



Reconstruir para variar: IA generativa y flexibilidad del storyboard en preproducción

A Change of Pace: Generative AI and Storyboard Flexibility in Pre-production

Guillermo Mejías Martínez¹, Sonia Blanco²

¹Universidad Complutense de Madrid, España

²Universidad de Málaga, España

KEYWORDS

Storyboard
Film pre-production
Generative artificial
intelligence
ComfyUI
FLUX

ABSTRACT

This article presents an exploratory study on the operational maturity of generative artificial intelligence applied to storyboarding in fiction film pre-production. Its aim is to examine whether a local, modular workflow based on ComfyUI can produce a functionally comparable version of a previously published professional storyboard from a script fragment and, from that visual base, generate variations in framing, composition, and action resolution without having to restart the process from scratch. To this end, the study adopts a single-case design focused on six panels from a commercial feature-film sequence, reconstructed through a local *image-to-image* and *inpainting* workflow using the FLUX.2 [klein] 9B fp8 model, without employing LoRAs or ControlNet. The results show that functional reconstruction was viable for all six panels, while controlled variations were successfully produced in four of them. The main limitations emerged when the image required precise representation of highly specific spatial and causal relations, such as a projectile striking an aircraft or the imminence of a crash, an area in which recent literature has identified persistent weaknesses in open diffusion models. The study concludes that ComfyUI, as a node-based, local, and configurable environment, already offers relevant professional utility for rapid storyboard ideation, reformulation, and alternative exploration, although its full professional use still requires additional control mechanisms in sequences demanding greater spatial or kinetic precision.

PALABRAS CLAVE

Storyboard
Preproducción
cinematográfica
Inteligencia artificial
generativa
ComfyUI
FLUX

RESUMEN

Este artículo presenta un estudio exploratorio sobre la madurez operativa de la inteligencia artificial generativa aplicada al *storyboard* en la preproducción cinematográfica de ficción. El objetivo es analizar si un flujo de trabajo local y modular basado en ComfyUI permite obtener una versión funcionalmente comparable a un *storyboard* profesional previamente publicado a partir de un fragmento de guion y, a partir de esa base, generar variaciones de encuadre, composición y resolución de la acción sin rehacer íntegramente el proceso. Para ello, se adopta un diseño de caso único centrado en seis viñetas de una secuencia de largometraje comercial, reconstruidas mediante un flujo local de *image-to-image* e *inpainting* con el modelo FLUX.2 [klein] 9B en formato fp8, sin emplear LoRAs ni ControlNet. Los resultados muestran que la reconstrucción funcional fue viable en las seis viñetas, mientras que la producción de variaciones controladas resultó satisfactoria en cuatro. Las principales limitaciones surgieron cuando la imagen exigía representar con precisión relaciones espaciales y causales muy específicas, como el impacto de un proyectil sobre una aeronave o la inminencia de un accidente, un ámbito en el que la literatura reciente ha señalado debilidades persistentes en los modelos de difusión abiertos. Se concluye que ComfyUI, como entorno nodal, local y configurable, ofrece ya una utilidad profesional relevante para tareas rápidas de ideación, reformulación y exploración alternativa del *storyboard*, aunque su uso pleno requiere mecanismos adicionales de control en secuencias de mayor precisión espacial o cinética.

RECIBIDO: 23/06/2026
ACEPTADO: 03/10/2026

Cómo citar este artículo / Referencia normalizada: (Norma APA 7ª)

Mejías Martínez, G., Blanco, S. (2026). Reconstruir para variar: IA generativa y flexibilidad del storyboard en preproducción. *Prisma Social revista de ciencias sociales*, 55, 1-16. <https://doi.org/10.65598/rps.6287>

1. Introducción

El *storyboard* ocupa una posición central en la preproducción audiovisual porque convierte el guion en una secuencia visual de planos, acciones, relaciones espaciales y continuidades narrativas antes del rodaje. Su función no es únicamente ilustrativa. En la práctica industrial, permite anticipar la puesta en escena, discutir decisiones con dirección, fotografía, arte y efectos, y reducir incertidumbre cuando la secuencia contiene acción compleja. La bibliografía sobre el medio ha insistido en que este elemento es simultáneamente un instrumento de comunicación y una herramienta de racionalización del proceso de producción. Pallant y Price subrayan su relevancia histórica en la organización de la narración audiovisual, mientras que marcos contemporáneos lo definen como un artefacto de preproducción orientado a visualizar la acción y planificar los tiros de cámara. (Pallant & Price, 2015; Skills England, 2025).

En paralelo, la expansión reciente de la inteligencia artificial generativa ha desplazado parte del debate desde la automatización de tareas aisladas hacia la integración de herramientas de apoyo dentro de flujos de trabajo creativos. Los estudios más recientes sobre cine y producción audiovisual coinciden en que la IA generativa ya no aparece únicamente como objeto de especulación futurista, sino como infraestructura operativa en ámbitos como postproducción, creación de imágenes de referencia y la previsualización. Trabajos sobre el contexto español y europeo muestran que estas herramientas empiezan a institucionalizar nuevos perfiles profesionales y nuevas formas de colaboración entre especialistas técnicos y creativos. Al mismo tiempo, la literatura advierte que su valor no reside tanto en una sustitución integral del trabajo autoral como en la aceleración de fases exploratorias y la apertura de escenarios intermedios de previsualización que antes requerían más tiempo o más personal (Dueñas Mohedas & Jiménez Alcarria, 2025; Caballero & Sora-Domenjó, 2024; Caballero et al., 2025).

Dentro de ese ecosistema, los estudios sobre *storyboard* asistido por IA generativa se han orientado mayoritariamente hacia sistemas experimentales más amplios o hacia entornos específicamente diseñados para colaboración humano-máquina. Es el caso de ReelFramer, concebido para ayudar a periodistas a traducir noticias en guiones y *storyboards*, o de CineVision, que propone una plataforma de previsualización interactiva para la colaboración entre dirección y dirección de fotografía. Ambos trabajos son relevantes porque muestran que la utilidad de la IA generativa en *storyboard* no depende tanto de producir imágenes “terminadas” como de permitir iteraciones legibles, más rápidas y negociables dentro del proceso creativo. Sin embargo, buena parte de esa literatura se ha desarrollado sobre sistemas ad hoc. Falta todavía investigación aplicada sobre la conveniencia profesional de flujos de trabajo base construidos con herramientas abiertas, localmente ejecutables y ya presentes en la práctica de comunidades creativas (Wang et al., 2024; Wei et al., 2025).

