Prisma Social revista de ciencias sociales , No. 55
Octubre 2026 e-ISSN 1989-3469. pp 1-16
DOI: https://doi.org/10.65598/rps.6287

Reconstructing to Vary: Generative AI and Storyboard Flexibility in Film Pre-production

Guillermo Mejías Martínez[1] , Sonia Blanco[2]
[1] Complutense University of Madrid, Spain
[2] University of Málaga, Spain

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. The 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, based on this visual foundation, generate variations in framing, composition, and action resolution without having to restart the entire 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 across all six panels, while the production of controlled variations proved successful in four of them. The main limitations emerged when the image required the 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 significant professional utility for rapid storyboard ideation, reformulation, and alternative exploration. However, its full professional implementation 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: 26/06/2026
ACEPTADO: 03/10/2026

How to cite this article / Standard reference: (APA 7th edition)
Mejías Martínez, G., Blanco, S.(2026). Reconstructing to Vary: Generative AI and Storyboard Flexibility in Film Pre-production. *Prisma Social revista de ciencias sociales*, *55*, 1-16. https://doi.org/10.65598/rps.6287

1. Introduction

Storyboard occupies a central position in audiovisual pre-production because it transforms the screenplay into a visual sequence of shots, actions, spatial relations, and narrative continuities before filming begins. Its function is not merely illustrative. In industrial practice, it makes it possible to anticipate mise-en-scène, discuss decisions with directors, cinematographers, art, and effects departments, and reduce uncertainty when a sequence involves complex action. The literature on the medium has consistently emphasised that the storyboard functions simultaneously as a communication instrument and as a tool for rationalising the production process. Pallant and Price underline its historical relevance in the organisation of audiovisual storytelling, while contemporary frameworks define it as a pre-production artefact aimed at visualising action and planning camera shots (Pallant & Price, 2015; Skills England, 2025).

At the same time, the recent expansion of generative artificial intelligence has shifted part of the debate from the automation of isolated tasks towards the integration of support tools within creative workflows. The most recent studies on cinema and audiovisual production agree that generative AI no longer appears merely as an object of futuristic speculation, but rather as an operational infrastructure in areas such as post-production, reference-image creation, and previsualisation. Research focused on the Spanish and European contexts shows that these tools are beginning to institutionalise new professional profiles and new forms of collaboration between technical and creative specialists. At the same time, the literature warns that their value lies less in the complete replacement of authorial labour than in the acceleration of exploratory phases and the opening of intermediate previsualisation scenarios that previously required more time or larger teams (Dueñas Mohedas & Jiménez Alcarria, 2025; Caballero & Sora-Domenjó, 2024; Caballero et al., 2025).

Within this ecosystem, studies on AI-assisted storyboarding have focused predominantly on broader experimental systems or on environments specifically designed for human–machine collaboration. This is the case with ReelFramer, conceived to help journalists translate news into scripts and storyboards, and CineVision, which proposes an interactive previsualisation platform for collaboration between directors and cinematographers. Both studies are relevant because they demonstrate that the usefulness of generative AI in storyboarding depends less on producing "finished" images than on enabling more legible, faster, and more negotiable iterations within the creative process. However, much of this literature has been developed around ad hoc systems. There is still a lack of applied research into the professional suitability of foundational workflows built with open tools, executable locally, and already present within the practices of creative communities (Wang et al., 2024; Wei et al., 2025).

This gap becomes particularly significant at a moment when environments such as ComfyUI are becoming consolidated as node-based engines for inference and visual editing. The official documentation defines ComfyUI as an open-source, node-based inference engine and application that allows highly customisable combinations of models, operations, and visual workflows, all executable locally. In the case of the FLUX.2 [klein] family, both the official documentation and the developer materials emphasise its orientation towards interactive workflows and low-latency image editing, including *image-to-image* and *inpainting* tasks. This makes ComfyUI a particularly suitable platform for evaluating not so much the absolute quality of an isolated image as the operational maturity of a storyboard workflow understood as an iterative and modifiable process (Comfy-Org, n.d.-a; ComfyUI, 2026).

The history of the storyboard demonstrates that it has never been an exclusively graphic object. Its existence arises from the need to transform verbal text into a chain of visual elements shared across different departments. Consequently, its principal qualities do not reside solely in drawing, but in its ability to translate the sequential logic of the screenplay into a visual proposal that can be manipulated by diverse teams. In productions involving complex action or special effects, the storyboard also becomes a technical and interdepartmental coordination instrument (Pallant & Price, 2015; Skills England, 2025).

