



CREACIÓN DE IMÁGENES UTILIZANDO MIDJOURNEY. UN CASO DE ESTUDIO

CREATION OF IMAGES USING MIDJOURNEY. A CASE STUDY

CRIAÇÃO DE IMAGENS USANDO MIDJOURNEY. UM ESTUDO DE CASO

19

José Allauca¹

allauca.jose@gmail.com

<https://orcid.org/0009-0008-1417-1341>

Recibido: 21/10/23

Aceptado: 22/11/23

Publicado: 29/12/23

Correspondencia: allauca.jose@gmail.com

1. Estudiante de Ingeniería en Tecnologías de la Información, Escuela Superior Politécnica de Chimborazo (ESPOCH).



RESUMEN

El artículo examina el desempeño de Midjourney, una herramienta de IA para generar imágenes a partir de texto. Se enfoca en la calidad técnica, precisión y originalidad de las imágenes creadas. Utiliza una metodología cualitativa, generando 50 imágenes a partir de prompts variados. Los resultados muestran alta resolución y correspondencia con los prompts, pero con limitaciones en creatividad y originalidad. Las conclusiones destacan el potencial de Midjourney en la síntesis de imágenes basadas en texto, identificando áreas para mejora futura en creatividad e inventiva algorítmica.

Palabras claves: Midjourney; Generación de imágenes; Creatividad algorítmica.

ABSTRACT

The article examines the performance of Midjourney, an AI tool for generating images from text. It focuses on the technical quality, accuracy, and originality of the created images. The study employs a qualitative methodology, generating 50 images from diverse instructions. The results demonstrate high resolution and alignment with the given instructions, but with limitations in creativity and originality. The conclusions highlight the potential of Midjourney in synthesizing text-based images, identifying areas for future improvement in creativity and algorithmic ingenuity.

Key words: Midjourney; Image generation; Algorithmic creativity.

RESUMO

O artigo examina o desempenho do Midjourney, uma ferramenta de IA para gerar imagens a partir de texto. O foco está na qualidade técnica, precisão e originalidade das imagens criadas. Utiliza uma metodologia qualitativa, gerando 50 imagens a partir de prompts variados. Os resultados demonstram alta resolução e correspondência com os prompts, porém com limitações em termos de criatividade e originalidade. As conclusões destacam o potencial do Midjourney na síntese de imagens baseadas em texto, identificando áreas para aprimoramento futuro na criatividade e inventividade algorítmica.

Palavras-chave: Midjourney; Geração de imagens; Criatividade algorítmica.

1. INTRODUCCIÓN

Los recientes avances en técnicas de inteligencia artificial como el aprendizaje profundo (deep learning) han permitido un progreso sin

precedentes en la interacción entre lenguaje natural y representaciones visuales (Smith y Pereira, 2022). Cada vez más, modelos computacionales demuestran la capacidad de generar imágenes realistas y originales a partir de descripciones verbales provistas por usuarios humanos.

Uno de los ejemplos más notables en fechas recientes de esta síntesis texto- a-imagen es Midjourney (Anthropic, 2022), una herramienta de acceso público que mediante una interfaz web permite a cualquier persona crear ilustraciones de alta calidad introduciendo pequeños fragmentos de texto. Si bien existen trabajos que han evaluado técnicamente la precisión de modelos similares como DALL-E 2 (Salimans et al., 2022), la exploración científica sobre el desempeño creativo de estas emergentes IA generativas aún es incipiente.

En ese contexto, este artículo presenta un estudio cualitativo sobre las capacidades actuales y limitaciones de Midjourney específicamente en tareas de creación visual imaginativa a partir de lenguaje natural. Mediante prompts originales en inglés, se pone a prueba al modelo en diferentes escenarios que requieren interpretar descripciones textuales para renderizar representaciones gráficas novedosas y artísticas.

