ISSN 2529-9824

Artículo de Investigación

Creación de ambientes para niveles de videojuegos utilizando Stable Diffusion

Optimizing video game level atmosphere with Stable Diffusion-assisted set dressing

Pedro Meira-Rodríguez: Universidade da Coruña, España.
pedro.meira.rodriguez@udc.es

Fecha de Recepción: 11/06/2024

Fecha de Aceptación: 03/09/2024

Fecha de Publicación: 27/01/2025

Cómo citar el artículo:

Meira-Rodríguez, P. (2025). Creación de ambientes para niveles de videojuegos utilizando Stable Diffusion [Optimizing video game level atmosphere with Stable Diffusion-assisted set dressing]. *European Public & Social Innovation Review*, 10, 1-18. https://doi.org/10.31637/epsir-2025-1354

Resumen:
Introducción: Este estudio explora la optimización de tiempos en desarrollo de videojuegos, evaluando la capacidad del modelo de Stable Diffusion para apoyar en la ambientación de niveles o *set dressing*. **Metodología**: Para ello se realiza un estudio comparativo que parte de tres imágenes del mismo entorno tridimensional, realizando solicitudes iterativas en Stable Diffusion XL con las extensiones de ControlNet, ConfyUI y un LoRA para unificar el estilo. **Resultados:** La evaluación de la creatividad formal, el apego al diseño espacial y la definición de detalles permitió obtener puntuaciones medias por indicación. **Discusión**: Se evidenció que el uso de variables y pesos específicos genera imágenes con cualidades predecibles, confirmando la correlación directa entre parámetros de entrada y resultados. **Conclusiones**: Se destacó la necesidad de controlar con precisión estas variables y se subrayó la utilidad de las herramientas de IA, aunque se recomienda definir una ruta de uso para su integración efectiva en la producción de videojuegos.

Palabras clave: ambientación de niveles; entornos de videojuegos; diseño de niveles; Stable Difusión; ControlNet; inteligencia artificial; modelos de difusión; arte digital.

Abstract:
Introduction: This study investigates the optimization of time in video game development by evaluating the capability of the Stable Diffusion model to assist in set dressing. **Methodology**: A comparative study was conducted using three images of the same three-dimensional environment, performing iterative requests in Stable Diffusion XL with ControlNet and ConfyUI extensions, and a LoRA to unify the style. **Results:** The evaluation considered formal creativity, adherence to spatial design, and detail definition, resulting in average scores per indication. **Discussions:** The study demonstrated that the use of specific variables and weights produces images with predictable qualities, confirming a direct correlation between input parameters and outcomes. **Conclusions**: The importance of precisely controlling these variables was highlighted, and the utility of AI tools was emphasized. However, it is recommended to define a clear usage pathway for their effective integration into video game production.

Keywords: setdressing; video game environments; level design; Stable Diffusion; ControlNet; artificial intelligence; diffusion models; digital art.

1. Introducción

La industria de los videojuegos destaca por ser una de las más multidisciplinares al combinar el trabajo diseñadores, artistas y programadores (Wade, 2007). La accesibilidad de diversos softwares de desarrollo ha permitido al público general crear y lanzar nuevos productos de entretenimiento con mayor frecuencia. No obstante, estos factores han incrementado la competitividad, afectando tanto en los tiempos de producción como la calidad de los resultados. La necesidad de acelerar la producción ha llevado al fenómeno conocido como "crunch" (Bulut, 2023), que en muchos casos implica semanas o incluso meses de trabajo intensivo. A pesar de los esfuerzos regulatorios, la optimización del tiempo de producción sigue siendo un desafío durante todas las fases del desarrollo (Ahmad *et al*., 2017; Torres-Ferreyros *et al*., 2017).

En este escenario, surge la pregunta de qué papel podrían desempeñar las nuevas herramientas de generación visual desarrolladas con modelos de difusión, que ya se han utilizado en campos como el diseño artístico y espacial tales como la arquitectura (del Campo, 2021; del Campo *et al*., 2020; del Campo y Leach, 2022; Lorenzo-Eiroa y Sprecher, 2013) o el cine (Song y Yip, 2023). Aunque se han compartido algunos avances en la creación digital de entornos y elementos en tres dimensiones en redes como LinkedIn o Instagram (Eisendorf, 2024; Fraunberger, 2023; Seleit, 2024), estos se han llevado a cabo con fines exploratorios más que con la intención de realizar un estudio analítico.

2. Objetivos

Esta investigación se centra en la optimización de los tiempos de desarrollo en la industria de los videojuegos, específicamente en la ambientación de niveles o "*set dressing*". El objetivo principal es investigar cómo las inteligencias artificiales generativas de imágenes pueden asistir en la visualización y toma de decisiones durante la creación de un videojuego. Aunque desde antes de 2018 se utilizaban extensiones de inteligencia artificial para agilizar la creación de escenarios (Williams y Wuetherick, 2018), la capacidad de previsualizar casi instantáneamente un posible resultado es una ventaja que los modelos como *Stable Diffusion* pueden ofrecer (Petráková y Šimkovič, 2023). Este estudio, se evalúa la capacidad del modelo de difusión TTI (de texto a imagen) e ITI (de imagen a texto) para transformar fotografías de *blockout* de niveles en propuestas gráficas de "*set dressing*" basadas en indicaciones específicas.

3. Marco teórico

En una aproximación al tema investigado desde el campo de la arquitectura, Fernández Álvarez y López Chao (2023) destacaron la importancia de dominar las herramientas de IA para controlar y dirigir los resultados visuales. En la misma línea, Petráková y Šimkovič (2023) realizaron un trabajo exploratorio en el que compararon como diferentes softwares generativos de texto a imagen procesaban indicaciones de entrada con el fin de observar las posibles bondades y carencias de estas IAs.

Limitado a la iteración de la indicación textual, Chang *et al*. (2023) intentaron definir qué expresiones o términos parecían dar como resultado una serie de imágenes más creativas. No obstante, no encontraron una fórmula suficientemente consistente más allá de concluir que el uso de palabras más sugerentes y descripciones precisas parecían conducir a imágenes menos sesgadas.

Estas investigaciones en las que se abordó el posible control útil de las herramientas de IA podría resultar de especial relevancia en el marco de la industria de los videojuegos, donde la ambientación de niveles es un aspecto fundamental tanto para la narrativa como para la jugabilidad, consumiendo una parte significativa del tiempo de producción (Game Developers Conference, 2022; Grepl-Malmgren y Hallenbom, 2023).