Recent studies on generative AI in audiovisual production outline an analogous scenario: the central issue is not merely image generation, but the reorganisation of labour around tools capable of accelerating decision-making and producing intermediate versions. Dueñas Mohedas and Jiménez Alcarria, in their study of the emerging AI artist profile within Spanish post-production, show that generative tools are integrated more effectively when positioned as specialised support within existing workflows rather than as substitutes for established professions. Caballero, Sora-Domenjó, and Codina, in their analysis of AI-generated filmography, likewise argue that these technologies expand methods of production and exploration, although they still carry problems related to coherence, ethics, and control. Within the field of editing, Caballero and Sora-Domenjó identify a professional demand for more interactive and dynamic interfaces capable of improving workflow without compromising the artistic integrity of the process (Dueñas Mohedas & Jiménez Alcarria, 2025; Caballero & Sora-Domenjó, 2024; Caballero et al., 2025).

The authors of ReelFramer and CineVision agree that generative AI provides value when it introduces intermediate stages of visual formulation and experimentation, reducing testing time and increasing the range of alternatives without losing intelligibility. The issue is not simply that the machine creates, but that the system enables users to test different formulations and arrive at a proposal sufficiently useful for subsequent discussion. This logic aligns with broader models of co-creation with generative AI that distribute agency between automated exploration and human correction. For film pre-production, this means understanding AI-assisted storyboarding not as a final image, but as an infrastructure for visual iteration (Wang et al., 2024; Wei et al., 2025).

At the same time, recent literature clearly warns about the current limitations of diffusion models, even in their advanced versions. Feng et al. point out that text-to-image models often misinterpret complex instructions when scenes combine multiple subjects or attribute–object relations. Zhang et al. explicitly state that diffusion models excel at photorealism but still struggle to represent precise spatial relations described in prompts. This observation is particularly relevant for storyboarding, where many images must represent not only “what appears”, but also “how elements are spatially related” and “which causal event occurs at that precise moment”. In other words, the difficulty does not always lie in appearance, but in the spatial and kinetic syntax of the image (Feng et al., 2024; Zhang et al., 2025).

This limitation connects with two technical approaches that should be distinguished. On the one hand, ControlNet was proposed precisely to add spatial conditioning to pretrained diffusion models, allowing them to be guided through edges or pose information. On the other hand, LoRA may be understood as a lightweight adaptation technique applicable to different generative models in order to introduce specialisation at a lower cost than full fine-tuning. The literature and technical documentation show that these additional layers improve control and specialisation, but they also add preparation requirements and new variables to the workflow. From a methodological perspective, therefore, excluding both resources from an initial test is not a limitation but an experimental decision: it allows observation of how far a foundational workflow can go before more advanced corrective mechanisms are introduced (Hu et al., 2021; Zhang et al., 2023).

In addition to the growing academic evidence, contemporary discussions surrounding AI-assisted audiovisual creation emphasise that the greatest value of these tools emerges when they enable faster ideation, the proposal of alternatives, and cheaper access to preliminary forms of previsualisation. This promise, however, must be assessed carefully, since cost reductions are not homogeneous and quality is not always sufficient. The work of Sookhom et al. reminds us that this technology does not automatically reduce costs unless there is an artist behind the process with knowledge of cinematic language and the capacity to direct it. The relevant question is not whether generative AI replaces the storyboard artist, but whether certain generative AI configurations are already sufficiently mature to enter usefully into a specific area of the professional workflow: the rapid construction of visual foundations and the generation of variations from them (Sookhom et al., 2023).

In this context, the interest of node-based environments lies not only in image generation, but also in the reconfiguration of pre-production workflows into a less linear and more reversible regime. In contrast to sequential models in which a late modification forces entire phases of the process to be redone, tools such as ComfyUI make it possible to reintroduce framing or compositional changes into an already constructed base while maintaining traceability of the applied operations. From this perspective, generative AI acts not merely as an instrument for iconographic production, but as a layer of articulation between creative and technical decisions capable of facilitating communication between storyboard, directing, cinematography, and art departments during the early stages of a project.

Against this background, the present article proposes a deliberately delimited case study. It does not seek to demonstrate that generative AI can reproduce a professional storyboard exactly, nor that it can replace the storyboard artist in general terms. Its purpose is more modest and, precisely for that reason, more useful for the current industrial debate: to determine whether a foundational ComfyUI workflow, without fine-tuning modules, is capable of functionally approximating a published sequence and, above all, whether it allows operational variations to be developed from that base with plausible levels of speed and flexibility for professional pre-production contexts. Rather than measuring final aesthetic excellence, the article focuses on the possibility of reaching a comparable visual scenario and subsequently manipulating it without restarting the work from scratch.

This study pursues two principal objectives.

The first is to evaluate whether a foundational workflow executed in ComfyUI can reconstruct a previously published professional storyboard sequence in a functionally comparable manner. The term "functionally comparable" is used here to designate a practical equivalence in shot logic, action arrangement, figure–background relations, and sequential continuity, without requiring exact reproduction of the original drawing.