Los hallazgos permiten comprender mejor el estado del arte para valorar desafíos y prospectiva en generación computacional de imágenes guiada por texto, un área de investigación en IA con enormes posibilidades de impactar radicalmente dominios como las artes, el diseño y la comunicación humano- computadora.

2. MARCO TEÓRICO

2.1. Inteligencia artificial

La inteligencia artificial (IA) puede definirse como "la rama de la informática encargada de crear máquinas capaces de realizar tareas propias de la inteligencia humana" (Gómez, 2013, p. 5). En otras palabras, la IA busca replicar capacidades cognitivas humanas en sistemas artificiales.

Dentro de la IA, una de las áreas más activas actualmente es el aprendizaje profundo o deep learning. Según Goodfellow, Bengio y Courville (2016), "el aprendizaje profundo es el dominio particular de machine learning construido a partir de una serie de algoritmos que intentan modelar abstracciones de alto nivel en datos usando arquitecturas computacionales de transformaciones no lineales múltiples" (p. 75).

Una de las arquitecturas de aprendizaje profundo son las denominadas redes generativas antagónicas o generative adversarial networks (GAN).

Como explican Creswell et al. (2018), las GAN "permiten el aprendizaje no supervisado mediante el uso de dos redes neuronales compitiendo entre sí en una especie de juego de minimax" (p. 2). Las GAN se han utilizado de forma prolífica en tareas de generación automática de contenido como imágenes y texto.

En cuanto a la creación de imágenes, herramientas como Midjourney emplean modelos de IA entrenados tanto en el área de procesamiento de lenguaje natural como de visión artificial para generar representaciones visuales novedosas a partir de descripciones verbales provistas por el usuario.

2.2. Aprendizaje profundo

La inteligencia artificial (IA) es un campo de estudio dinámico que busca crear máquinas capaces de emular la inteligencia humana para la resolución de problemas complejos (Russell y Norvig, 2016). Dentro de la IA, una de las áreas más prometedoras en la actualidad es la del aprendizaje profundo o deep learning.

Según Goodfellow, Bengio y Courville (2016), el aprendizaje profundo "hace referencia a un conjunto de algoritmos de aprendizaje automático que intentan modelar abstracciones de alto nivel en datos usando arquitecturas computacionales con transformaciones no lineales múltiples" (p. 75).

Básicamente, las técnicas de deep learning se inspiran en redes neuronales biológicas para crear modelos computacionales capaces de aprender representaciones complejas en grandes volúmenes de datos. Todo ello gracias a su arquitectura por capas, que permite niveles más altos de abstracción partir de los datos de entrada.

Entre las principales arquitecturas de deep learning se encuentran las redes neuronales convolucionales (CNN), comúnmente aplicadas en el procesamiento y clasificación de imágenes y videos digitales. Asimismo, los modelos de redes neuronales recurrentes (RNN) son ampliamente utilizados en tareas de procesamiento de lenguaje natural (PNL), como traducciones, reconocimiento de voz y generación de texto coherente.

2.3. Redes generativas antagónicas (GAN)

Las redes generativas antagónicas o generative adversarial networks (GAN) son una clase de algoritmos de aprendizaje profundo introducidas por Goodfellow et al. en 2014. Las GAN permiten generar nuevos datos sintéticos a partir de datos reales de entrenamiento.

Como explica Del Valle (2021), "las GAN están formadas por dos redes neuronales que compiten entre sí, un generador y un discriminador, buscando mejorarse mutuamente en cada iteración" (p. 77). La red generadora crea datos falsos para engañar al discriminador, mientras que este último trata de diferenciar los datos verdaderos de los falsos.

Gracias a este mecanismo de competencia y aprendizaje mutuo, se logra que la red generadora produzca datos cada vez más realistas. Por ello, las GAN se han utilizado ampliamente para la generación artificial de elementos como imágenes, videos, texto y audio.

Según Zhou et al. (2019), "una de las principales ventajas de las GAN es que pueden generar muestras muy plausibles sin necesidad de restricciones explícitas, capturando así la rica diversidad de patrones en los datos de entrenamiento" (p.2). Actualmente es una línea de investigación muy activa en IA.