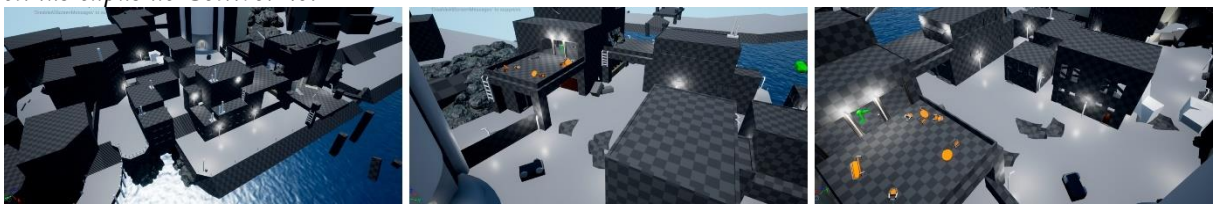
Directamente relacionado con la necesidad de controlar el resultado con el fin generar material útil que acelere la producción de un modo menos demandante para los trabajadores, ha de mencionarse el problema del sesgo del instrumento, presente en mayor o menor medida en los mencionados estudios. Al contar los modelos de difusión como *Midjourney*, *Stable Diffusión* o *Dall-E* con unas bases de datos generales influenciadas por el conocimiento colectivo y no específico, podría incurrirse en resultados menos ricos y precisos (López-Chao *et al*., 2023). A fin de paliar este efecto, y en pos de guiar la producción del material audiovisual, se valoró especialmente la posibilidad de seleccionar modelos preentrenados en áreas específicas, destacando el caso de *Stable Diffusion XL* con *Comfy UI*, cuya versatilidad recalcan diferentes profesionales del sector (Nuray, 2024a, 2024b; Žnidarič, 2024).

4. Metodología

Este estudio se ha diseñado para evaluar las capacidades y limitaciones de las herramientas de generación de imágenes mediante una metodología en dos partes: seguimiento de los parámetros implicados y evaluación de la eficiencia de las respuestas obtenidas.

Figura 1.

Capturas de un nivel creado en Unreal Engine que se emplearon como referencia para el procesamiento en las capas de ControlNet



Fuente: Elaboración propia (2024).

2.1. Proceso de generación

Para minimizar sesgos, se generaron imágenes de tres vistas diferentes de un mismo entorno urbano a nivel de *blockout* en Unreal Engine. Se probaron veinticuatro combinaciones de parámetros en *Stable Diffusion XL*, utilizando la extansión *ComfyUI*, el controlador multicapa *ControlNet* y un modelo auxiliar LoRA preentrenado, produciendo un total de 720 imágenes. Esta estrategia de combinación aditiva de variables en cada indicación se testeó previamente en otro estudio exploratorio que mostraba que las capas e *ControlNet* permitían un mayor control sobre el resultado del que se conseguía con el uso aislado de las indicaciones textuales y los *checkpoints* (Meira-Rodríguez y López-Chao, 2024).

En el proceso de generación se ha empleado el software *ComfyUI*, una extensión de *Stable Diffusion* que propicia la implicación directa del usuario con los agentes presentes en la generación al ofrecer la posibilidad de seleccionar y organizar los nodos activos (Strossmayera y Fakultet, 2023). La versatilidad de esta aplicación, ya compartida por profesionales de la creación digital en medios de divulgación no científica como *LlikedIn* o *Youtube* (Abdelsalam Soliman, 2024; Musli, 2024; Nuray, 2024b), se hace patente al observar la enorme cantidad de variables que pueden alterarse a la hora de establecer una indicación de entrada. De entre todas ellas, en esta investigación se han iterado solamente algunas de las más destacables, manteniendo los valores del resto en los intervalos de influencia recomendados (Andrew, 2023a; Du *et al*., 2023).

En lo referente a la codependencia de la indicación, las variables que se han empleado podrían clasificarse entre persistentes, aquellas cuyos valores no pueden ser nulos para obtener cualquier respuesta visual, y complementarias, aquellas que no siendo imprescindibles permiten aumentar significativamente el grado de control que se tiene sobre las propuestas generadas.

2.1.1. Variables persistentes

La primera de estas variables configurables el *checkpoint*, el modelo preentrenado en sí mismo que se encarga de decodificar el ruido para generar las imágenes (Andrew, 2024; Petráková y Šimkovič, 2023; Zhang *et al*., 2023). En el caso de *Stable Difusión*, a diferencia de otras IAs generativas como *Dall-E* o *Midjourney*, cabe destacar la ventaja que supone poder elegir con total libertad el modelo a emplear, dado que el afinamiento de estos ayuda a guiar la generación.

Para este estudio se ha seleccionado un *checkpoint* diseñado para SDXL de la plataforma *civitai.com*, una base de modelos en línea donde diferentes profesionales publican sus modelos; *animagineXLV31*, especializado en la creación de imágenes con una estética de animación en 2D con más de ciento cuarenta y tres mil descargas y más de doce mil "me gusta".

Con respecto a la indicación textual, se manutuvo inalterada durante el proceso con la intención de determinar cuál de los otros parámetros influía en mayor medida a la hora de obtener una respuesta óptima a una solicitud en el marco de la ambientación de escenarios para videojuegos. Así pues, para todas las indicaciones se utilizó el prompt:

"A top view of a destroyed city with scattered concrete blocks and broken windows, dimly illuminated by street lamps at night, rendered in a 3D video game style similar to Super Mario, with saturated dark bluens and ocher and whimsical elements amidst the destruction"

Con respecto a las variables incluidas en el *KSampler* (Andrew, 2024; Strossmayera y Fakultet, 2023), solamente se iteró el valor de los pasos de muestreo, parámetro para el que se observaron las salidas al aumentar el índice de veinte a cincuenta. En lo referente a la semilla y a la escala CFG, la primera se mantuvo con un valor aleatorio y a la segunda se le asigno el valor constante de diez dado otros muy superiores o inferiores podrían alterar negativamente a las propuestas (Andrew, 2023b; Nuray, 2024b).

2.1.2. Variables complementarias

Las variables complementarias, por otro lado, son aquellos campos que admiten variaciones en sus valores pero que, en caso de ser estos nulos, no impedirían al modelo proponer una respuesta.

Dentro de ellos, podrían destacarse el *prompt* negativo, que para este estudio se ha integrado mediante la indicación "*low quality details, low resolution, ugly textures*", los LoRA, modelos de menor tamaño que los *checkpoints* con los que ajustar y afinar la generación (Ameneh y Microsoft, 2023; Hu *et al*., 2022) y las los modos de control de la extensión multicapa *ControlNet*.

Como modelo LoRA, también obtenido de la plataforma *civitai.com,* se ha probado el efecto que ha tenido sobre las propuestas la implementación de *Picture illustration*, un modelo con seis mil descargas afinado para la proposición de imágenes en un estilo colorido y de película de animación 2D.