The second objective is to determine whether this initial reconstruction can become a platform for visual exploration, that is, a basis from which controlled variations in framing or action emphasis may be generated without completely rebuilding the sequence.

From these objectives arise two research questions:

RQ1. Is it viable to obtain, through a local and modular ComfyUI workflow without LoRAs or ControlNet, a reconstruction functionally comparable to a published professional storyboard?

RQ2. Does this reconstruction make it possible to introduce useful variations for reviewing mise-en-scène with an operational workflow of practical value for fiction film pre-production contexts?

2. Methodology

The article is structured as an exploratory comparative single-case study. This design is considered appropriate for three reasons. First, because the purpose is not to generalise statistically to the industry as a whole, but rather to observe in a controlled manner the behaviour of a specific workflow applied to a limited corpus. Second, because the object of analysis — the professional viability of a foundational workflow — requires close attention to process detail rather than broad sample size. Third, because the comparison between a published professional reference and an AI-assisted reconstruction requires operational traceability compatible with a case-study methodology.

A six-panel sequence from the aerial climax of *Batman* (1989), directed by Tim Burton, was selected. The analysed panels correspond to a scene in which the Joker draws an oversized weapon and fires at Batman's aircraft (the Batwing). The visual corpus was obtained from the public portfolio of David Russell, concept and storyboard artist credited on the film, together with a contextualised reproduction of the same sequence published in a specialised online repository.

The combination of both sources makes it possible to identify the authorship and attribution of the sequence (Burton, 1989; Russell, n.d.; Filmsketchr, 2012).

The six selected panels present the following action milestones: an initial aerial approach; a closer shot of the Joker reaching for the weapon; the visual interaction between character and aircraft; the firing of the shot; the impact on the aircraft; and the final pass of the burning vehicle above the character. The sequence therefore constitutes an especially suitable micro-corpus for observing how generative AI responds both to relatively stable shots and to others in which causal relations between objects and trajectories become critically important.

The reconstruction was carried out locally using ComfyUI. According to its official documentation, ComfyUI is an open-source, node-based inference engine and application that allows models, operations, and visual workflows to be combined through nodes, with the possibility of local execution. For this study, a foundational workflow composed of *image-to-image* and *inpainting* operations was selected, deliberately avoiding the use of advanced spatial-control extensions or fine-tuning mechanisms. This decision made it possible to observe the performance of a foundational and reproducible configuration before introducing additional layers of control or specialisation (Comfy-Org, n.d.-a).

The model employed was FLUX.2 [klein] 9B fp8. Official documentation from ComfyUI and Black Forest Labs characterises the FLUX.2 [klein] family as a compact architecture oriented towards interactive workflows, image editing, and low-latency previsualisation, with foundational versions designed for flexibility and customisation. The use of an fp8 variant proved optimal for a mid-to-high-range computer and made it possible to execute the study on a GeForce RTX 3060 GPU with 12 GB of GDDR6 VRAM (ComfyUI, 2026).

The exclusion of LoRAs and ControlNet followed a methodological rationale. LoRA is generally used to adapt models to specific styles at lower cost than full fine-tuning, while ControlNet introduces explicit spatial-conditioning mechanisms. Both resources are potentially relevant in production contexts, but they would also have introduced supplementary variables that would make it more difficult to evaluate the operational maturity of the foundational workflow itself. The question posed by this article is not how much the results may be improved through the addition of extra control systems, but rather what an individual operator can already achieve using a minimal yet reasonably professional configuration (Hu et al., 2021; Zhang et al., 2023).

The installation was carried out using ComfyUI-Easy-Install, a simplified local deployment system integrating ComfyUI Manager for dependency, model, and node management. Official documentation describes ComfyUI-Manager as an extension intended to install, update, remove, disable, or enable custom nodes, as well as to manage ComfyUI environment resources. In order to maintain experimental stability, no updates were applied during the generation process, which took place throughout March 2026. This decision prevented version changes, newly introduced nodes, or unexpected incompatibilities from altering a workflow that needed to remain constant throughout the case study (Comfy-Org, n.d.-b).

In order to keep the study within as open, local, and methodologically controllable a framework as possible, commercial cloud-based platforms (such as Freepik or Higgsfield) were not used despite their evident professional relevance. These services offer greater ease of use, but they also introduce variables that are less transparent for an exploratory single-case study, such as undisclosed underlying models, content moderation policies, credit-purchase systems, and service updates beyond the researcher's control. Local execution through ComfyUI made it possible to specify explicitly the model employed, the type of operation applied, the sequence of nodes, and the approximate temporal cost of each iteration, thereby maintaining greater analytical control over the image-production conditions.