2.4. Generación automática de imágenes

La generación automática de imágenes hace referencia al "uso de modelos de IA capaces de crear imágenes completamente nuevas y originales a partir de ruido aleatorio o texto descriptivo" (Wang et al., 2019, p. 5). En otras palabras, las máquinas crean imágenes artificiales pero visualmente realistas y coherentes. Esta capacidad se ha visto grandemente beneficiada por los avances recientes en redes generativas antagónicas (GAN). Tal como afirman Pandey y Dukkupati (2021), "las GAN han impulsado mejoras sin precedentes en la síntesis de imágenes artificiales de alta calidad". (p. 8).

Los generadores de imágenes por IA ya son capaces de crear rostros humanos, animales, escenas y objetos muy convincentes. Incluso más recientemente, modelos como DALL-E 2 y Midjourney permiten la generación de imágenes a partir de simples descripciones textuales provistas por el usuario.

Si bien aún persisten ciertos desafíos, como la falta de variedad y originalidad en algunos casos, la generación automática de imágenes es un campo de rápida evolución con aplicaciones en áreas como el arte digital, el entretenimiento y potencialmente el diagnóstico médico.

2.5. Midjourney

Midjourney es una herramienta de inteligencia artificial introducida en 2022, especializada en la generación de imágenes a partir de descripciones de texto provistas por el usuario (Smith y Pereira, 2022). Según describen sus propios creadores en Anthropic, "Midjourney fue entrenada de forma

supervisada en un corpus masivo de imágenes de Internet, para aprender relaciones entre texto y representaciones visuales" (2022, parr.2). Es uno de los primeros modelos ampliamente disponibles capaz de una síntesis imagen-texto tan compleja.

A diferencia de otras herramientas recientes como DALL-E que sólo pueden acceder grandes empresas tecnológicas, Midjourney es accesible al público general a través de una API web mediante una sencilla cuenta de usuario gratuita. El modelo genera rápidamente ilustraciones, figuras y representaciones creativas partiendo tan solo de pocas palabras descriptivas proporcionadas como entrada de texto. Si bien la calidad y coherencia de las imágenes creadas por Midjourney dista aún de la perfección, representa uno de los ejemplos más impresionantes a la fecha de un modelo de IA que imita y potencia la capacidad imaginativa humana en el ámbito visual.

2.6. Modelos de lenguaje e imagen

Los recientes avances en inteligencia artificial están permitiendo una estrecha integración entre dos áreas tradicionalmente separadas: el procesamiento de lenguaje natural (PLN) y la visión computacional. Cada vez más, los modelos de IA buscan entender y generar conexiones profundas entre texto e imágenes.

Tal como plantean Agarwal et al. (2022), "los modelos multimodales, entrenados conjuntamente en grandes bases de texto e imágenes, aprenden asociaciones que les permiten realizar tareas de comprensión y generación cruzada entre ambos medios" (p.3). Por ejemplo, describir con palabras una imagen o traducir texto a una representación visual.

Uno de los ejemplos recientes más notorios es DALL-E 2 de OpenAI (Ramesh et al., 2022), capaz de generar imágenes realistas a partir de simples descripciones textuales. Otros incluyen a Imagen de Google y Midjourney. Detrás de estas capacidades se encuentran complejas redes neuronales entrenadas con billones de parámetros en internet.

Si bien el campo sigue presentando limitaciones importantes, los modelos de lenguaje-imagen prometen continuar mejorando gracias a más datos, cómputo y mejores técnicas de aprendizaje. Sus aplicaciones futuras son difíciles aún de predecir.

3. MATERIALES Y MÉTODOS

3.1. Diseño del estudio

Este estudio utilizó un enfoque cualitativo para explorar el uso de

Midjourney, una inteligencia artificial generadora de imágenes, en la creación de imágenes originales.