Ese vacío resulta especialmente importante en un momento en el que entornos como ComfyUI se consolidan como motores nodales de inferencia y edición visual. La documentación oficial define ComfyUI como una aplicación y motor de inferencia de código abierto, basada en nodos, que permite combinar modelos, operaciones y flujos visuales altamente personalizables, ejecutables además en local. En el caso de la familia FLUX.2 [klein], la propia documentación y los materiales del desarrollador insisten en su orientación hacia flujos interactivos y edición de imagen de baja latencia, incluyendo tareas de *image-to-image* e *inpainting*. Esto convierte a ComfyUI en una plataforma particularmente adecuada para evaluar no tanto la calidad absoluta de una imagen aislada, sino la madurez operativa de un flujo de *storyboard* entendido como proceso iterativo y modificable (Comfy-Org, s. f.-a; ComfyUI, 2026).

La historia del *storyboard* muestra que este jamás ha sido un objeto exclusivamente gráfico. Su existencia surge de la necesidad de transformar un texto verbal en una cadena de elementos visuales compartidas por departamentos distintos. De ahí que sus cualidades principales no sean únicamente el dibujo, sino la capacidad de traducir la lógica secuencial del guion en una propuesta visual manipulable por equipos diversos. En producciones con acción compleja o con efectos especiales, el *storyboard* se vuelve además un instrumento técnico y de coordinación interdepartamental (Pallant & Price, 2015; Skills England, 2025).

Los estudios recientes sobre IA generativa en el audiovisual dibujan un escenario análogo: la cuestión central no es únicamente la generación de imágenes, sino la reorganización del trabajo alrededor de herramientas capaces de acelerar decisiones y producir versiones intermedias. Dueñas Mohedas y Jiménez Alcarria, al estudiar el perfil emergente del *AI artist* en la posproducción española, muestran que las herramientas generativas se integran mejor cuando se sitúan como apoyo especializado dentro de flujos ya existentes y no como sustitutos de oficios previos. Caballero, Sora-Domenjó y Codina, en su análisis de filmografía creada con IA, insisten igualmente en que estas tecnologías expanden los métodos de producción y exploración, aunque arrastran todavía problemas de coherencia, ética y control. En el ámbito de la edición, Caballero y Sora-Domenjó identifican una demanda profesional de interfaces más interactivas y dinámicas, capaces de mejorar el flujo sin comprometer la integridad artística del proceso (Dueñas Mohedas & Jiménez Alcarria, 2025; Caballero & Sora-Domenjó, 2024; Caballero et al., 2025).

Los autores de los softwares ReelFramer y CineVision coinciden en que la IA generativa aporta valor cuando introduce etapas intermedias de formulación y tanteo visual, reduciendo tiempos de prueba y aumentando el rango de alternativas sin perder inteligibilidad. No se trata solo de que la máquina cree, sino de que el sistema permita al usuario ensayar diferentes formulaciones y encontrar una propuesta suficientemente útil para la discusión posterior. Esta lógica coincide con modelos más amplios de co-creación con IA generativa que distribuyen agencia entre exploración automática y corrección humana. Para la preproducción cinematográfica, eso supone entender el *storyboard* asistido por IA no como imagen definitiva, sino como infraestructura de iteración visual (Wang et al., 2024; Wei et al., 2025).

Ahora bien, la literatura reciente advierte con claridad sobre los límites actuales de los modelos de difusión, incluso en sus versiones avanzadas. Feng et al. señalan que los modelos *text-to-image* suelen interpretar mal instrucciones complejas cuando la escena combina múltiples sujetos o vínculos atributo-objeto. Zhang et al. afirman de manera explícita que los modelos de difusión destacan en fotorealismo, pero siguen teniendo dificultades para representar relaciones espaciales precisas descritas en el prompt. Esta observación es especialmente pertinente para el *storyboard*, donde muchas imágenes deben representar no solo “qué aparece”, sino “cómo se relacionan exactamente los elementos en el espacio” y “qué evento causal se produce en ese instante”. En otras palabras: la dificultad no reside siempre en la apariencia, sino en la sintaxis espacial y cinética de la imagen (Feng et al., 2024; Zhang et al., 2025).

Este límite enlaza con dos líneas técnicas que conviene distinguir. Por un lado, ControlNet fue propuesto precisamente para añadir condicionamiento espacial a modelos de difusión preentrenados, permitiendo guiarlos con bordes o pose. Por otro, LoRA puede entenderse como una técnica de adaptación ligera aplicable a distintos modelos generativos para introducir especialización con un coste menor que un ajuste completo. La literatura y la documentación muestran que estas capas adicionales mejoran el control o la especialización, pero también añaden preparación y nuevas variables al flujo. Desde el punto de vista metodológico, por tanto, excluir ambos recursos en una primera prueba no es una carencia sino una decisión experimental: permite observar hasta dónde llega un flujo base antes de introducir correctores más avanzados (Hu et al., 2021; Zhang et al., 2023).

Además de la evidencia académica creciente, los discursos contemporáneos sobre creación audiovisual asistida por IA generativa subrayan que el mayor valor de estas herramientas aparece cuando permiten idear más deprisa, proponer alternativas y abaratar el acceso a formas preliminares de previsualización. Esa promesa debe ser contrastada con cuidado, porque los ahorros no son homogéneos y la calidad no siempre es suficiente. Los trabajos de Sookhom et al. recuerdan que esta tecnología no reduce automáticamente los costes si no existe un artista detrás con conocimiento de lenguaje cinematográfico y capacidad para dirigir el proceso. La pregunta correcta no es si sustituye al *storyboard artist*, sino si determinadas configuraciones de IA generativa ya son suficientemente maduras como para entrar de forma útil en una zona concreta del flujo profesional: la construcción rápida de bases visuales y la generación de variantes sobre ellas (Sookhom et al., 2023).

En este contexto, el interés de los entornos nodales no se limita a la generación de imágenes, sino a la reconfiguración del flujo de preproducción en un régimen menos lineal y más reversible. Frente a modelos secuenciales donde una modificación tardía obliga a rehacer fases enteras del proceso, herramientas como ComfyUI permiten reinyectar cambios de encuadre o composición sobre una base ya construida, manteniendo la trazabilidad de las operaciones aplicadas. Desde esta perspectiva, la IA generativa no actúa únicamente como instrumento de producción iconográfica, sino como una capa de articulación entre decisiones creativas y técnicas que puede facilitar la comunicación entre *storyboard*, dirección, fotografía y arte en fases tempranas del proyecto.

A partir de este contexto, el presente artículo propone un estudio de caso deliberadamente acotado. No pretende demostrar que la IA generativa reproduzca exactamente un *storyboard* profesional ni que sustituya al *storyboard artist* en términos generales. Su propósito es más modesto y, precisamente por ello, más útil para el debate industrial actual: comprobar si un flujo base de ComfyUI, sin módulos de ajuste fino, es capaz de aproximarse funcionalmente a una secuencia publicada y, sobre todo, si permite abrir variaciones operativas sobre esa base con una economía de tiempo y una flexibilidad plausibles para contextos profesionales de preproducción. En lugar de medir la excelencia estética final, el artículo pone el foco en la posibilidad de llegar a un escenario visual comparable y manipularlo después sin reiniciar el trabajo desde cero.