In order to situate this methodological decision within the current landscape of AI-assisted visual-generation tools, the following table summarises the most relevant operational differences

between a local workflow based on ComfyUI and cloud-based SaaS (Software as a Service) platforms oriented towards audiovisual generation and editing. This is not intended as an empirical performance comparison between services, but rather as a methodological synthesis useful for justifying the choice of the experimental environment.

Table 1.

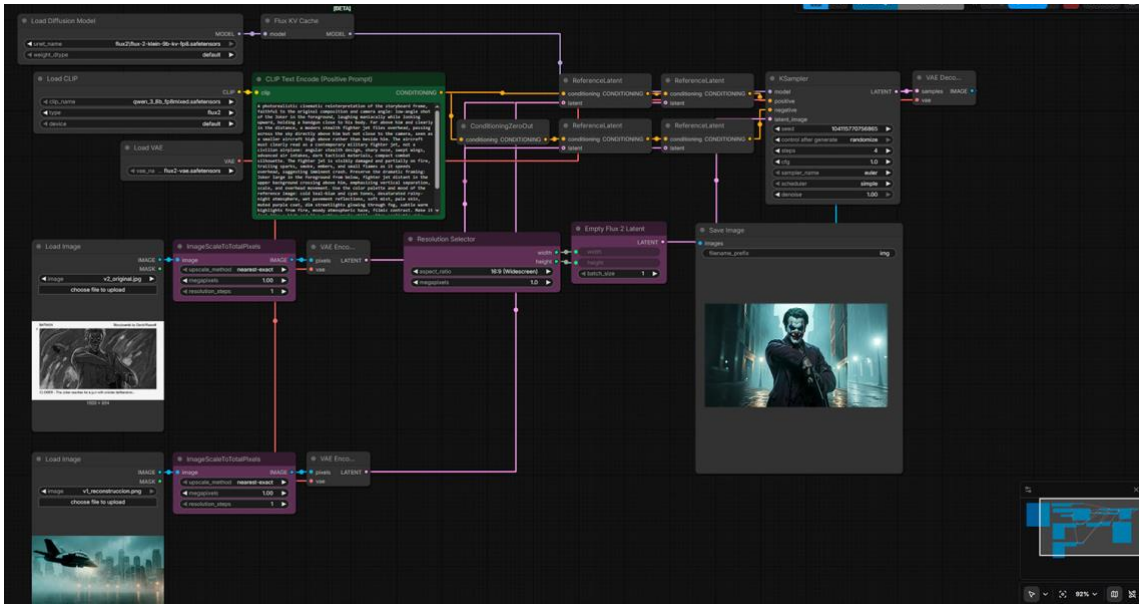
Comparison between a local ComfyUI workflow and cloud-based SaaS platforms.

Dimension	Local workflow with ComfyUI	Cloud-based SaaS platforms
Transparency	High: models, nodes and operations are fully identifiable	Medium: backend and combination of services less visible
Ease of use	Lower: requires installation and technical expertise	High: quick, prompt access and guided interface
Content moderation	Less centralised mediation of the service	Moderation and restrictions vary depending on the platform or model
Stability	Depends on versions, nodes and local maintenance	Compatibility managed more by the platform
Suitability	High, due to methodological control and traceability	Professionally interesting, but less suitable for isolating variables

Source: Authors' own elaboration, 2026.

Figure 1 shows one of the general configurations of the base workflow used in the study.

Figure 1. Base ComfyUI workflow used for the *image-to-image*, latent reference, and corrective *inpainting* operations employed in the reconstructions and variations.



Source: Authors' own elaboration, 2026.

The procedure was developed in four phases.

In the first phase, the six original panels were organised and the screenplay fragment corresponding to the sequence was extracted, together with the visual indications contained in the storyboard captions themselves. In the second phase, a base reconstruction was generated for each of the six panels. The prompts were not formulated as isolated instructions, but rather as descriptive structures composed of several recurring fields: shot type, main subject, dominant action,

spatial relations between elements, visual atmosphere, cinematic style, and negative constraints (Table 2). This strategy made it possible to maintain a coherent logic across panels without requiring literal repetition of every instruction. Since the aim of the article is not to analyse each prompt linguistically, but rather to evaluate the operational viability of the workflow, a synthesis of their structure is presented below.