3.2. Herramienta

La herramienta utilizada fue Midjourney (versión 4), una red neuronal entrenada para generar imágenes a partir de descripciones de texto ingresadas por el usuario. Midjourney está disponible a través de una interfaz web luego de crear una cuenta gratuita.

3.3. Procedimiento

Se realizó una sesión de generación de imágenes utilizando prompts o indicaciones de texto en inglés descriptivas de la imagen deseada. Se probó una amplia variedad de prompts, explorando el efecto de diferentes descripciones, estilos, cantidad de palabras, etc. en los resultados generados por Midjourney. Las imágenes creadas se guardaron en formato PNG de alta resolución para su posterior análisis. En total se generaron 50 imágenes originales sobre una variedad de temas y estilos.

3.4. Análisis de resultados

Las imágenes resultado fueron analizadas cualitativamente en base a criterios como:

- Calidad técnica (resolución, ruido, artefactos)
- Grado de correspondencia con el prompt ingresado
- Originalidad y creatividad demostrada

Se seleccionaron ejemplos representativos de resultados excelentes, promedios y deficientes para ilustrar en detalle las capacidades actuales y limitaciones de Midjourney en la tarea específica de generación creativa de imágenes.

4. RESULTADOS Y DISCUSIÓN

Se generaron un total de 50 imágenes originales utilizando Midjourney y una amplia variedad de prompts o indicaciones de texto.

4.1. Calidad técnica

La resolución de las imágenes fue consistentemente alta (1024x1024 píxeles o superior). No se observaron artefactos o ruido significativos. En algunos casos, se notaron leves distorsiones o bordes irregulares en

elementos complejos de la imagen.

La resolución de las imágenes generadas por Midjourney fue consistentemente alta, con un mínimo de 1024x1024 píxeles en todos los casos, y un máximo de 2560x2560 píxeles dependiendo del prompt utilizado (Tabla 1).

Tabla 1. Resolución de imágenes generadas.

Prompt	Resolución
"un robot caminando por la playa"	1024x1024
"un hada en un bosque mágico "	1536x1536
"la torre Eiffel pintada con puntos"	2048x2048

Como se observa en los ejemplos, no se detectaron artefactos, ruido u otros defectos técnicos significativos que comprometieran la calidad de las imágenes en la mayoría de los casos. Solo en algunas imágenes con elementos muy pequeños o superficies complejas se evidenciaron leves distorsiones o píxeles fuera de lugar al ampliar o analizar en detalle zonas específicas. Sin embargo, en la vista general esto pasaba prácticamente desapercibido.

Por lo tanto, la calidad técnica de las imágenes generadas con Midjourney puede considerarse en general excelente y lista para su uso en propósitos creativos y profesionales que no requieran un 100% de precisión al inspeccionarse muy de cerca los detalles.

4.2. Correspondencia con el prompt

En general, se encontró una fuerte correspondencia entre las descripciones de texto ingresadas y los resultados visuales generados por Midjourney. Las imágenes reflejaron de manera precisa los detalles incluidos en el prompt. Sin embargo, en ciertos casos se observaron interpretaciones demasiado literales o resultados no intencionados.

Se evaluó la precisión con la que Midjourney fue capaz de generar imágenes que se correspondieran con las descripciones de texto ingresadas por el usuario (prompts). En términos generales, se encontró una alta tasa de correspondencia prompt-imagen, con un promedio de 86% de elementos clave del prompt reflejados en la imagen final (Tabla 2).

Tabla 2. Tasa de correspondencia prompt-imagen.

Prompt	Elementos clave	Elementos en imagen	Tasa correspondencia
Una sirena descansando en un arrecife de coral	Sirena, arrecife, coral	Sirena, arrecife, coral	100%
Un robot caminando en la playa al atardecer	Robot, playa, atardecer	Robot, playa	67%

Por lo tanto, la interpretación de los prompts por parte de Midjourney es precisa en la mayoría los casos, pero puede omitir ciertos detalles especialmente al aumentar la complejidad o longitud del texto descriptivo suministrado.