Finalmente, en lo referente a la extensión de *ControlNet*, se ha observado la variación de las imágenes al incorporar de manera alternada tres capas de procesamiento diferentes; la capa *scrible* (entrenada en la interpretación de las líneas generales de una referencia), la *canny* (especializada en el procesamiento del trazado de borde) y la capa *Midas Depth* (encargada de extraer e interpretar la profundidad de una imagen).

2.2. Proceso de evaluación

Una vez obtenidas las setecientas veinte imágenes, cada una de ellas se sometió a un test conformado por tres categorías con el fin de determinar cuál de los componentes presentes en las diferentes indicaciones de entrada resultó más beneficioso a la hora de proponer una ambientación.

2.2.1. Elaboración del test

El test realizado se dividió en tres ítems medidos de uno a cinco (siendo uno la respuesta más baja y cinco la mayor) que se organizaron según el área de estudio.

El primero, *comprensión de la forma*, permitió determinar la capacidad de interpretación de las referencias de entrada según el procesador de *ControlNet* empleado. Con el segundo, *adaptación del estilo*, se estudió la capacidad del modelo de difusión para dotar a la salida de la apariencia visual solicitada y, con la tercera, *integración de la paleta cromática*, se cuantificó la eficiencia de la IA al incorporar los colores solicitados.

2.2.2. Obtención de los resultados

Con el fin de determinar la eficacia de las diferentes variables en la generación visual y propositiva de ambientes para niveles de videojuegos, se realizaron dos tipos de análisis estadísticos con la muestra obtenida.

En primer lugar, se aplicó el análisis de varianza (ANOVA). Éste permitió identificar cuáles de los parámetros afectaban de modo significativo los resultados en las tres categorías evaluadas. Luego, se utilizaron tablas cruzadas con el software SPSS para observar cuáles de los valores de entrada condujeron con mayor consistencia a resultados favorables.

Este segundo estudio complementa los hallazgos identificados del ANOVA porque facilita la identificación de patrones y tendencias.

El planteamiento de análisis es que permitan identificar los parámetros más influyentes, así como establecer recomendaciones claras sobre cómo ajustar las variables para optimizar la calidad de las imágenes generadas en futuros desarrollos de videojuegos.

5. Resultados

Tras realizar el análisis de varianza de las diferentes variables de entrada con respecto a las tres competencias evaluadas pudo determinarse cuales afectaron en mayor medida a los resultados medios obtenidos por cada indicación.

3.1. Análisis de varianza (ANOVA)

El análisis de varianza (ver Tabla 1) evidencia que el uso de una u otra capa de *ControlNet* ha sido la variable más determinante para la evaluación al obtener los valores p más bajos, implicando la falsedad de la hipótesis nula. Seguida de esta, la presencia o ausencia tanto del LoRA como del prompt negativo han afectado directamente a los resultados obtenidos en la evaluación; especialmente en los casos de adaptación del estilo e integración de la paleta cromática con respecto al LoRA y de la comprensión de la forma y la adaptación del estilo en lo referente al uso de indicación textual negativa.

Por otro lado, el incremento de los pasos de muestreo ha tenido una implicación mínima en los resultados obtenidos por las imágenes; sobre todo, en lo que respecta a la integración de la paleta cromática, donde el valor de p se acercó notablemente a uno.

Tabla 1.

Cálculo de la varianza entre grupos mediante ANOVA

Variables de entrada	Comprensión de la forma	Adaptación del estilo	Integración de la paleta cromática
Imagen empleada	0,012	0,03	<0,001
Capa de ControlNet	<0,001	<0,001	<0,001
Pasos de muestreo	0,443	0,331	0,941
Uso de LoRA	0,021	<0,001	<0,001
Uso de prompt negativo	<0,001	<0,001	0,026

Fuente: Elaboración propia al aplicar ANOVA sobre los datos del test realizado (2024).

3.2. Comparación de las variables con tablas cruzadas

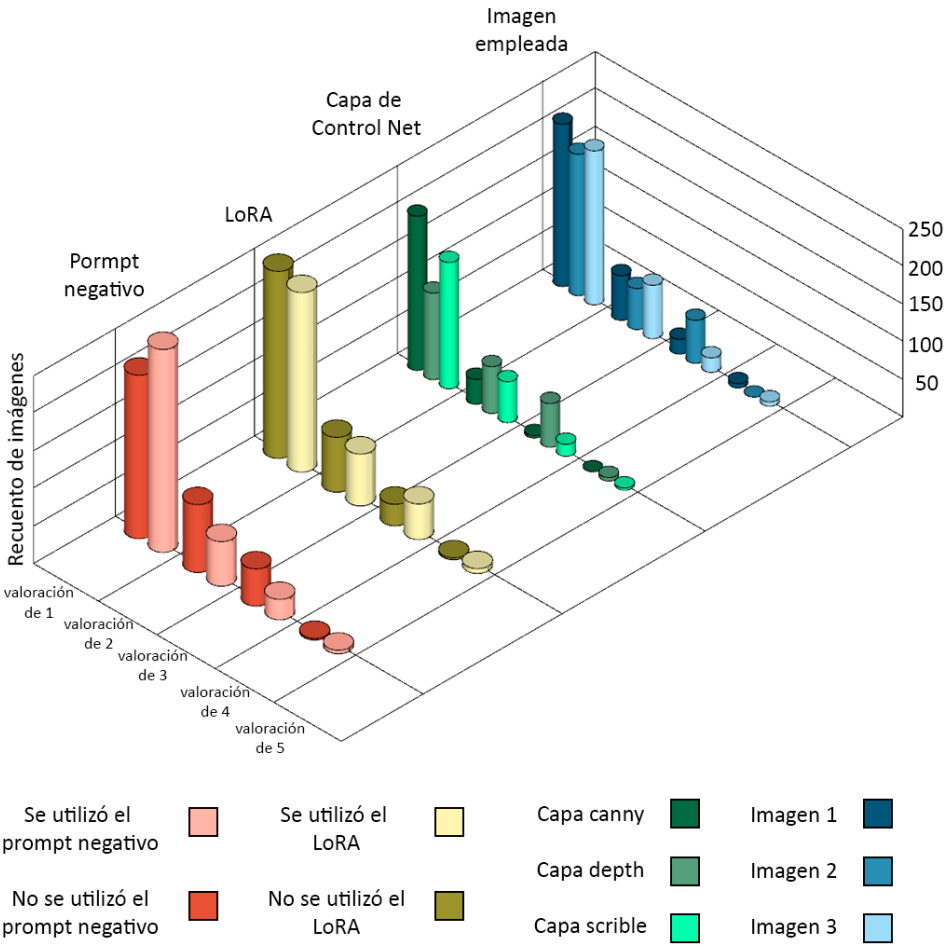
Gracias a la comparación de datos mediante tablas cruzadas, pudieron observarse los resultados obtenidos con más frecuencia para cada una de las competencias definidas en el test.