Table 2

Summary of the prompt structure used for storyboard panel reconstruction and variation

Prompt component	Role within the workflow	Example of formulation
Shot type	Guide the general composition of the image	wide shot; medium shot; low angle shot; aerial view
Main subject	Identify the dominant visual element	military aircraft; figure with a weapon; city at night
Dominant action	Translate the narrative milestone of the panel	character draws a weapon; aircraft approaches; shot fired into the sky
Spatial relation	Establish the relative position between elements	character in the foreground and aircraft in the background; aircraft flying over the character
Visual atmosphere	Maintain overall aesthetic coherence	night lighting; cinematic aesthetic; high contrast
Output style	Pre-production visualisation	cinematic storyboard frame; production still; dramatic composition
Negative constraints	Reduce common errors or unwanted elements	avoid anatomical distortions; illegible text; duplicate objects; confusing composition
Requested variation	Introduce controlled changes to the previous base	change camera angle; increase the character's scale; intensify gesture; modify composition

Source: Authors' own elaboration

By way of example, a reconstruction prompt was structured from the description of the original shot, incorporating subject, action, and spatial relation: "night-time cinematic shot of a military aircraft approaching above a city, dramatic composition, high contrast, pre-production aesthetic". By contrast, the variation prompts took the previously generated image as their starting point and introduced a specific modification, for example: "maintain the same action and atmosphere, but modify the framing towards a lower angle, increasing the prominence of the character in the foreground". These examples are included solely to illustrate the criteria followed in constructing the prompts, without reproducing exhaustively all the instructions used during the process.

In practice, each storyboard panel was developed through an average of four workflow generations, after which the most useful output was selected for comparison. In the panels that required it, *inpainting* operations were used to correct localised errors or reinforce specific graphic elements (for example, the inclusion of movement arrows in a manner similar to the original storyboard). This combination of selection across generations and localised intervention reinforces the non-linear nature of the workflow: rather than recreating the entire image, the artist can reopen specific areas and modulate the proposal already obtained.

In the third phase, an attempt was made to produce a controlled variation of each base reconstruction by modifying framing, action orientation, or spatial distribution. This phase shifts the evaluation from the mere capacity for imitation towards the capacity for reformulation, which is where the principal industrial advantage of these tools lies.

In the fourth phase, simple operational indicators were recorded: workflow preparation time, average execution time per image, average number of generations per storyboard panel, occasional use of corrective *inpainting*, and success or failure in the generation of variations.

Given the exploratory nature of the study, the results were interpreted through descriptive analysis and qualitative visual comparison. A reconstruction was considered functionally successful when it fulfilled four criteria:

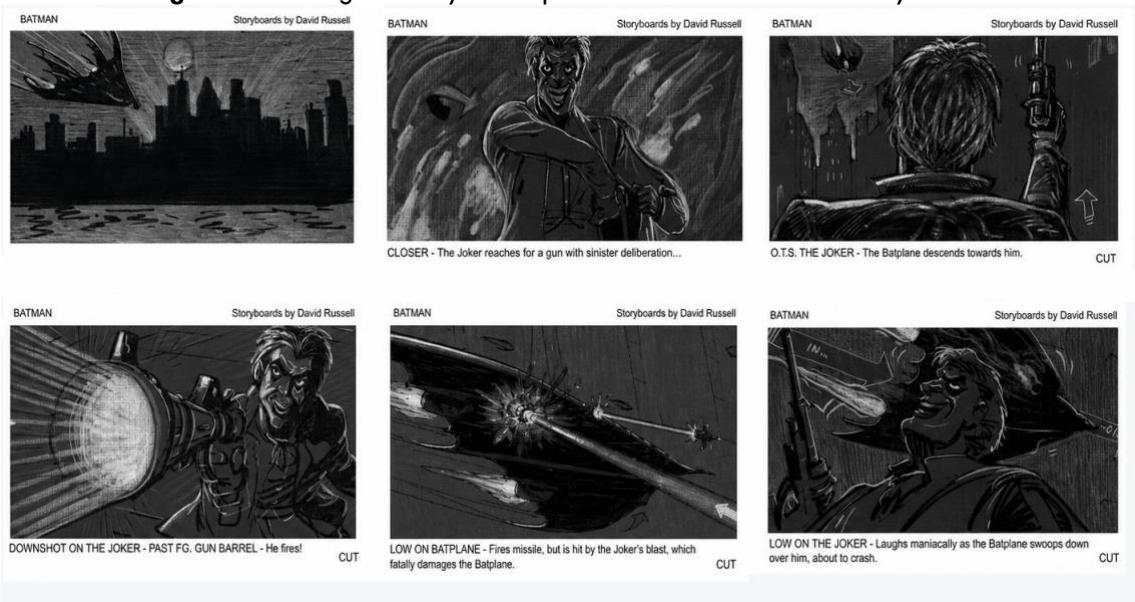
a) unequivocal identification of the shot
b) clear legibility of the action
c) maintenance of basic continuity with the rest of the sequence
d) preservation of the dominant dramatic vector of the original storyboard panel

A variation was considered satisfactory when it introduced a recognisable change in framing, composition, or visual resolution without compromising the intelligibility of the scene or rebuilding the shot logic from scratch. No external evaluators or usability scales were employed, because the aim of the article is to document the viability of an actually executed foundational workflow and to discuss its potential professional suitability.