Este estudio persigue dos objetivos principales.

El primero consiste en evaluar si un flujo base ejecutado en ComfyUI permite reconstruir de forma funcionalmente comparable una breve secuencia de *storyboard* profesional previamente publicada. El término “funcionalmente comparable” se utiliza aquí para designar una equivalencia práctica en la lógica del plano, la disposición de la acción, la relación figura-fondo y la continuidad secuencial, sin exigir una reproducción exacta del dibujo original.

El segundo objetivo consiste en determinar si esa reconstrucción inicial puede convertirse en una plataforma de exploración visual, es decir, en una base desde la cual generar variaciones controladas de encuadre o énfasis de la acción sin rehacer completamente la secuencia.

De estos objetivos se desprenden dos preguntas de investigación (research questions):

RQ1. ¿Es viable obtener, mediante un flujo local y modular de ComfyUI sin LoRAs ni ControlNet, una reconstrucción funcionalmente comparable a un *storyboard* profesional publicado?

RQ2. ¿Permite esa reconstrucción introducir variantes útiles para la revisión de la puesta en escena con una operativa apreciable para contextos de preproducción de ficción?

2. Metodología

El artículo se articula como un estudio exploratorio comparativo de caso único. Este diseño se considera adecuado por tres razones. En primer lugar, porque el propósito no es generalizar estadísticamente a toda la industria, sino observar de forma controlada el comportamiento de un flujo concreto sobre un corpus reducido. En segundo lugar, porque el objeto de análisis (la viabilidad profesional de un flujo base) exige atender al detalle del proceso más que a la amplitud de la muestra. En tercer lugar, porque la comparación entre referencia profesional publicada y reconstrucción asistida por IA generativa requiere una trazabilidad operativa compatible con una metodología de caso.

Se seleccionó una secuencia de seis viñetas perteneciente al clímax aéreo de Batman (1989), largometraje dirigido por Tim Burton. Las viñetas analizadas corresponden a una escena en la que el personaje del Joker extrae un arma desproporcionada y dispara contra la aeronave de Batman (Batwing). El corpus visual procede del portfolio público de David Russell, artista de concepto y *storyboard* acreditado en la película, y de una reproducción contextualizada de la misma secuencia en un repositorio divulgativo especializado. La combinación de ambas fuentes

permite identificar la autoría y la adscripción de la secuencia (Burton, 1989; Russell, s. f.; Filmsketchr, 2012).

Las seis viñetas seleccionadas presentan los siguientes hitos de acción: una aproximación aérea inicial; un plano más cercano del Joker alcanzando el arma; la interacción visual entre personaje y aeronave; el disparo; el impacto sobre la aeronave; y la pasada final del vehículo en llamas por encima del personaje. Se trata, por tanto, de un microcorpus idóneo para observar cómo responde la IA tanto ante planos relativamente estables como ante otros donde la relación causal entre objetos y trayectorias adquiere un peso decisivo.

La reconstrucción se realizó en local mediante ComfyUI. Según su documentación oficial, se trata de una aplicación y motor de inferencia nodal de código abierto que permite combinar modelos, operaciones y flujos visuales mediante nodos, con posibilidad de ejecución local. Para este estudio se eligió un flujo base compuesto por operaciones de *image-to-image* e *inpainting*, evitando el uso de extensiones de control espacial avanzado o de ajuste fino. Esta elección permitió observar el rendimiento de una configuración base y reproducible, antes de introducir capas adicionales de control o especialización (Comfy-Org, s. f.-a). El modelo empleado fue FLUX.2 [klein] 9B fp8. La documentación oficial de ComfyUI y Black Forest Labs caracteriza la familia FLUX.2 [klein] como una arquitectura compacta orientada a flujos interactivos, edición de imagen y previsualización de baja latencia, con versiones base orientadas a flexibilidad y personalización. El uso de una variante fp8 resultó óptimo para un ordenador de gama media-alta y permitió ejecutar el estudio en una GPU GeForce RTX 3060 con 12 GB de VRAM GDDR6 (ComfyUI, 2026).

La exclusión de LoRAs y ControlNet respondió a un criterio metodológico. LoRA se emplea normalmente para adaptar el modelo a estilos concretos con menos costes que un *fine-tuning* completo, mientras que ControlNet añade mecanismos explícitos de condicionamiento espacial. Ambos recursos son potencialmente relevantes en producción, pero habrían introducido variables suplementarias que dificultarían evaluar la madurez operativa del flujo base. La pregunta del artículo no es cuánto puede mejorarse el resultado añadiendo controles adicionales, sino qué puede hacer ya un operador individual con una configuración mínima pero razonablemente profesional (Hu *et al.*, 2021; Zhang *et al.*, 2023).

La instalación se realizó mediante ComfyUI-Easy-Install, una modalidad de despliegue local simplificado que integra ComfyUI Manager para la gestión de dependencias, modelos y nodos. La documentación oficial de ComfyUI-Manager lo describe como una extensión orientada a instalar, actualizar, eliminar, desactivar o habilitar nodos personalizados, así como a gestionar recursos del entorno ComfyUI. Con el fin de mantener estabilidad experimental, no se aplicaron actualizaciones durante el proceso de generación, el cual se realizó durante de marzo de 2026. Esta decisión evitó que cambios de versión, nodos nuevos o incompatibilidades sobrevenidas alteraran un flujo de trabajo que debía permanecer constante a lo largo del caso de estudio (Comfy-Org, s. f.-b). Para mantener el estudio en un marco lo más abierto, local y metodológicamente controlable posible, no se recurrió a plataformas comerciales en la nube (como Freepik o Higgsfield) pese a su evidente interés profesional. Este tipo de servicios ofrecen mayor facilidad de uso, pero también agregan variables menos transparentes para un estudio exploratorio de caso único como modelos subyacentes no siempre declarados, políticas de moderación del contenido, compras de crédito y actualizaciones del servicio fuera del control del investigador. La ejecución local en ComfyUI permitió fijar de manera explícita el modelo utilizado, el tipo de operación aplicada, la secuencia de nodos y el coste temporal aproximado de cada iteración, manteniendo así un mayor control analítico sobre las condiciones de producción de las imágenes.

Con el fin de situar esta decisión metodológica en el panorama actual de herramientas de generación visual asistida por IA, la siguiente tabla resume de forma sintética las diferencias operativas más relevantes entre un flujo local basado en ComfyUI y las plataformas SaaS (*Software as a Service*, software como servicio) en la nube orientadas a generación y edición

audiovisual. No se trata de una comparación de rendimiento empírico entre servicios, sino de una síntesis metodológica útil para justificar la elección del entorno experimental.