3.2.1. Comprensión de la forma

De forma general, y como se observa en la figura 2, la inmensa mayoría de las imágenes obtuvieron una puntuación igual o inferior a dos en esta categoría. El sesenta y siete por ciento de las salidas propuestas obtuvieron la valoración mínima y ninguna alcanzó la máxima. No obstante, la distribución de las puntuaciones varió significativamente entre los valores posibles de los parámetros estudiados.

Figura 2.

Gráfica de la distribución de puntuaciones en la competencia de "comprensión de la forma"



Fuente: Elaboración propia (2024).

En el caso de las imágenes de referencia utilizadas para la generación, se evidenció que la media resultante fue mayor al emplear la imagen dos al presentar una proporción ligeramente más equilibrada que la uno y la tres en cuanto a la mayor concentración de respuestas calificadas con un dos o un tres. Mientras que las generaciones en las que se partió de las referencias uno y tres alcanzaron un crecimiento progresivo y más regular (próximo a la relación de un tercio con respecto al término previo), que incluso les permitió llegar a alguna salida valorada con un cinco; la media de puntuación de las indicaciones en las que se incorporó la imagen dos superó al resto en más de 0,17 puntos.

Figura 3.

Comparación de las imágenes generadas con SDXL+ConfyUI variando la imagen de referencia

Fuente: Elaboración propia mediante el software SDXL+ConfyUI con ControlNet mediante el checkpoint y LoRA mencionados en la metodología (2024).

De un modo semejante, se comprobó que el uso de la capa de control *Midas Depth* favoreció a la comprensión de la forma de referencia por parte de la IA. Las imágenes resultantes de las indicaciones en las que se utilizó esta capa obtuvieron puntuaciones de dos, tres y cuatro en una proporción notablemente superior a aquellas en cuyas indicaciones habían participado las capas *scrible* o *canny*. Con una distribución más homogénea, el valor medio que obtuvieron las imágenes en las que se procesaron las referencias con la capa *Midas Depth* fue de casi 2 puntos, más de cuatro décimas por encima de la segunda, la capa *scrible*.

Figura 4.

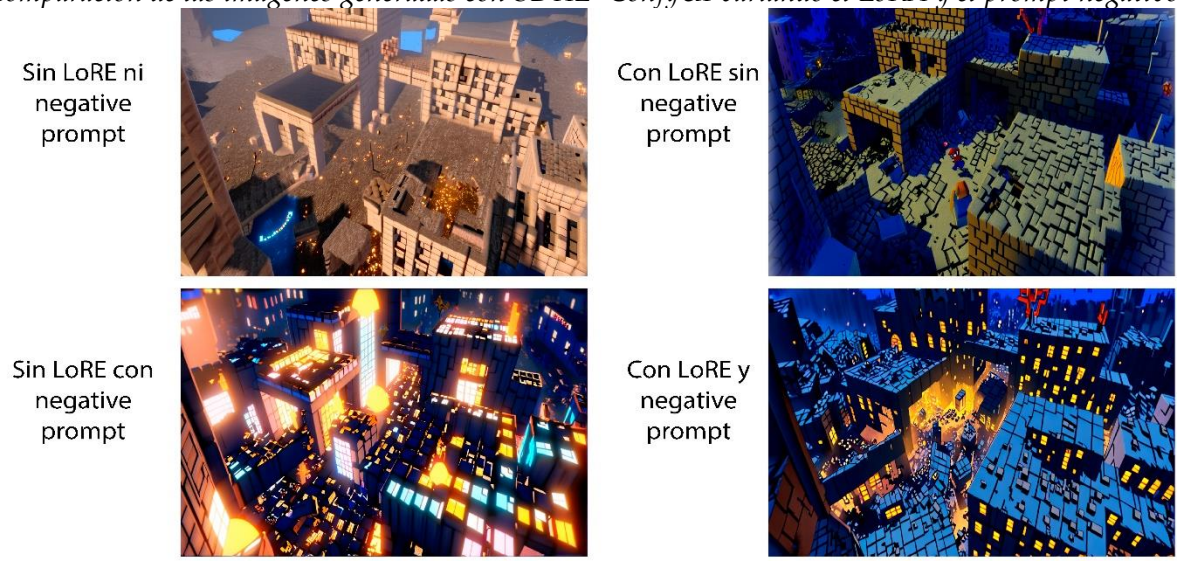
Comparación de las imágenes generadas con SDXL+ConfyUI variando del modo de control

Fuente: Elaboración propia mediante el software SDXL+ConfyUI con ControlNet mediante el checkpoint y LoRA mencionados en la metodología (2024).

Con respecto a la presencia o ausencia del modelo LoRA, el estudio reflejó que su incorporación a la indicación resultó beneficioso al ofrecer un mayor número de imágenes valoradas positivamente (con puntuaciones de tres y cuatro). Un caso muy diferente se dio con la integración del prompt negativo, cuya presencia favoreció la concentración de resultados cualificados con un uno a costa de aquellos valorados con doses y treses. Cabe destacar a mayores que, en ambos casos, la diferencia de las medias casi alcanzó los 0,2 puntos.

Figura 5.

Comparación de las imágenes generadas con SDXL+ConfyUI variando el LoRA y el prompt negativo



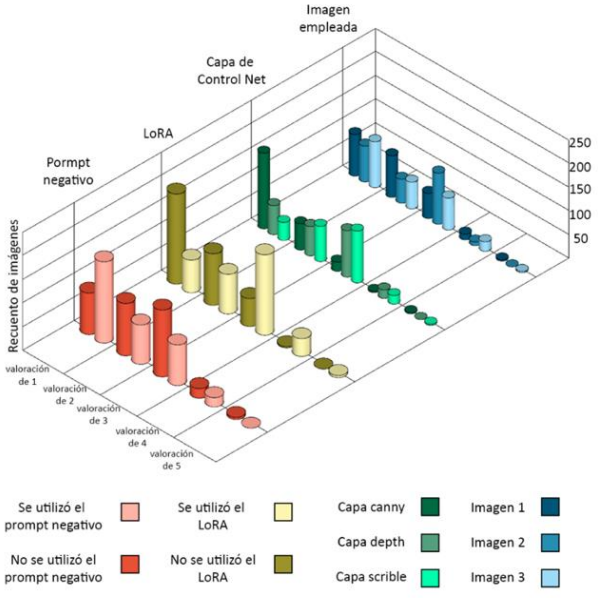
Fuente: Elaboración propia mediante el software SDXL+ConfyUI con ControlNet mediante el checkpoint y LoRA mencionados en la metodología (2024).