3. Results

The base reconstruction proved viable across all six selected storyboard panels. In every case, an image was obtained that allowed recognition of the narrative moment of the scene and the principal logic of the shot. It should be clarified that the original panels correspond to the functional language characteristic of production storyboards: high contrast, graphic economy, and prioritisation of narrative information over pictorial finish. This apparent simplicity does not diminish their professional complexity but rather demonstrates that the value of the storyboard does not necessarily reside in the image as a final piece, but in its capacity to be used during production. For this reason, the comparison developed in this study does not evaluate the visual quality of the generated image against the original drawing, but rather the preservation of the planning function that the professional storyboard resolves through a deliberately synthetic graphic formulation.
Equivalence was therefore not achieved at the level of graphic style or the exact tonal aesthetic of the original storyboard (drawn in white on black with highly marked expressiveness), but at the functional level of visual planning. The workflow proved capable of translating camera positions, subject–environment relations, and general action into contemporary cinematic-looking imagery. One relevant detail of the process was that, in order to represent operationally the concept of the Batwing or Batman's aircraft, it became necessary to approximate the prompt towards the more generic concept of a "fighter aircraft". This reformulation reveals an additional difficulty of the model in maintaining highly specific iconic references when object recognition and framing variation are combined.
The panels centred on the character's gesture or the basic confrontation between the aircraft and the Joker produced the strongest results. In these cases, the model responded effectively to relatively stable prompts: aerial approach, character drawing a weapon, gaze and direction of the threat, and the aircraft passing above the character. The principal strength of the system was not stylistic fidelity, but rather the rapid production of legible compositions sufficiently close to the intention of the original shot.

Figure 2. Six original storyboard panels from the reference storyboard.



Source: David Russell, professional portfolio; contextualised reproduction in Filmsketchr (2012). Used for comparative academic analysis.

Figure 3. Six base reconstructions generated with ComfyUI from the screenplay fragment and the reference storyboard panels.

Source: Authors' own elaboration.

The controlled-variation phase proved satisfactory in four of the six storyboard panels. In those four images, the workflow made it possible to modify the size of the subject, shift the point of view, or intensify the principal gesture with relative effectiveness without losing recognition of the scene. From a pre-production perspective, this constitutes the most significant finding: once a plausible visual base had been obtained, ComfyUI functioned as a flexible platform for exploring alternative formulations of the same dramatic moment.

By contrast, the fifth and sixth storyboard panels did not allow for convincing controlled variations. The fifth required representing the impact of a projectile fired from the ground against the Batwing in mid-flight. The sixth required a situation close to imminent collision: the burning Batwing passing above the Joker within a highly specific spatial dynamic. In both cases, successive generations tended to simplify the event, replace it with more generic alternatives, or degrade the precision of the scene. The problem, therefore, was not the impossibility of generating a visually striking image, but rather the difficulty of maintaining an exact spatial and causal relation between trajectory, impact, and the state of the vehicle when variations in the shot were requested.

In both unsuccessful cases, a highly specific semantic problem was also observed. In panel 5, the model failed to interpret consistently the causal connection between the bullet fired from the Joker's gun and the impact on the Batwing. In panel 6, the representation of the Batwing in a state of loss of control likewise could not be recreated when the shot was varied.

Figure 4. Four satisfactory variations generated from the base reconstruction.

Source: Authors' own elaboration.

The workflow preparation and initial debugging time was approximately 15 minutes. The average execution time per generation was around 3 minutes on the GPU used. Each image was generated, on average, four times before the most suitable result was selected. In some variations, additional *inpainting* was employed, with an approximate cost of 3 minutes per corrective intervention, both to correct minor errors and to reintegrate arrows or other graphic indicators present in the original storyboard.

Based on this record, the operational estimate may be summarised in Table 3.

Table 3. *Workflow record*

Phase	Units	Average number of generations	Time per generation	Estimated time
Workflow preparation and debugging	1	—	—	15 min
Base reconstruction of 6 storyboard panels	6	4	3 min	72 min
Variations on 4 storyboard panels	4	4	3 min	48 min
Occasional corrective *inpainting*	2–4 interventions	1	3 min	6–12 min
Estimated total	—	—	—	141–147 min

Source: Authors' own elaboration.

Although these times do not correspond to the total working time of a professional storyboard artist (because they do not include briefing, script reading, or external validation), they do show

that an individual operator can produce, in approximately two and a half hours, a sequence composed of six reconstructions and four alternatives.