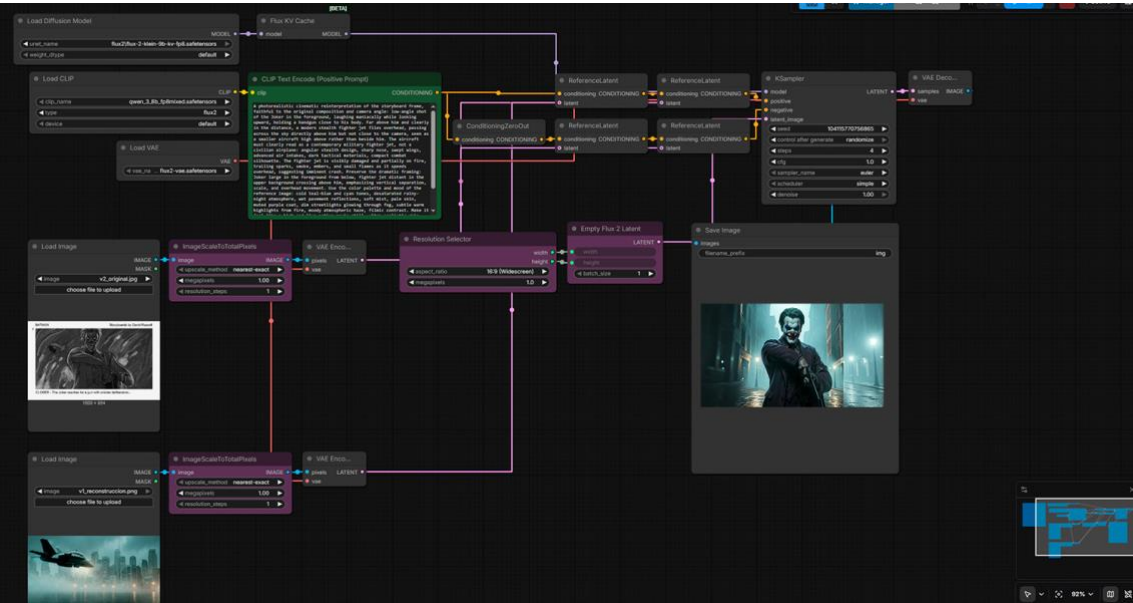
Tabla 1.

Comparación entre flujo local con ComfyUI y plataformas SaaS en la nube.		
Dimensión	Flujo local con ComfyUI	Plataformas SaaS en la nube
Transparencia	Alta: modelo, nodos y operaciones declarables	Media: backend y combinación de servicios menos visibles
Facilidad de uso	Menor: requiere instalación y manejo técnico	Mayor: acceso rápido, prompt e interfaz guiada
Moderación de contenidos	Menor mediación centralizada del servicio	Moderación y restricciones variables según plataforma o modelo
Estabilidad	Depende de versiones, nodos y mantenimiento local	Compatibilidad más gestionada por la plataforma
Adecuación	Alta, por control metodológico y trazabilidad	Interesante profesionalmente, pero menos limpia para aislar variables

Fuente: elaboración propia

La Figura 1 muestra una de las estructuras generales del workflow base utilizado en el estudio.

Figura 1. Workflow base de ComfyUI empleado para las operaciones de *image-to-image*, referencia latente e *inpainting* correctivo utilizadas en las reconstrucciones y variaciones



Fuente elaboración propia.

El procedimiento se desarrolló en cuatro fases.

En la primera fase se organizaron las seis viñetas originales y se extrajo el fragmento de guion correspondiente a la secuencia, junto con las indicaciones visuales presentes en las propias captions del storyboard. En la segunda fase se generó una reconstrucción base para cada una de las seis viñetas. Los prompts no se formularon como instrucciones aisladas, sino como estructuras descriptivas compuestas por varios campos recurrentes: tipo de plano, sujeto principal, acción dominante, relación espacial entre elementos, atmósfera visual, estilo cinematográfico y restricciones negativas (Tabla 2). Esta estrategia permitió mantener una lógica común entre viñetas sin exigir una repetición literal de cada instrucción. Dado que el objetivo del artículo no es analizar

lingüísticamente cada prompt, sino evaluar la viabilidad operativa del flujo, se presenta a continuación una síntesis de su estructura.

Tabla 2. Estructura sintética de los prompts empleados en la reconstrucción y variación de viñetas

Componente del prompt	Función dentro del flujo	Ejemplo de formulación
Tipo de plano	Orientar la composición general de la imagen	plano general; plano medio; contrapicado; vista aérea
Sujeto principal	Identificar el elemento visual dominante	aeronave militar; personaje con arma; ciudad nocturna
Acción dominante	Traducir el hito narrativo de la viñeta	personaje desenfunda un arma; aeronave se aproxima; disparo hacia el cielo
Relación espacial	Fijar la posición relativa entre elementos	personaje en primer término y aeronave al fondo; aeronave sobrevolando al personaje
Atmósfera visual	Mantener una coherencia estética general	iluminación nocturna; estética cinematográfica; alto contraste
Estilo de salida	Visualización preproducción	cinematic storyboard frame; production still; dramatic composition
Restricciones negativas	Reducir errores frecuentes o elementos no deseados	evitar deformaciones anatómicas; texto ilegible; objetos duplicados; composición confusa
Variación solicitada	Introducir cambios controlados sobre la base previa	cambiar el ángulo de cámara; aumentar la escala del personaje; intensificar el gesto; modificar la composición

Fuente: elaboración propia

A modo de ejemplo, un prompt de reconstrucción se estructuró a partir de la descripción del plano original, incorporando sujeto, acción y relación espacial: “plano cinematográfico nocturno de una aeronave militar aproximándose sobre una ciudad, composición dramática, alto contraste, estética de preproducción”. En cambio, los prompts de variación partieron de la imagen ya obtenida e introdujeron una modificación concreta, por ejemplo: “mantener la misma acción y atmósfera, pero modificar el encuadre hacia un ángulo más bajo, aumentando la presencia del personaje en primer término”. Estos ejemplos se incluyen únicamente para mostrar el criterio seguido en la construcción de los prompts, sin reproducir de forma exhaustiva todas las instrucciones utilizadas durante el proceso.

En la práctica, cada viñeta se trabajó mediante una media de cuatro generaciones del *workflow*, seleccionando después la salida más útil para la comparación. En las viñetas que lo exigieron se utilizaron operaciones de *inpainting* para corregir errores localizados o reforzar elementos gráficos concretos (por ejemplo, la inclusión de flechas de movimiento de forma similar al *storyboard* original). Esta combinación de selección entre generaciones e intervención localizada refuerza el carácter no lineal del flujo: en lugar de rehacer la imagen completa, el artista puede reabrir zonas concretas y modular la propuesta ya obtenida.