3.2.2. Adaptación del estilo

A diferencia de la competencia anterior, donde tan solo el doce por ciento de las imágenes alcanzaban una valoración de tres o superior, la gráfica de la figura 6 mostró una distribución más homogénea entre las diferentes puntuaciones. Aunque los resultados fueron más regulares, la variación en función del parámetro empleado en la indicación fue significativa.

Figura 6.

Gráfica de la distribución de puntuaciones en la competencia de "adaptación del estilo"



Fuente: Elaboración propia (2024).

Nuevamente, la segunda imagen de referencia produjo las propuestas visuales más acordes con el estilo solicitado. Pese aquellas indicaciones en las que se integró no llegaron a generar ninguna imagen valorada con cinco puntos y pese a haber dado lugar a solo ocho calificadas con un cuatro frente a las veinte que se generaron con la imagen tres, las ciento siete puntuadas con un tres fueron una cantidad suficientemente sólida como para llegar a la puntuación media de 2,2 puntos, dieciocho centésimas por encima de la segunda.

En lo referente a las capas de *ControlNet*, donde antes destacó la capacidad del procesador *Midas Depth* para la comprensión de la forma, en el caso de la adaptación del estilo, este ocupó un segundo lugar, siendo superado por la capa *scrible.* Esta no solo generó la menor cantidad de imágenes con una valoración de uno, sino que de las doscientas cuarenta imágenes procesadas con s*cribble*, ciento nueve obtuvieron una puntuación de tres, dieciocho de cuatro y una de cinco, logrando una media de 2,3 puntos por imagen.

Figura 7.

Comparación de las imágenes generadas con SDXL+ConfyUI variando la imagen de referencia

Fuente: Elaboración propia mediante el software SDXL+ConfyUI con ControlNet mediante el checkpoint y LoRA mencionados en la metodología (2024).

Con respecto a la presencia del LoRA y de la indicación textual negativa, los patrones observados en la primera competencia se acentuaron en la segunda. La integración del LoRA en la indicación de entrada dio lugar a imágenes con una nota media de 2,5 en términos de adaptación al estilo solicitado; mientras que el uso del prompt negativo condujo a imágenes valoradas en un punto y ocho décimas de media frente a la media de 2,3 puntos conseguida por las indicaciones sin prompt negativo.

Figura 8.

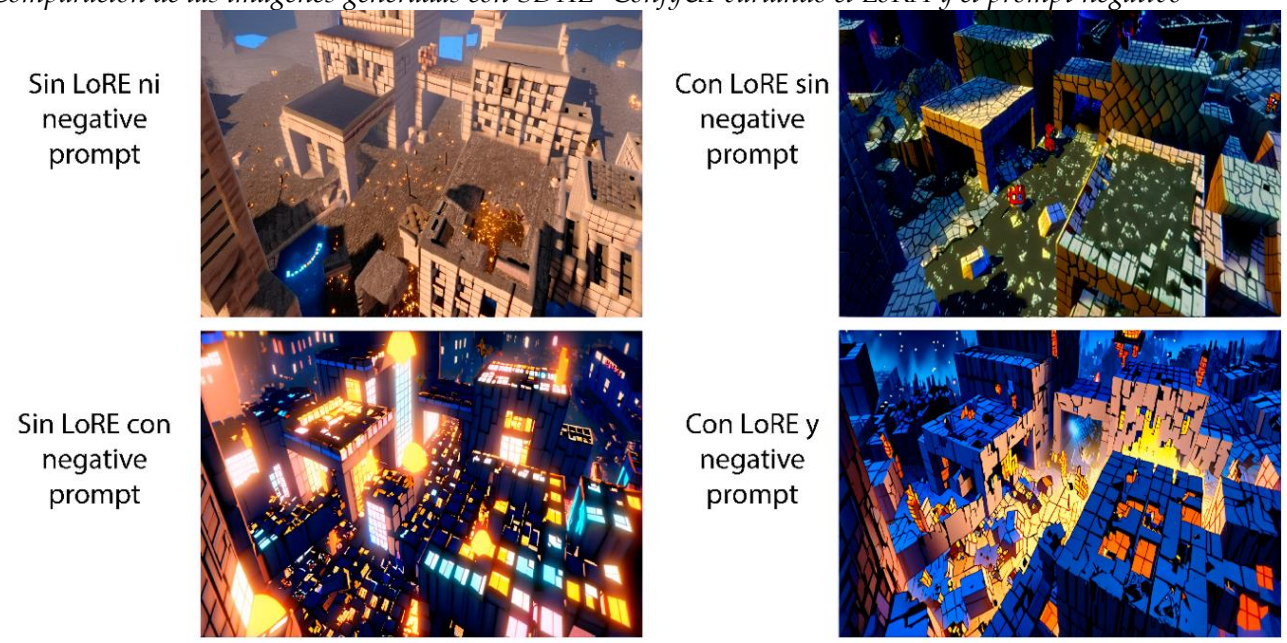
Comparación de las imágenes generadas con SDXL+ConfyUI variando del modo de control

Fuente: Elaboración propia mediante el software SDXL+ConfyUI con ControlNet mediante el checkpoint y LoRA mencionados en la metodología (2024).

En ambos casos, con la presencia de LoRA y ausencia de indicación textual negativa, se obtuvieron puntuaciones de tres o más en el cincuenta y siete por ciento y en el cuarenta y seis por ciento de los casos respectivamente. Sin embargo, solo la media de las propuestas visuales de ambientación generadas por la primera llegó a superar la valoración mínima positiva de 2,5 puntos.

Figura 9.

Comparación de las imágenes generadas con SDXL+ConfyUI variando el LoRA y el prompt negativo



Fuente: Elaboración propia mediante el software SDXL+ConfyUI con ControlNet mediante el checkpoint y LoRA mencionados en la metodología (2024).

3.2.3. Integración de la paleta cromática

En la cercera competencia, en la que se evaluó la capacidad de la IA para generar respuestas visuales de ambientación en el aspecto referido a correcta integración de la paleta de colores solicitada, se observaron dos aspectos relevantes.

En primer lugar, se comprobó que esta competencia fue, en términos generales, la que presentó una división más homogénea de los valores con una concentración mayor para las puntuaciones de dos y tres. De la muestra de setecientas veinte imágenes, el sesenta y cinco por ciento obtuvieron una calificación de dos o tres, siendo la segunda ligeramente más frecuente, como se muestra en la tabla 2.

Tabla 2.