As an external reference, and not as a calculation of the actual cost of the case study, the average storyboard artist rates published by YunoJuno for the British freelance market may be considered: £50 per hour and £400 per half-day. The source also notes that rates depend on the number of shots, the number of revision rounds, the complexity of the style, and usage rights. Consequently, this benchmark can only be interpreted as an indication of comparative economic plausibility rather than as a direct equivalence (YunoJuno, 2026).

If this average half-day rate is translated in a purely indicative manner into approximately four hours of work, the equivalent external labour cost would be around £200. The workflow tested here was completed within approximately that time range, but with two essential qualifications: first, because the comparison contrasts specialised human labour with locally assisted execution; and second, because the cost of the AI workflow does not include hardware amortisation, system learning time, or the author's cinematic expertise. What the benchmark does suggest is that the most plausible competitive advantage lies not in reducing the cost of the storyboard itself, but in lowering the marginal cost of variations once the base has been constructed. The greatest value of the workflow lies less in replacing the initial version than in avoiding the need to recreate it repeatedly when alternative *mise-en-scène* options are required.

4. Discussion

The results make it possible to answer the first research question affirmatively, albeit with important nuances. It is indeed viable to obtain, using ComfyUI and a foundational workflow without LoRAs or ControlNet, a reconstruction functionally comparable to a professional storyboard within a brief and highly controlled corpus. However, this equivalence must be understood in industrial rather than mimetic terms. The system does not reproduce the graphic style of the original storyboard but instead generates an alternative visualisation that preserves the logic of the shot and the sequence. This distinction is important because it situates the professional usefulness of generative AI within the previsualisation stage, rather than in the stylistic replacement of the storyboard profession.

The second research question also receives an affirmative answer: controlled variation is possible and, in fact, constitutes the principal strength of the workflow. The fact that four of the six storyboard panels admitted satisfactory reformulations without restarting the process suggests that ComfyUI is already operational as a tool for exploring alternatives. This connects directly with the literature on AI-assisted storyboarding and previsualisation. The present study confirms this intuition within a less experimental environment and one closer to self-implemented creative practices: a single author, a local workflow, and consumer-grade hardware (Wang et al., 2024; Wei et al., 2025).

From a technical perspective, the failures observed in the fifth and sixth storyboard panels do not invalidate the workflow but rather delimit the point at which its current maturity reaches its limits. The difficulty in generating coherent variations when the image required the exact representation of an impact, a trajectory, or the imminence of a crash is consistent with what recent literature has identified as a compositional deficit in diffusion models. This occurs when problems arise in following complex instructions and representing precise spatial relations. In the present study, this limitation manifested itself clearly: the workflow was capable of producing versions close to the narrative moment, but not of rigorously preserving the shot when the framing was varied (Feng et al., 2024; Zhang et al., 2025).

This point makes it possible to qualify the most optimistic hypothesis regarding the immediate professional use of generative AI in storyboarding. If the task consists of rapidly generating shot concepts, atmospheres, or character arrangements, the workflow tested here appears sufficiently mature to be useful. If, by contrast, the sequence requires strict precision in trajectories, collisions, or highly specific physical details of the action, the foundational workflow ceases to be fully reliable and requires additional support.

It is at this point that the discussion of ControlNet and LoRA becomes relevant, albeit only as a reasoned projection. ControlNet was designed to add spatial guidance to diffusion models and therefore appears particularly suitable for scenes in which pose or the direction of an action vector cannot be left to the model's discretion. Complementarily, LoRA makes it possible to adapt the model to specific iconographic or stylistic domains at lower cost than full fine-tuning. Documentation associated with FLUX and ComfyUI further shows that the ecosystem already includes examples of fill and inpainting operations, additional depth or canny control, and even LoRA-format variants for certain workflows. However, incorporating these elements would have transformed the present study into something different: an analysis of advanced optimisation rather than a test of the operational maturity of a minimal workflow. The results therefore indicate that the foundational workflow is professionally useful for a well-defined range of tasks, whereas situations requiring high spatial precision justify the use of additional layers of control (Hu et al., 2021; Zhang et al., 2023).

From an economic and organisational perspective, the most valuable finding is likewise not the direct reduction of costs, but rather the reduction in exploration costs. This distinction aligns with Sookhom et al., who had already warned that AI does not automatically reduce expenditure unless there is a competent author and a clearly defined technical framework for its use. In the present case, the potential saving lies primarily in the possibility of rapidly achieving a useful visualisation and producing variations at low marginal cost. In a real production context, this capability may prove decisive for the different departments involved in a shoot when they need to compare alternatives before transferring tasks to specialised artistic departments (Sookhom et al., 2023).