En la tercera fase se intentó producir una variación controlada de cada reconstrucción base, modificando encuadre, orientación de la acción o distribución espacial. Esta fase traslada la evaluación desde la mera capacidad de imitación hacia la capacidad de reformulación, que es donde se sitúa la principal ventaja industrial de estas herramientas.

En la cuarta fase se registraron indicadores operativos simples: tiempo de preparación del flujo, tiempo medio de ejecución por imagen, número medio de generaciones por viñeta, uso puntual de *inpainting* correctivo y éxito o fracaso en la generación de variantes.

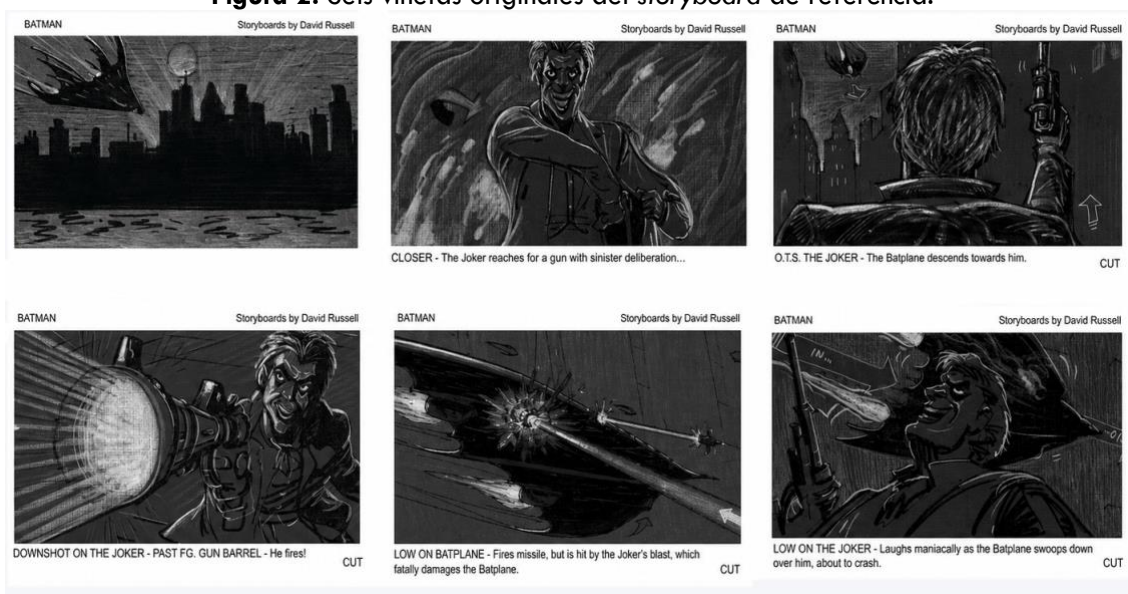
Dado el carácter exploratorio del estudio, los resultados se interpretaron mediante análisis descriptivo y comparación visual cualitativa. La reconstrucción se consideró funcionalmente lograda cuando cumplía cuatro criterios:

- a) identificación inequívoca del plano
- b) legibilidad clara de la acción
- c) mantenimiento de la continuidad básica con el resto de la secuencia
- d) conservación del vector dramático dominante de la viñeta original

La variación se consideró satisfactoria cuando introducía un cambio reconocible de encuadre, composición o resolución visual sin romper por ello la inteligibilidad de la escena ni rehacer desde cero la lógica del plano. No se recurrió a jueces externos ni a escalas de usabilidad, porque el objetivo del artículo es documentar la viabilidad de un flujo base efectivamente ejecutado y discutir su conveniencia profesional potencial.

3. Resultados

La reconstrucción base fue viable en las seis viñetas seleccionadas. En todos los casos se obtuvo una imagen que permitía reconocer el momento narrativo de la escena y la lógica principal del plano. Conviene precisar que las viñetas originales responden al lenguaje funcional propio del *storyboard* de producción: alto contraste, economía gráfica y prioridad de la información narrativa sobre el acabado pictórico. Esta aparente sencillez no desmerece su complejidad profesional, sino que evidencia que el valor del *storyboard* no reside necesariamente en la imagen como pieza final, sino en su capacidad para ser utilizada en el rodaje. Por ello, la comparación desarrollada en este estudio no evalúa la calidad plástica de la imagen generada frente al dibujo original, sino la conservación de la función planificadora que el *storyboard* profesional resuelve mediante una formulación gráfica deliberadamente sintética. La equivalencia no se produjo en el nivel del grafismo ni del estilo tonal exacto del *storyboard* original (dibujado en blanco sobre negro con una expresividad muy marcada), sino en el nivel funcional de la planificación visual. El flujo fue capaz de traducir posiciones de cámara, relaciones sujeto-entorno y acción general a una imaginería contemporánea de apariencia cinematográfica. Un detalle relevante del proceso fue que, para representar de forma operativa el concepto de Batwing o aeronave de Batman, fue necesario aproximar el prompt al concepto más genérico de “caza militar”. Esta reformulación evidencia una dificultad adicional del modelo para mantener referencias icónicas muy específicas cuando se combinan reconocimiento de objeto y variación de encuadre. Las viñetas centradas en el gesto del personaje o la confrontación básica entre el avión y Joker ofrecieron los resultados más sólidos. En estos casos, el modelo respondió bien a prompts relativamente estables: aproximación aérea, personaje desenfundando, mirada y dirección de la amenaza, paso de la aeronave sobre el personaje. La mayor fortaleza del sistema no fue la fidelidad estilística, sino la rápida obtención de composiciones legibles y suficientemente cercanas a la intención del plano original.

Figura 2. Seis viñetas originales del *storyboard* de referencia.

Fuente: David Russell, portfolio profesional; reproducción contextualizada en Filmsketchr (2012).
Uso con fines de análisis académico comparativo.

Figura 3. Seis reconstrucciones base obtenidas con ComfyUI a partir del fragmento de guion y de las viñetas de referencia.

Fuente: elaboración propia.

La fase de variación controlada fue satisfactoria en cuatro de las seis viñetas. En esas cuatro imágenes, el flujo permitió alterar con relativa eficacia el tamaño del sujeto, desplazar el punto de vista o intensificar el gesto principal sin perder el reconocimiento de la escena. Desde una perspectiva de preproducción, este es el hallazgo más relevante: una vez obtenida una base visual verosímil, ComfyUI se comportó como plataforma flexible para estudiar formulaciones alternativas del mismo momento dramático.

En cambio, las viñetas quinta y sexta no admitieron variantes controladas convincentes. La quinta requería representar el impacto de un proyectil disparado desde tierra sobre Batwing en pleno vuelo. La sexta exigía una situación cercana al accidente inminente: el Batwing en llamas pasando por encima del Joker con una dinámica espacial muy concreta. En ambos casos, las generaciones sucesivas tendieron a simplificar el evento, sustituirlo por otros más genéricos o degradar la precisión de la escena. El problema, por tanto, no fue la imposibilidad de generar una imagen vistosa, sino la dificultad de mantener una relación espacial y causal exacta entre trayectoria, impacto y estado del vehículo cuando se pedía variar el plano.