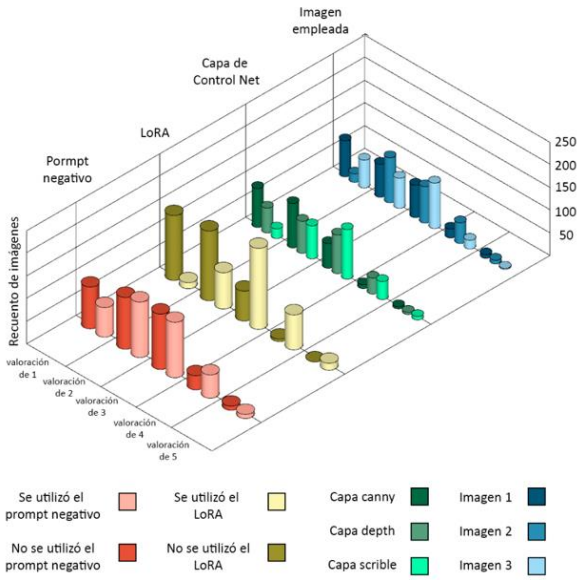
Comparación de la distribución total de imágenes según sus puntuaciones para cada competencia

Puntuaciones	Comprensión de la forma	Adaptación del estilo	Integración de la paleta cromática
1	486	258	154
2	149	193	232
3	77	227	240
4	8	40	80
5	0	2	14

Fuente: Elaboración propia al aplicar las tablas cruzadas sobre el test (2024).
En segundo lugar, y como se puede comprobar en la figura 10, la variación de las notas de las imágenes varió se forma muy notoria con relación a los parámetros que se integraron en las indicaciones. Así pues, la homogeneidad en el reparto de resultados se relacionó con el menor grado de coincidencia de las curvas calculadas para cada posible parámetro seleccionado.

Figura 6.

Gráfica de la distribución de puntuaciones en la competencia de "adaptación del estilo"



Fuente: Elaboración propia (2024).

De las imágenes de referencia, la primera volvió a ser la que produjo resultados con un mayor número de errores al haber obtenido estos una evaluación de dos o inferior en más del setenta por ciento de las ocasiones. Con respecto a las imágenes dos y tres, ambas alcanzaron medias favorables. La segunda imagen de referencia, pese a que la mayoría de las indicaciones en las que se empleó condujeron a respuestas visuales valoradas con dos puntos, su escasa tasa de imágenes puntuadas con un uno y su elevado número de generaciones calificadas con un cuatro le permitieron alcanzar una nota media de 2,5 puntos.

La tercera, por otra parte, a pesar de haberse relacionado con un gran número de imágenes evaluadas con un uno, presentó una curva de crecimiento creciente hasta el valor de tres, el obtenido con más solidez a lo largo de las doscientas cuarenta imágenes generadas para esta

referencia visual del *blockout*. De esta manera, fue, en lo referente a la integración de la paleta cromática, la tercera imagen la que obtuvo una mejor valoración con una media de dos puntos y setenta y cinco centésimas.

Figura 11.

Comparación de las imágenes generadas con SDXL+ConfyUI variando la imagen de referencia

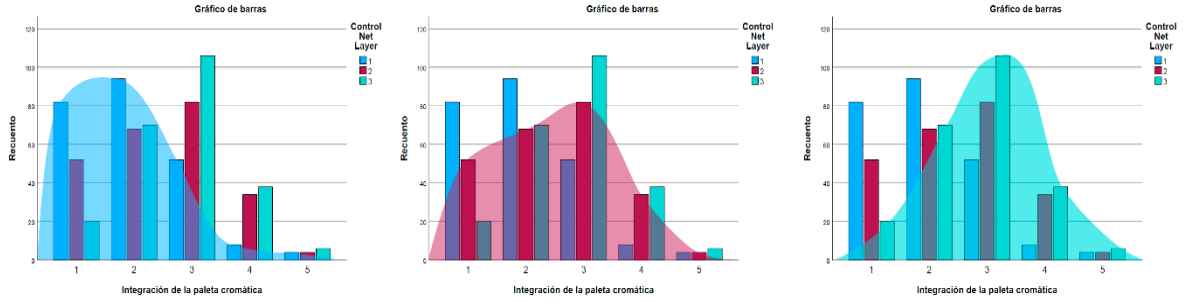
Fuente: Elaboración propia mediante el software SDXL+ConfyUI con ControlNet mediante el checkpoint y LoRA mencionados en la metodología (2024).

Con respecto a las Capas de *ControlNet*, el procesador *scrible* volvió a ser el que generó las doscientas cuarenta imágenes con una mayor nota media; 2,8 puntos, tres décimas por encima de la media correspondiente a las imágenes resultantes de emplear la capa *Midas Depth*.

La curva que representa las evaluaciones de las imágenes generadas con el procesador de profundidad tiene una pendiente menos pronunciada (Figura 12), lo que indica que la frecuencia de evaluaciones favorables (puntuaciones altas) fue baja en comparación con las evaluaciones desfavorables (puntuaciones inferiores a tres). De manera similar, las imágenes producidas con la capa *canny* tampoco obtuvieron buenos resultados, ya que la mayoría recibió puntuaciones de uno o dos.

Figura 12.

Comparación de las curvas de evaluación obtenidas por SPSS para las tres capas de ControlNet



Fuente: Elaboración propia mediante el software SPSS (2024).

En lo referente al LoRA, de nuevo, su incorporación a la indicación como parámetro de entrada adicional resultó de gran utilidad, pues de las trescientas imágenes en cuyas generaciones se incluyó, se observó que más del setenta y dos por ciento de las imágenes alcanzaban o superaban la calificación de tres en la competencia de integración de la paleta cromática.

La instrucción textual negativa, que anteriormente había demostrado ser perjudicial cuando se incluía como información adicional en las solicitudes, resultó en valores de variable poco diferenciados entre su presencia y ausencia. En ambos casos, la mayoría de las imágenes

fueron calificadas con dos o tres puntos. Sin embargo, al incluir el prompt negativo, se observó un aumento en la cantidad de imágenes calificadas con cuatro o cinco puntos. Esto condujo a que el promedio de calificaciones en los trescientos sesenta casos evaluados superara en catorce centésimas a los casos en los que no se había añadido en el procesamiento.