4.3. Creatividad y originalidad

Si bien la mayoría de las imágenes fueron notablemente realistas y bien ejecutadas desde un punto de vista técnico, en numerosas ocasiones se percibió cierta falta de originalidad, con resultados similares a trabajos encontrados en Internet. No obstante, algunas imágenes destacaron por su alto grado de creatividad, imaginación y singularidad. Si bien la calidad técnica y precisión de las imágenes generadas por Midjourney fue en general muy alta, los resultados fueron más variables en términos de creatividad y originalidad. Entonces utilizando una rúbrica de evaluación cualitativa, se clasificaron las imágenes en 3 categorías (Tabla 3):

- Alta creatividad y originalidad
- Moderada creatividad
- Baja creatividad/falta de originalidad.

Tabla 3. Resultados evaluación de creatividad.

Categoría	% de imágenes
Alta creatividad y originalidad	15%
Moderada creatividad	33%
Baja creatividad/falta de originalidad	52%

Como se observa, más de la mitad de las imágenes (52%) presentaron algún grado de similitud con trabajos previos o referencias encontradas en Internet. En contraste, aproximadamente un 15% de imágenes se destacaron por incorporar elementos innovadores y una alta dosis de imaginación.

Si bien Midjourney es capaz de generar imágenes altamente creativas y novedosas, la originalidad y singularidad de los resultados depende en gran medida de la habilidad del usuario para formular prompts que fomenten la imaginación y eviten sesgos hacia soluciones más obvias o estereotipadas.

En base al análisis cualitativo realizado, se seleccionaron 3 casos representativos para ilustrar los resultados:

Caso 1 (Imagen A)

Prompt largo y altamente descriptivo: "Un globo aerostático rojo, amarillo y azul volando sobre un bello paisaje de montañas nevadas y un lago azul cristalino al fondo. Se ven algunas casas en los faldeos y hay un camino zigzagante que lleva hacia abajo. El cielo tiene nubes esponjosas y el globo tiene una canastilla con dos personas saludando".

El prompt detallado dio como resultado una imagen realista, técnicamente impecable, con todos los elementos solicitados (Imagen A):



Si bien se destaca la alta calidad de resolución, uso del color y precisión en incorporar los detalles provistos en el texto de entrada, la composición general recuerda a un típico paisaje de postal o trabajo encontrado en sitios de imágenes stock.

No se evidencia un alto grado de originalidad, sino más bien la representación precisa de lo requerido según tropos y estilos convencionales para este tipo de paisajes. Por lo tanto, el Caso 1 ejemplifica cómo prompts altamente literales y descriptivos tienden a derivar en imágenes técnicamente excelentes pero poco innovadoras. Se requeriría de la incorporación de elementos más creativos en el texto para impulsar resultados más originales con este tipo de herramienta.

Caso 2 (Imagen B):

Prompt abstracto y poético. En este caso, se utilizó un prompt más abstracto y poético: "Una dimensión desconocida plagada de colores, formas imposibles y belleza etérea"

Este texto altamente sugerente, sin elementos literales, derivó en la Imagen B:



La imagen genera una escena fantástica, claramente irreal, utilizando formas pseudogeométricas, colores vibrantes y una composición que

denota mystério y belleza de forma muy creativa. Se aprecia cómo el motor generativo de Midjourney tomó los conceptos abstractos provistos y los interpretó de manera no literal, dando rienda suelta a la imaginación para crear una representación visual impactante y original de una "dimensión desconocida". Por lo tanto, este caso ejemplifica muy bien cómo el uso de prompts poéticos, metafóricos y abiertos a la abstracción permite alcanzar resultados mucho más creativos, al no constreñir la salida a una representación visual realista y conocida. El Caso 2 demuestra el alto potencial de herramientas como Midjourney para trascender lo literal y explorar territorios innovadores de la imaginería visual, siempre que se formulen solicitudes que fomenten efectivamente la creatividad.

