escenarios, hacerse nuevas preguntas, plantear discusiones e imaginar paradigmas.

Finalmente, en cuanto al futuro de la IA, es de inmensa importancia recalcar que la falta de atención a preguntas como las que se plantearon aquí le es indiferente a la rapidez con la que los avances en el campo toman lugar. Hoy puede que la vasta mayoría de decisiones sea aún

tomada por humanos, pero por cómo se plantea la tendencia, mañana la IA será el agente principal en la toma de decisiones elementales: transacciones automáticas en la bolsa, control vehicular, diagnósticos médicos [29] e incluso procesos democráticos; no sería la primera vez que tecnologías tienen influencia en ellos, como lo ocurrido en las elecciones estadounidenses en 2016 [30].

El poder que deriva del rol clave que jugarán estas tecnologías exige -y este punto debe ser enfatizado- que la IA sea democratizada a través de su entendimiento de la mano de iniciativas como la de “IA explicable” o “XAI” en inglés, o los marcos definidos por la ISO/IEC descritos anteriormente. Se señala así a estos campos (el de la XAI y los estándares de la ISO/IEC como promotores de la transparencia dentro de la IA [29]) como futuro objetos de estudio para investigaciones que pretendan -como esta- describir los procesos y las consecuencias del formato de caja negra con la que hoy se nos presenta la inteligencia artificial.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] A. Hertzmann, “Can computers create art?,” *Arts*, vol. 7, no. 2, p. 18, 2018.
- [2] Y. Haik, *Engineering design process*. Pacific Grove, CA: Cengage Learning, 2018.
- [3] Tate, “Generative art,” *Tate*. [Online]. Available: <https://www.tate.org.uk/art/art-terms/g/generative-art>. [Accessed: 12-Dec-2022].
- [4] C. Paul, “Introduction,” *A Companion to Digital Art*, pp. 1–19, 2016.
- [5] Z. Epstein, S. Levine, D. G. Rand, and I. Rahwan, “Who gets credit for AI-generated art?” *iScience*, vol. 23, no. 9, p. 101515, 2020.
- [6] “Is artificial intelligence set to become art's next medium?: Christie's,” *The first piece of AI-generated art to come to auction / Christie's*, 12-Dec-2018. [Online]. Available: <https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>. [Accessed: 12-Dec-2022].
- [7] N. Tiku, “The Google Engineer who thinks the company's AI has come to life,” *The Washington Post*, 17-Jun-2022. [Online]. Available: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>. [Accessed: 12-Dec-2022].
- [8] A. H. Doan, A. Halevy, and Z. Ives, *Principles of Data Integration*. Saint Louis: Elsevier Science, 2014.
- [9] K. P. Murphy, *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press, 2021.
- [10] “Introduction | machine learning | google developers,” *Google*. [Online]. Available: <https://developers.google.com/machine-learning/gan>. [Accessed: 12-Dec-2022].

- [11] J. Brownlee, “A gentle introduction to generative adversarial networks (Gans),” *MachineLearningMastery.com*, 19-Jul-2019. [Online]. Available: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>. [Accessed: 12-Dec-2022].
- [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, “adaptive computation and machine learning series,” in *Deep learning*, Cambridge, MA: The MIT Press, 2017.
- [14] N. Ketkar, *Deep learning with python a hands-on introduction*. New York: APRESS, 2017.
- [15] *Overview of trustworthiness in artificial intelligence*, ISO/IEC TR 24028:2020, 2020.
- [16] P. Esser, R. Rombach, and B. Ommer, “Taming transformers for high-resolution image synthesis,” *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [17] A. Dertat, “Applied deep learning - part 3: Autoencoders,” *Medium*, 08-Oct-2017. [Online]. Available: <https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798>. [Accessed: 12-Dec-2022].
- [18] E. Tiu, “Understanding latent space in machine learning,” *Medium*, 04-Feb-2020. [Online]. Available: <https://towardsdatascience.com/understanding-latent-space-in-machine-learning-de5a7c687d8d>. [Accessed: 12-Dec-2022].