6. Discusión

Los resultados obtenidos sugieren que ciertas combinaciones de variables pueden ofrecer mejores resultados en la ambientación visual de niveles de videojuegos; algo que trabajos como el de Petráková y Šimkovič (2023) ya parecían adelantar en el campo de la arquitectura. No obstante, en el caso concreto, aunque se esperaba que la combinación de la segunda imagen con la capa *scrible,* el LoRA y sin prompt negativo fuera la mejor, las diez imágenes resultantes obtuvieron una valoración media de 2,6 puntos. En contraste, otras indicaciones similares, pero con ajustes diferentes, como la sustitución de la capa *scrible* por la *Midas Depth* (que había quedado en segundo lugar en dos de las tres competencias), alcanzaron valoraciones medias más altas, alcanzando 3 y 2,9 puntos en función de la escala CFG. Este mismo caso es extrapolable a otras variaciones, como la resultante de mantener la capa *scrible* pero tomando la imagen tres como referencia, lo que condujo a un valor medio de 2.7 puntos para ambas escalas CFG.

Así pues, se evidencia la importancia de entender estos resultados no como las combinaciones con las que obtener de manera consistente las mejores respuestas, sino como un apoyo y referencia de cuales podrían ser las combinaciones y variables que con mayor probabilidad generen una respuesta útil pero no última; lo que confirma que la IA parece actuar como un agente guiable pero no completamente controlable (Chang *et al.*, 2023).

Por otro lado, es crucial reconocer que el estudio no abarcó una amplia variedad de parámetros, como la combinación simultánea de varias capas, sus respectivos pesos de influencia o el sesgo dependiente de los modelos seleccionados. Explorar estos factores de manera más detallada podría mejorar significativamente la precisión de los resultados. El software ConfyUI ofrece numerosas opciones de configuración, lo que subraya la necesidad de una investigación más exhaustiva para determinar con mayor precisión cómo ajustar estos parámetros.

En el contexto de la generación de ambientes mediante inteligencia artificial para videojuegos, existen factores menos obvios que influyen significativamente. Por ejemplo, aunque dos imágenes de un mismo escenario sean idénticas en detalle, textura e iluminación, el análisis estadístico ANOVA reveló que la elección de una imagen de referencia en lugar de otra similar puede resultar en diferencias notables en los resultados obtenidos. Esto subraya la importancia de seleccionar cuidadosamente las imágenes de referencia, ya que puede impactar significativamente en la calidad y coherencia de los ambientes generados por los modelos de inteligencia artificial.

Durante el proceso de generación de imágenes, se observó que los modelos preentrenados tienen la capacidad de conservar la información de la semilla utilizada. Esto implica que si un usuario utiliza una configuración específica con una semilla particular, otro usuario puede replicar los resultados al utilizar la misma configuración inicial (Andrew, 2023a). Aunque este aspecto no afecta directamente a las valoraciones obtenidas en las pruebas realizadas, es crucial considerarlo para integrar eficazmente la herramienta en un entorno de trabajo real, donde la consistencia y la reproducibilidad son fundamentales para mantener la calidad del producto final (Fernández-Álvarez y López-Chao, 2023; Torres-Ferreyros *et al.*, 2017).

Además, se encontró que la escala CFG, aunque fue desechada después del análisis de varianza, tuvo un impacto localizado en la evaluación de algunas imágenes, especialmente en la evaluación del estilo visual. Este efecto no fue uniforme; en algunos casos, la escala CFG aumentó la puntuación de las imágenes, mientras que en otros casos la redujo. Este hallazgo resalta la importancia de comprender cómo ciertos parámetros, aparentemente secundarios, pueden influir en los resultados finales de la generación de ambientes visuales.

En general, a partir de los datos analizados, se ha comprobado que sí es posible intuir que tipo de variables podrían llegar a favorecer en mayor medida la obtención de imágenes útiles en el contexto de la ambientación visual de niveles de videojuegos asistida por inteligencia artificial generativa (Meira-Rodríguez y López-Chao, 2024). No obstante, también se ha evidenciado que la magnitud de los factores presentes en la misma, sumados a la cierta aleatoriedad con la que se decodifican las imágenes, obligan a mantener un flujo iterativo que, si bien llevado podría agilizar realmente el trabajo, la falta de conocimiento o la mala praxis podrían resultar muy perjudiciales de cara a la optimización del trabajo.

7. Conclusiones

Tras analizar los datos de más de setecientas imágenes generadas y evaluadas, este estudio ha revelado parte del potencial de las herramientas visuales generativas de inteligencia artificial basadas en modelos de difusión para asistir en el diseño de entornos y ambientación en videojuegos. Se ha demostrado que estas herramientas ofrecen una capacidad inicial para esbozar solicitudes completas con cierto grado control y atractivo visual. Sin embargo, se ha observado la complejidad inherente en el ecosistema de estos softwares, donde la iteración de parámetros es crucial debido a la vasta cantidad de variables presentes a lo largo del proceso generativo.

A pesar de la limitación de este estudio acotado, se ha destacado el potencial emergente de estas herramientas. Por consiguiente, para el futuro, sería prudente explorar líneas de investigación adicionales que contribuyan a establecer un marco de trabajo preciso. Esto facilitaría la integración cómoda y controlada de estos softwares en diversas fases de la producción de entornos digitales.

8. Referencias

Abdelsalam Soliman, M. (abril de 2024). *LinkedIn post April 2024*. LinkedIn. https://www.linkedin.com/posts/mai-abdelsalam_comfyui-grasshopper-ai-activity-7180160257367629824-QtCZ/

Ahmad, N. B., Barakji, S. A. R., Shahada, T. M. A., y Anabtawi, Z. A. (2017). How to launch a successful video game: A framework. *Entertainment Computing*, 23, 1-11. https://doi.org/10.1016/j.entcom.2017.08.001

Ameneh, P. y Microsoft, G. (2023). *LoRA-Enhanced Distillation on Guided Diffusion Models*. https://arxiv.org/abs/2312.06899v1

Andrew. (16 de noviembre de 2023). *Conoce estos parámetros importantes para imágenes impresionantes de IA - Arte de difusión estable*. https://stable-diffusion-art.com/know-these-important-parameters-for-stunning-ai-images/#CFG_Scale

Andrew. (19 de marzo de 2024). *Stable Diffusion Models: a beginner's guide*. Stable Diffusion Art. https://stable-diffusion-art.com/models/

Andrew. (2023a). *How to use Stable Diffusion to create AI-generated images*. STABLE DIFFUSION ART. https://stable-diffusion-art.com/beginners-guide/

Bulut, E. (2023). The Fantasy of Do What You Love and Ludic Authoritarianism in the Videogame Industry. *Television & New Media*, *24*(8), 851-869.