En los dos casos fallidos se observó, además, un problema semántico muy concreto. En la viñeta 5 el modelo no comprendió de forma estable la relación causal entre la bala disparada desde la pistola del Joker y el impacto sobre el Batwing. En la viñeta 6, la representación del Batwing en situación de pérdida de control tampoco pudo ser recreada variando el plano.

Figura 4. Cuatro variaciones satisfactorias generadas sobre la reconstrucción base.



Fuente: elaboración propia.

El tiempo de preparación del flujo y depuración inicial fue de aproximadamente 15 minutos. El tiempo medio de ejecución por generación fue de unos 3 minutos en la GPU utilizada. Cada imagen se generó, de media, cuatro veces antes de seleccionar el resultado más adecuado. En algunas variantes se empleó *inpainting* adicional, con un coste aproximado de 3 minutos por intervención correctiva, tanto para corregir pequeños errores como para reintegrar flechas u otros indicadores gráficos presentes en el storyboard original.

A partir de ese registro, la estimación operativa puede resumirse en la Tabla 3.

Tabla 3. Registro del flujo de trabajo

Fase	Unidades	Media de generaciones	Tiempo por generación	Tiempo estimado
Preparación y depuración del workflow	1	—	—	15 min
Reconstrucción base de 6 viñetas	6	4	3 min	72 min
Variaciones sobre 4 viñetas	4	4	3 min	48 min
<i>Inpainting</i> correctivo puntual	2-4 intervenciones	1	3 min	6-12 min
Total estimado	—	—	—	141-147 min

Fuente: elaboración propia

Aunque estos tiempos no equivalen al tiempo total de trabajo de un *storyboard artist* profesional (porque no incluyen briefing, lectura ni validación externa), sí muestran que un operador individual puede elaborar en unas dos y media horas una secuencia compuesta por seis reconstrucciones y cuatro alternativas.

Como referencia externa y no como cálculo de coste real del caso, puede utilizarse el promedio de tarifas de *storyboard artist* publicado por YunoJuno para el mercado freelance británico: 50 libras esterlinas por hora y 400 libras por jornada media. Esta fuente advierte además que la tarifa depende del número de planos, de las rondas de revisión, de la complejidad del estilo y de los derechos de uso. Por ello, el benchmark solo puede leerse como indicio de plausibilidad económica comparativa, no como equivalencia directa (YunoJuno, 2026).

Si se traslada de forma puramente orientativa ese promedio diario a media jornada, el coste laboral externo equivalente rondaría las 200 libras para una dedicación de unas cuatro horas. El flujo probado aquí se completó aproximadamente en ese rango temporal, pero con dos matices esenciales: primero, porque la comparación enfrenta trabajo humano especializado con ejecución asistida en local; segundo, porque el coste del flujo IA no incorpora amortización de hardware, aprendizaje del sistema ni el conocimiento cinematográfico del autor. Lo que sí sugiere el benchmark es que la ventaja competitiva más plausible no reside en abaratar el *storyboard*, sino en reducir el coste marginal de las variaciones una vez construida la base. Donde más valor aporta el flujo no es tanto en reemplazar la primera versión como en evitar rehacerla varias veces cuando se desean alternativas de puesta en escena.

4. Discusión

Los resultados permiten responder con matices afirmativamente a la primera pregunta de investigación. Sí es viable obtener con ComfyUI y un flujo base sin LoRAs ni ControlNet una reconstrucción funcionalmente comparable a un *storyboard* profesional en un corpus breve y muy controlado. Sin embargo, esa equivalencia ha de entenderse en términos industriales y no miméticos. El sistema no reproduce el trazo del *storyboard* original, sino que genera una visualización alternativa que conserva la lógica del plano y la secuencia. Esta matización es importante porque sitúa la utilidad profesional de la IA generativa en la fase de previsualización, no en la sustitución estilística del oficio del *storyboarding*.

La segunda pregunta también obtiene una respuesta afirmativa: la variación controlada es posible y, de hecho, constituye el principal rendimiento del flujo. Que cuatro de las seis viñetas hayan admitido reformulaciones satisfactorias sin recomenzar el proceso sugiere que ComfyUI ya es operativo como herramienta de exploración de alternativas. Esto conecta directamente con la literatura sobre *storyboard* y previsualización asistida por IA. El presente estudio confirma esa intuición en un entorno menos experimental y más cercano a prácticas de autoimplantación creativa: un único autor, un flujo local y hardware de consumo (Wang *et al.*, 2024; Wei *et al.*, 2025).

Desde el punto de vista técnico, los fallos observados en las viñetas quinta y sexta no invalidan el flujo, sino más bien acotan dónde termina su madurez actual. La dificultad para generar variaciones coherentes cuando la imagen exige representar con exactitud un impacto, una trayectoria o una inminencia de accidente encaja con lo que la literatura reciente ha diagnosticado como déficit compositivo de los modelos de difusión. Esto se produce cuando surgen problemas de seguimiento de instrucciones complejas y de relaciones espaciales precisas. En este estudio, esa limitación se manifestó con claridad: el flujo podía ofrecer versiones próximas al momento narrativo, pero no preservar de forma rigurosa el plano al variar el encuadre (Feng *et al.*, 2024; Zhang *et al.*, 2025).

Este punto permite matizar la hipótesis más optimista sobre el uso profesional inmediato de la IA generativa en *storyboard*. Si la tarea consiste en levantar con rapidez conceptos de plano, atmósferas o disposiciones de personajes, el flujo probado se muestra suficientemente maduro para ser útil. Si, en cambio, la secuencia requiere una precisión estricta en trayectorias, choques o detalles de acción físicamente muy específicos, el flujo base deja de ser plenamente fiable y necesita apoyo adicional.

Es aquí donde resulta pertinente, aunque solo como proyección razonada, la discusión sobre ControlNet y LoRA. ControlNet fue diseñado para añadir guiado espacial a modelos de difusión, y por tanto parece especialmente adecuado para escenas donde la pose o la dirección de un vector de acción no pueden quedar al arbitrio del modelo. De forma complementaria, LoRA permite adaptar el modelo a dominios iconográficos o estilísticos específicos con menos coste que un ajuste completo. La documentación asociada a FLUX y ComfyUI muestra, además, que el ecosistema ya contempla tanto ejemplos de *fill/inpainting* como de control adicional por profundidad o *canny*, e incluso variantes en formato LoRA para algunos flujos. Ahora bien, incorporarlos habría transformado este estudio en otra cosa: un análisis de optimización avanzada y no una prueba de madurez operativa de un flujo mínimo. Por tanto, los resultados nos indican que el flujo base resulta profesionalmente útil para una franja bien definida de tareas, mientras que las situaciones de alta precisión espacial justifican el uso de capas adicionales de control (Hu et al., 2021; Zhang et al., 2023).