- [19] A. Abdelaal, "Autoencoders for Image Reconstruction in python and keras," *Stack Abuse*, 18-Feb-2020. [Online]. Available: <https://stackabuse.com/autoencoders-for-image-reconstruction-in-python-and-keras/>. [Accessed: 12-Dec-2022].
- [20] C. Snell, "Understanding VQ-vae (dall-e explained pt. 1)," *Understanding VQ-VAE (DALL-E Explained Pt. 1) - ML@B Blog*, 09-Feb-2021. [Online]. Available: <https://ml.berkeley.edu/blog/posts/vq-vae/>. [Accessed: 14-Dec-2022].
- [21] L. Tanzi, "VQ-Gan, explained," *Medium*, 02-Apr-2022. [Online]. Available: <https://medium.com/geekculture/vq-gan-explained-4827599b7cf2>. [Accessed: 14-Dec-2022].
- [22] A. van den Oord, O. Vinyals, and koray kavukcuoglu, "Neural Discrete Representation Learning," in *Advances in neural information processing systems: 31st annual conference on neural information process ...*, LA JOLLA: NEURAL INFO PROCESS SYS F, 2017.
- [23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems: 31st annual conference on neural information process ...*, LA JOLLA: NEURAL INFO PROCESS SYS F, 2017.
- [24] Maxime, "What is a Transformer?," *Medium*, 05-Mar-2020. [Online]. Available: <https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04>. [Accessed: 13-Dec-2022].
- [25] D. Marx, "PyTTI-Tools," *PyTTI*, 2021. [Online]. Available: <https://pytti-tools.github.io/pytti-book/intro.html>. [Accessed: 13-Dec-2022].

- [26] E. Cetinic and J. She, “Understanding and creating art with AI: Review and outlook,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 2, pp. 1–22, 2022.
- [27] M. Newberry, “What is Chaos? A Complex Systems scientist explains,” *The Conversation*, 13-Sep-2022. [Online]. Available: <https://theconversation.com/what-is-chaos-a-complex-systems-scientist-explains-169423>. [Accessed: 13-Dec-2022].
- [28] A. Dorin, “Chance and complexity,” *Proceedings of the Virtual Reality International Conference: Laval Virtual*, 2013.
- [29] W. Samek, Montavon Grégoire, A. Vedaldi, L. K. Hansen, and Müller Klaus-Robert, *Explainable AI: Interpreting, explaining and Visualizing Deep Learning*. Cham: Springer International Publishing, 2019.
- [30] S. Detrow, “What did Cambridge Analytica do during the 2016 election?,” *NPR*, 20-Mar-2018. [Online]. Available: <https://www.npr.org/2018/03/20/595338116/what-did-cambridge-analytica-do-during-the-2016-election>. [Accessed: 13-Dec-2022].

**ANEXO A: PORTRAIT OF EDMOND BELAMY**

Nombre: Retrato de Edmond Belamy.

Fuente:

Z. Epstein, S. Levine, D. G. Rand, and I. Rahwan, "Who gets credit for AI-generated art?" *iScience*, vol. 23, no. 9, p. 101515, 2020.

## **ANEXO B: DOCUMENTACIÓN PYTTI-TOOLS Y CÓDIGO USADO**

Documentación: <https://pytti-tools.github.io/pytti-book/intro.html>

Código y metodología: <https://github.com/PedroGuerrero5150/pytti-book>

## **ANEXO C: PROPOTIPOS Y OBRA FINAL**

Prototipos 1 a 5 con su respectiva configuración de PyTTI-Tools y obra final:

[https://drive.google.com/drive/folders/136sV3ZK8svWUpnPF\\_2oMDp8uTV3DzM9Z?usp=sharing](https://drive.google.com/drive/folders/136sV3ZK8svWUpnPF_2oMDp8uTV3DzM9Z?usp=sharing)

**ANEXO D: ARTSTATION ART BY AZAT NURGALEEV**

Link a la obra de Azat Nurgaleev:

<https://www.artstation.com/azatnurgaleev>

**ANEXO E: LINK DESCARGA MODELO ADABINS**

Link de descarga:

<https://drive.google.com/drive/folders/1kKCJrwcxnIGDOjNe3sByYjnmiRfXOyq?usp=sharing>

**ANEXO F: OBRAS DE STEVE REICH, ERWIN DRIESSENS Y MARIA  
VERSTAPPEN, JACKSON POLLOCK**

Pendulum music de Steve Reich:

<https://www.youtube.com/watch?v=fU6qDeJPT-w>

Top-down Bottom-up de Erwin Driessens y María Verstappen:

<https://notnot.home.xs4all.nl/top-down-bottom-up/drip.html>

Jackson Pollock:

<https://www.moma.org/artists/4675>