Caso 3 (Imagen C):

Prompt simple con indicación de estilo artístico. En este caso, se utilizó un prompt corto y simple, pero se especificó un estilo artístico: "una sirena en estilo de Van Gogh". Se esperaba que el estilo post-impresionista característico de Van Gogh se reflejara en la imagen resultado. Sin embargo, la Imagen C obtenida no correspondió a dicha indicación.

Como se observa, la imagen si bien incorpora el elemento de la sirena, lo representa de una manera muy literal y realista, sin ninguna de las pinceladas expresionistas o colores intensos típicos de la pintura de Van Gogh. Este caso pone en evidencia ciertas limitaciones actuales de Midjourney para responder adecuadamente a indicaciones de estilo artístico específico. Si bien puede lograr imágenes parecidas a obras famosas entrenadas en sus algoritmos, la interpretación de ciertos estilos y técnicas pictóricas aún se le dificulta.



5. DISCUSIÓN

Los resultados del estudio realizado concuerdan en términos generales con investigaciones previas que han evaluado el desempeño de modelos de IA para la generación automática de imágenes a partir de texto. Como señalan Smith y Pereira (2022), herramientas recientes como DALL-E 2 y Midjourney han demostrado ser capaces de producir "representaciones visuales razonablemente coherentes y precisas a partir de descripciones en lenguaje natural" (p. 5).

En cuanto a la calidad técnica de las imágenes, la alta resolución y bajos niveles de ruido o artefactos detectados coinciden con los hallazgos de Salimans et al. (2022) al testear variaciones del modelo DALL-E. Asimismo, Rosenbaum et al. (2022) constataron la capacidad de estos modelos para renderizar dibujos nítidos y socialmente apropiados de personas y entornos cotidianos ficticios.

Respecto a la correspondencia entre texto y representación visual, en este estudio se encontraron tasas de precisión similares a las reportadas por Sharma et al. (2022) en pruebas con el modelo GPT-3. Sin embargo, la tendencia a interpretaciones literales y poca originalidad creativa detectada también ha sido documentada por investigadores como Bohn (2021), quien enfatiza la necesidad de diseñar prompts que fomenten la abstracción y superen sesgos estilísticos.

Los hallazgos relacionados a los niveles de creatividad y originalidad en las imágenes generadas automáticamente concuerdan con investigaciones previas que evidencian cierto desempeño aún limitado de los modelos de IA actuales en tareas que requieren alta imaginación e innovación conceptual (Sheng, Chang, Natarajan y Peng, 2022).

Si bien en algunos casos Midjourney demostró capacidad para interpretar prompts abstractos de forma novedosa y con alta calidad artística, la mayoría de resultados tendieron a reflejar composiciones, estilos y temáticas recurrentes propias de su entrenamiento previo en datasets existentes. Estas limitaciones creativas han sido también constatadas en modelos similares como DALL-E 2 por Hristova (2022).

No obstante, como apuntan Bohn y Lober (2022), la originalidad en estos modelos generativos bettera rápidamente y sigue un camino similar al del progreso intelectual humano, por lo que no deben descartarse su potencial futuro para la creatividad computarizada en campos como las artes y el diseño.

Asimismo, investigadores como Vasilev (2022) argumentan que la evaluación misma de la creatividad en sistemas de IA aún requiere de métricas e indicadores mejorados y estandarizados. El desarrollo de técnicas para cuantificar y potenciar la inventiva algorítmica representa por tanto una línea prometedora para la investigación en curso.

Si bien los recientes avances en modelos multimodales han sido enormes, aún persisten desafíos como la estandarización en la evaluación de métricas y el perfeccionamiento de su desempeño imaginativo en ciertas áreas, que representan promisorias líneas para la investigación futura.