Desde la perspectiva económica y organizativa, el hallazgo más valioso no es tampoco la reducción directa de costes, sino la disminución del coste de exploración. Este matiz coincide con Sookhom et al., que ya advertían que la IA no reduce automáticamente el gasto si no existe un autor competente y un encuadre técnico claro del uso. En el presente caso, el ahorro potencial aparece sobre todo en la posibilidad de llegar rápido a una visualización útil y en producir variaciones con un coste marginal bajo. En un contexto de producción real, esa capacidad puede resultar determinante para los diferentes departamentos de un rodaje cuando necesitan contrastar opciones antes de trasladar encargos a departamentos artísticos especializados (Sookhom et al., 2023).

La conveniencia profesional de un flujo local como el aquí estudiado no puede evaluarse al margen del coste técnico que comporta. Las plataformas en la nube presentan ventajas claras de acceso y velocidad de uso, especialmente para perfiles no técnicos o para entornos que priorizan la inmediatez operativa. En cambio, el trabajo en ComfyUI exige mayor alfabetización técnica, comprensión de nodos y atención a incidencias potenciales asociadas a actualizaciones o dependencias. Ahora bien, ese mayor coste de entrada se compensa con un grado superior de control sobre el pipeline, la selección de modelos, la trazabilidad del proceso y una menor mediación de políticas externas de servicio. A este coste técnico debe añadirse el coste material del hardware. Aunque el presente estudio se ejecutó en una GPU de consumo, el uso de modelos más potentes, resoluciones superiores, mayores lotes de generación o pipelines con más capas de control puede exigir tarjetas gráficas con mayor memoria VRAM y, por tanto una inversión inicial más elevada. En consecuencia, la adopción profesional de flujos locales no debe evaluarse solo por el coste por imagen o por el tiempo de generación, sino también por la amortización del hardware y la actualización periódica de los equipos (ComfyUI, 2026).

Por tanto, este trabajo sostiene que la conveniencia profesional actual de ComfyUI no debe medirse con el patrón de la sustitución. Su valor reside en otra parte: en servir de interfaz nodal para una preproducción más iterativa, local, controlable y asequible, especialmente en fases tempranas o en proyectos que necesitan muchas decisiones visuales preliminares en poco tiempo. Esta conclusión dialoga con el diagnóstico de Caballero y Sora-Domenjó sobre la demanda de herramientas más interactivas y con la documentación oficial de ComfyUI y FLUX, centrada precisamente en flujos editables y de baja latencia. Este trabajo aporta una evidencia práctica que ayuda a situar estas herramientas no en el terreno del espectáculo tecnológico, sino en el de la utilidad operativa condicionada (Caballero & Sora-Domenjó, 2024; Comfy-Org, s. f.-a; ComfyUI, 2026).

Como limitaciones, el presente trabajo establece cuatro. La primera es el carácter de caso único. Se ha analizado una sola secuencia, de solo seis viñetas, perteneciente a un film concreto y a una escena de acción muy singular. Los resultados no pueden extrapolarse sin cautela a otros géneros, estilos o tipos de *storyboard*. La segunda es la ausencia de evaluación o de validación con profesionales externos. La comparación se ha realizado mediante análisis cualitativo descriptivo, adecuado para el carácter exploratorio del estudio. La tercera es la dependencia del hardware

disponible y del modelo escogido. Los tiempos observados se refieren a una GPU concreta y a una variante concreta de FLUX. Con recursos distintos, el rendimiento podría variar. La cuarta es que el estudio se centra deliberadamente en un flujo base. No incorpora LoRAs, ControlNet ni otras capas de control que probablemente mejorarían parte de los resultados problemáticos. Esta exclusión es una fortaleza metodológica para el objetivo perseguido, pero limita el techo de calidad alcanzable.

Una extensión natural de este trabajo consistiría en incorporar mecanismos adicionales de control y especialización para medir hasta qué punto aumentan su consistencia interdepartamental, así como ampliar la muestra con distintos estilos de imágenes de origen y géneros diferentes. En un contexto profesional real, podrían emplearse ControlNet de pose, bordes o profundidad para asegurar relaciones espaciales más estables, así como LoRAs de estilo vinculadas a materiales ya legitimados por el propio proyecto (por ejemplo, concept art, diseño de producción o biblias visuales aprobadas por dirección de arte) con el fin de acercar el *storyboard* generado al ecosistema estético del resto de departamentos. Metodológicamente, esta posibilidad no debilita el presente estudio, sino que refuerza su interés como línea base sobre la que futuras investigaciones podrían medir incrementos de control, coherencia visual y adecuación industrial. También apunta a la consolidación de perfiles híbridos, situados entre el artista técnico, el integrador de pipeline y el especialista en visualización, capaces de traducir necesidades creativas en infraestructuras concretas.

5. Conclusiones

El estudio permite formular cuatro conclusiones principales. La primera es que un flujo base de ComfyUI ejecutado en local, con operaciones de *image-to-image* e *inpainting* y sin módulos de control avanzado, ya permite reconstruir de forma funcionalmente comparable una secuencia breve de *storyboard* profesional. Esto sitúa a la herramienta en un nivel de madurez suficiente para usos reales de previsualización temprana. La segunda es que el principal valor industrial del flujo no se encuentra en la imitación exacta del *storyboard* de referencia, sino en la posibilidad de construir sobre él. La generación de variaciones satisfactorias en cuatro de las seis viñetas demuestra que la IA generativa puede funcionar como mecanismo de revisión y reformulación, que es precisamente donde más fricción temporal existe en la práctica de preproducción. La tercera es que la madurez operativa del flujo es desigual. Funciona mejor cuando el plano se organiza alrededor de gestos claros y relaciones espaciales relativamente estables. Se debilita cuando la imagen exige precisión causal y espacial de alta resolución, como estados transicionales muy concretos del objeto representado. La cuarta es que esta limitación no invalida la conveniencia profesional de ComfyUI, sino que delimita su zona óptima de uso. Para *storyboard* rápido, conceptualización secuencial, tanteo de encuadres y exploración de alternativas, la herramienta es ya útil y manejable por un operador individual. Para secuencias donde la exactitud espacial o la consistencia icónica resultan críticas, parece aconsejable incorporar en el futuro mecanismos de control adicional como ControlNet o adaptaciones ligeras mediante LoRA, siempre valorando el coste añadido de complejidad.