Chang, M., Druga, S., Fiannaca, A. J., Vergani, P., Kulkarni, C., Cai, C. J. y Terry, M. (2023). *The Prompt Artists*. Proceedings of the 15th Conference on Creativity and Cognition, 75–87. https://doi.org/10.1145/3591196.3593515

del Campo, M. (2021). *Architecture, Language and Ai*. Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia, Volumen 1, 1, pp. 211–220.

del Campo, M. y Leach, N. (2022). Unleashing New Creativities. *Architectural Design*, *92*(3), 122-135. https://doi.org/10.1002/ad.2823

del Campo, M., Carlson, A. y Manninger, S. (2020). *How machines learn to plan: A critical interrogation of machine vision techniques in architecture*. Proceedings of the 40th Annual Conference of the Association for Computer Aided Design in Architecture: Distributed Proximities, ACADIA 2020, 1, pp. 272–281.

Du, C., Li, Y., Qiu, Z. y Xu, C. (2023). *Stable Diffusion is Unstable*. https://github.com/duchengbin8/Stable_Diffusion_is_Unstable

Eisendorf, M. [@marcusisntcool]. (marzo de 2024). *Instagram post March 2024* [Video]. Instagram. https://www.instagram.com/reel/C3_Tv-ixmc3/

Fernández-Álvarez, Á. J. y López-Chao, V. (2023). Drawing, scripting, prompting. A critical approach from architectural graphics. *Diségno*, 13, 143-152. https://doi.org/10.26375/disegno.13.2023.16.13.2023.16

Fraunberger, A. (2023). *Linkedin post August 2023*. LinkedIn. https://www.linkedin.com/posts/dr-andreas-fraunberger_xr-avatars-genai-ugcPost-7095326558659260416-8a6H/

Game Developers Conference. (2022). *Physiological Effects of Crunch: A Look at the Science* [Video]. YouTube. https://www.youtube.com/watch?v=Sb2U_9IGgc0

Grepl-Malmgren, L. y Hallenbom, L. (2023). How Does the Level Design in The Last of Us Part 2 Use Lighting and Set Dressing to Suggest Whether an Area Is Safe or Dangerous?

Hu, E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. y Chen, W. (2022). *Lora: Low-Rank Adaptation of Large Language Models*. ICLR 2022 - 10th International Conference on Learning Representations, 1–4. https://doi.org/10.2312/pg.20231276

López-Chao, V., Fernández-Álvarez, Á. J. y Grela, M. R. (2023). Unreal memories: The collective image of architecture in visual social networks. *SOBRE: Prácticas Artísticas y Políticas de La Edición*, *9*(1), 31-42.

Lorenzo-Eiroa, P. y Sprecher, A. (2013). *Architecture in Formation: On the Nature of Information in Digital Architecture*. https://doi.org/10.4324/9781315890128

Meira-Rodríguez, P. y López-Chao, V. (2024). Diffusion Models for Environment Visualization: Leveraging Stable Diffusion as a Generator for Architectural Spatial Design. En L. Hermida González, J. P. Xavier, A. Amado Lorenzo y Á. J. Fernández-Álvarez (Eds.), *Graphic Horizons. EGA 2024. Springer Series in Design and Innovation* (pp. 417-426). Springer. https://doi.org/10.1007/978-3-031-57575-4_49

Musli, H. (marzo de 2024). *Explorando la intersección de la inteligencia artificial y la arquitectura en Roma* [Publicación de estado]. LinkedIn. https://www.linkedin.com/posts/hmusli_aiarchitecture-aiart-rome-ugcPost-7174513123406602240-ceM-

Nuray, Ö. (enero de 2024). *LinkedIn post January 2024* [Publicación en LinkedIn]. LinkedIn. https://www.linkedin.com/posts/omer-nuray_install-comfyui-ugcPost-7148360838360342528-fDyg/

Nuray, Ö. (marzo de 2024). *ComfyUI for Everything (other than stable diffusion)* [Video]. YouTube. https://www.youtube.com/watch?v=fUcDAExxndQ&t=1559s

Petráková, L. y Šimkovič, V. (2023). Architectural alchemy: Leveraging Artificial Intelligence for inspired design–a comprehensive study of creativity, control, and collaboration. *Architecture Papers of the Faculty of Architecture and Design STU*, *28*(4), 3-14.

Seleit, I. (marzo de 2024). *LinkedIn post March 2024*. LinkedIn. https://www.linkedin.com/posts/ismailseleit_ai-aiarchitecture-architecture-activity-7169239970254151680-G9KU/?utm_source=share&utm_medium=member_ios

Song, J. y Yip, D. (2023). *Exploring the Intersection of AI Art and Film: A Case Study of Giant*. Proceedings - 2023 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2023, pp. 347–352. https://doi.org/10.1109/ICMEW59549.2023.00066

Strossmayera, J. J. y Fakultet, O. (2023). *Text-to-image Stable Diffusion model*. University of Osijek.

Torres-Ferreyros, C. M., Festini-Wendorff, M. A. y Shiguihara-Juarez, P. N. (2017). *Developing a videogame using unreal engine based on a four stages methodology*. Actas del IEEE ANDESCON 2016. https://doi.org/10.1109/ANDESCON.2016.7836249

Wade, A. C. (2007). *The State of the Art: Western Modes of Videogame Production*. https://www.researchgate.net/publication/229035979

Williams, G. y Wuetherick, M. (2018). *Artist Workflow Improved: Digital Content Creation Tools Roundtripping y Worldbuilding* (Presented by Unity Technologies). GDC Vault. https://www.gdcvault.com/play/1024836/Artist-Workflow-Improved-Digital-Content

Zhang, Z., Fort, J. M. y Mateu, L. G. (2023). Exploring the Potential of Artificial Intelligence as a Tool for Architectural Design: A Perception Study Using Gaudí's Works. *Buildings*, *13*(7), 1863. https://doi.org/10.3390/buildings13071863

Žnidarič, M. (enero de 2024). *LinkedIn post January 2024*. LinkedIn. https://www.linkedin.com/posts/znidaricmarcel_architecture-characterdesign-ai-ugcPost-7147147404998336512-SODW/?utm_source=share&utm_medium=member_ios

AUTOR:

Pedro Meira-Rodríguez:
Universidade da Coruña.

Graduado en estudios de arquitectura en 2021 por la Universidade da Coruña. Máster en Diseño y Desarrollo de Videojuegos en 2022 por la Universidad Complutense de Madrid. Estudiante de Doctorado en el Programa Oficial de Doutorado en Novas Perspectivas en Documentación, Comunicación e Humanidades de la Universidade da Coruña desde 2023. Personal docente investigador de la Universidade da Coruña en el Departamento de Ingeniería Civil (septiembre 2023-actualidad).

pedro.meira.rodriguez@udc.es

Orcid ID: https://orcid.org/0009-0000-8355-5324
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=58891660300
Google Scholar: https://scholar.google.es/citations?user=aODiVOYAAAAJ&hl=es&oi=ao
ResearchGate: https://www.researchgate.net/profile/Pedro-Meira-Rodriguez-2