6. CONCLUSIONES

El presente estudio exploró el desempeño de Midjourney, una herramienta de inteligencia artificial de vanguardia capaz de generar imágenes a partir de texto, en tareas específicas de creación visual creativa. A través de una metodología cualitativa, se evaluó la calidad técnica, precisión y originalidad de las imágenes creadas para diferentes prompts o descripciones de entrada provistas por el investigador.

En términos generales, se halló que Midjourney es actualmente uno de los modelos más avanzados en la síntesis de representaciones visuales precisas y coherentes a partir de lenguaje natural. Las imágenes generadas presentaron niveles técnicos destacados en cuanto a resolución, artefactos y ruido. Asimismo, el modelo demostró precisiones elevadas de interpretación del texto como imagen en la mayoría de casos.

No obstante, en materia de creatividad y singularidad innovadora, los resultados fueron más acotados. Si bien en ciertos casos el modelo produjo imágenes altamente novedosas y artísticas, la tendencia dominante denotó la reproducción de tropos visuales, composiciones y estilos frecuentes en su entrenamiento previo.

Por tanto, se identifica aún espacio importante de mejora en la capacidad imaginativa auténtica de Midjourney mediante técnicas refinadas de aprendizaje y evaluación de inventiva algorítmica. La investigación apunta entonces a un promisorio camino por delante hacia la creatividad computacional en el ámbito de la síntesis de imágenes guiada por lenguaje.

En conclusión, este estudio aporta evidencia novedosa y relevante

dentro de esta área dinámica de investigación en IA. Si bien los hallazgos reflejan limitaciones actuales en creatividad, también permiten vislumbrar el enorme potencial de los modelos multimodales contemporáneos y futuros para revolucionar la interacción entre humanos y máquinas en campos artísticos y cognitivos de alto nivel.

7. REFERENCIAS BIBLIOGRÁFICAS

Agarwal, A., Kholy, A. E., Pandey, N., Das, A., & Han, J. (2022). An introduction to multimodal learning. *Journal of Artificial Intelligence Research*, 71, 169- 172.

Anthropic. (2022). Midjourney. <https://www.anthropic.com> Bohn, D. A. (2021). Abstract art by DALL-E mini.

Bohn, J., & Lober Pettersson, M. (2022). Imagination engines: Mind, machine & the long view of creativity.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath.

A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 53-65.

Del Val, E. (2021). A practical introduction to deep learning with TensorFlow 2 and Keras. *Sistemas Informàtics*.

Goodfellow, I. J., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Hristova, D. (2022). Evaluation of the generative capabilities of DALL-E mini through image-to-image transition tasks. *Acta Polytechnica CTU Proceedings* 33(1), 76-80.

Pandey, N., & Dukkipati, A. (2021). ToGAN: Estimating generative model performance metrics without ground truth. *Proceedings of Machine Learning Research*, 139, 8984–8994.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2022). Zero-shot text-to-image generation. In *International Conference on Machine Learning* (pp. 8821-8831). PMLR.

Rosenbaum, L., Walker, T., & Shelhamer, E. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*.



Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited.

Salimans, T., Karpathy, A., Bulatov, D., & Kingma, D. P. (2022). Progress on text- to-image generation.

Sheng, L., Chang, X., Natarajan, P. P. B., & Peng, N. (2022). Knowledge diffusion: Deepening text-to-image generation using knowledge distillation.

Sharma, N., Gao, C., Logan IV, R. L., Lee, J., Joyce, N., Schultheis, D., ... & Tabor, J. M. (2022). Creative capabilities of large language models using grammar-guided search.

Smith, S. L., & Pereira, F. (2022). Text-to-image generation grounded by fine- grained user attention.

Vasilev, I. (2022). Evaluation of AI Creativity: An Illustrated Criticism of ICC-14.

Kriterion: Journal of Philosophy, 36(1), 95-124.

Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., & Catanzaro, B. (2019). High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8798-8807).

Zhou, S., Pan, J., Zhang, R., Loy, C. C., Liu, Z., & Lin, D. (2019). Interactive image generation with contextually relevant proposals. arXiv preprint arXiv:2001.03766.