En términos generales, el trabajo no defiende una sustitución del *storyboard artist*, sino una reasignación de funciones dentro del flujo de preproducción. La IA generativa, desplegada a través de ComfyUI, aparece aquí como un instrumento suficientemente maduro para acelerar la fase de exploración visual y abrir una zona intermedia entre el guion y el *storyboard* finalizado. Esa zona, históricamente costosa en tiempo y revisiones, es precisamente donde la herramienta demuestra hoy una conveniencia profesional más clara.

Referencias

- Burton, T. (Director). (1989). *Batman* [Película]. Warner Bros.
- Caballero, J., & Sora-Domenjó, C. (2024). Automation and creativity in AI-driven film editing: The view from the professional documentary sector. *Communication & Society*, 37(3), 201–218. <https://doi.org/10.15581/003.37.3.201-218>
- Caballero, J., Sora-Domenjó, C., & Codina, L. (2025). Implications of generative artificial intelligence in cinematic narrative and aesthetics: A study of entries to OpenDocs and the +RAIN film festival. *Anàlisi*, 71, 55–76. <https://doi.org/10.5565/rev/analisi.3773>
- Comfy-Org. (s. f.-a). ComfyUI official documentation. <https://docs.comfy.org/>
- Comfy-Org. (s. f.-b). ComfyUI-Manager installation. <https://docs.comfy.org/manager/install>
- ComfyUI. (2026, 15 de enero). FLUX.2 [klein] 4B & 9B: Fast local image editing and generation. ComfyUI Blog. <https://blog.comfy.org/p/flux2-klein-4b-fast-local-image-editing>
- Dueñas Mohedas, S., & Jiménez Alcarria, F. (2025). Generative artificial intelligence in media production: The emerging role of artificial intelligence artist in Spain. *Comunicação e Sociedade*, 47, e025011. [https://doi.org/10.17231/comsoc.47\(2025\).6212](https://doi.org/10.17231/comsoc.47(2025).6212)
- Feng, Y., Gong, B., Chen, D., Shen, Y., Liu, Y., & Zhou, J. (2024). Ranni: Taming *text-to-image* diffusion for accurate instruction following. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4744–4753). https://openaccess.thecvf.com/content/CVPR2024/html/Feng_Ranni_Taming_Text-to-image_Diffusion_for_Accurate_Instruction_Following_CVPR_2024_paper.html
- Filmsketchr. (2012, 18 de junio). *Batman 1989 Batwing concept art and storyboards* by David Russell. <https://filmsketchr.blogspot.com/2012/06/batman-1989-batwing-concept-art-and.html>
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-rank adaptation of large language models [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2106.09685>
- Pallant, C., & Price, S. (2015). *Storyboarding: A critical history*. Palgrave Macmillan. <https://doi.org/10.1057/9781137027603>
- Prisma Social. (s. f.). Política sobre uso de inteligencia artificial. <https://revistaprismasocial.es/ps/politicaia>
- Russell, D. (s. f.). *Batman* [Portfolio de *storyboard*]. <https://www.dynamicimagesdr.com/batman>
- Skills England. (2025). *Storyboard artist* (ST0489). <https://skillsengland.education.gov.uk/apprenticeships/st0489-v1-1>
- Sookhom, A., Klinthai, P., A-masiri, P., & Kerdvibulvech, C. (2023). A new study of AI artists for changing the movie industries. *Digital Society*, 2, Article 37. <https://doi.org/10.1007/s44206-023-00065-z>

- Wang, S., Menon, S., Long, T., Henderson, K., Li, D., Crowston, K., Hansen, M., Nickerson, J. V., & Chilton, L. B. (2024). ReelFramer: Human-AI co-creation for news-to-video translation. In Proceedings of the CHI Conference on Human Factors in Computing Systems. ACM. <https://doi.org/10.1145/3613904.3642868>
- Wei, Z., Wu, H., Zhang, L., Xu, X., Zheng, Y., Hui, P., Agrawala, M., Qu, H., & Rao, A. (2025). CineVision: An interactive pre-visualization storyboard system for director-cinematographer collaboration. In Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology. ACM. <https://doi.org/10.1145/3746059.3747793>
- YunoJuno. (s. f.). Storyboard artist freelance rates. Recuperado el 26 de abril de 2026 de <https://www.yunojuno.com/freelancer-rates-job-role/storyboard-artist>
- Zhang, G., Fu, B., Fan, Q., Zhang, Q., Liu, R., Gu, H., Zhang, H., & Liu, X. (2025). CoMPaSS: Enhancing spatial understanding in *text-to-image* diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 15253–15265). https://openaccess.thecvf.com/content/ICCV2025/html/Zhang_CoMPaSS_Enhancing_Spatial_Understanding_in_Text-to-image_Diffusion_Models_ICCV_2025_paper.html
- Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to *text-to-image* diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3836–3847). <https://doi.org/10.1109/ICCV51070.2023.00355>
- Zhang, R., Yu, B., Min, J., Xin, Y., Wei, Z., Shi, J. N., Huang, M., Kong, X., Xin, N. L., Jiang, S., Bahuguna, P., Chan, M., Hora, K., Yang, L., Liang, Y., Bian, R., Liu, Y., Campillo Valencia, I., Morales Tredinick, P., Kozlov, I., Jiang, S., Huang, P., Chen, N., Liu, X., & Rao, A. (2025). Generative AI for film creation: A survey of recent advances [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2504.08296>



CONTRIBUCIONES DE AUTORES/AS, FINANCIACIÓN Y AGRADECIMIENTOS

Contribuciones de los/as autores/as:

Conceptualización: Mejías Martínez, Guillermo y Blanco, Sonia

Software: Mejías Martínez, Guillermo

Validación: Mejías Martínez, Guillermo y Blanco, Sonia

Metodología: Mejías Martínez, Guillermo y Blanco, Sonia.

Curación de datos: Mejías Martínez, Guillermo

Redacción-Preparación del borrador original: Mejías Martínez, Guillermo

Redacción-Re- visión y Edición: Mejías Martínez, Guillermo y Blanco, Sonia.

Visualización: Mejías Martínez, Guillermo

Supervisión: Blanco, Sonia.

Administración de proyectos: Mejías Martínez, Guillermo y Blanco, Sonia.

Todos los/as autores/as han leído y aceptado la versión publicada del manuscrito: Mejías Martínez, Guillermo y Blanco, Sonia.

Financiación: DesinfoperIA Aplicaciones periodísticas de la IA para reducir la desinformación: tendencias, usos y percepciones de profesionales y audiencias Referencia: PID2023-147486OB-I00, financiado por Ministerio de Ciencia e Innovación.

Agradecimientos: gen-IA: Biblioteca de herramientas de inteligencia artificial para la creación de contenidos en medios de comunicación. Ref: JA.B3-23. PPRO-B3-2023-03. Financiado por el II Plan Propio de Investigación, Transferencia y Divulgación Científica. Universidad de Málaga

Conflicto de intereses: no existen conflictos de interés