The professional suitability of a local workflow such as the one studied here cannot be evaluated independently of the technical cost it entails. Cloud-based platforms present clear advantages in terms of accessibility and speed of use, particularly for non-technical users or environments that prioritise operational immediacy. By contrast, working with ComfyUI requires greater technical literacy, an understanding of node-based systems, and attention to potential issues associated with updates or dependencies. However, this higher entry cost is compensated for by a greater degree of control over the pipeline, model selection, process traceability, and reduced mediation by external service policies.
To this technical cost must be added the material cost of hardware. Although the present study was carried out on a consumer-grade GPU, the use of more powerful models, higher resolutions, larger generation batches, or pipelines incorporating additional layers of control may require graphics cards with greater VRAM capacity and, consequently, a higher initial investment. Professional adoption of local workflows should therefore not be evaluated solely in terms of cost per image or generation time, but also in relation to hardware amortisation and the periodic updating of equipment (ComfyUI, 2026).

This work therefore argues that the current professional suitability of ComfyUI should not be measured according to a logic of replacement. Its value lies elsewhere: in functioning as a node-based interface for a more iterative, local, controllable, and accessible form of pre-production, particularly in early stages or in projects requiring numerous preliminary visual decisions within a limited time frame. This conclusion aligns with Caballero and Sora-Domenjó's diagnosis regarding the demand for more interactive tools, as well as with the official documentation of ComfyUI and FLUX, which is specifically oriented towards editable, low-latency workflows. The present study provides practical evidence that helps situate these tools not within the realm of technological spectacle, but within that of conditional operational usefulness (Caballero & Sora-Domenjó, 2024; Comfy-Org, n.d.-a; ComfyUI, 2026).

As limitations, the present study identifies four main constraints. The first is its single-case nature. Only one sequence, consisting of six storyboard panels from a specific film and a highly singular action scene, was analysed. The results therefore cannot be extrapolated without caution to other genres, styles, or types of storyboards. The second limitation is the absence of evaluation or validation by external professionals. The comparison was conducted through descriptive qualitative analysis, which is appropriate given the exploratory nature of the study. The third

limitation is the dependence on the available hardware and the selected model. The observed times refer to a specific GPU and a specific FLUX variant. With different resources, performance could vary. The fourth limitation is that the study deliberately focuses on a foundational workflow. It does not incorporate LoRAs, ControlNet, or other layers of control that would probably improve some of the problematic results. This exclusion constitutes a methodological strength for the objective pursued, but it also limits the maximum achievable quality.

A natural extension of this work would consist in incorporating additional mechanisms of control and specialisation in order to measure the extent to which they increase interdepartmental consistency, as well as expanding the sample to include different styles of source imagery and different genres. In a real professional context, pose-, edge-, or depth-based ControlNet systems could be employed to ensure more stable spatial relations, alongside style LoRAs linked to materials already validated by the project itself (for example, concept art, production design, or visual bibles approved by the art department), with the aim of bringing the generated storyboard closer to the aesthetic ecosystem of the other departments. Methodologically, this possibility does not weaken the present study but rather reinforces its value as a baseline against which future research could measure increases in control, visual coherence, and industrial suitability. It also points towards the consolidation of hybrid professional profiles situated between the technical artist, the pipeline integrator, and the visualisation specialist, capable of translating creative needs into concrete infrastructures.

5. Conclusions

The study makes it possible to formulate four main conclusions. The first is that a foundational ComfyUI workflow executed locally, using *image-to-image* and *inpainting* operations and without advanced control modules, already makes it possible to reconstruct a brief professional storyboard sequence in a functionally comparable manner. This places the tool at a sufficient level of maturity for real early-stage previsualisation uses. The second conclusion is that the principal industrial value of the workflow does not lie in the exact imitation of the reference storyboard, but in the possibility of building upon it. The generation of satisfactory variations in four of the six storyboard panels demonstrates that generative AI can function as a mechanism for revision and reformulation, which is precisely where the greatest temporal friction exists within pre-production practice. The third conclusion is that the operational maturity of the workflow is uneven. It performs more effectively when the shot is organised around clear gestures and relatively stable spatial relations. It becomes weaker when the image requires high-resolution causal and spatial precision, such as highly specific transitional states of the represented object. The fourth conclusion is that this limitation does not invalidate the professional suitability of ComfyUI but rather defines its optimal zone of use. For rapid storyboarding, sequential conceptualisation, framing exploration, and the testing of alternatives, the tool is already useful and manageable for an individual operator. For sequences in which spatial accuracy or iconic consistency are critical, however, it appears advisable in future to incorporate additional control mechanisms such as ControlNet or lightweight adaptations through LoRA, always weighing the added cost in complexity.

In general terms, the study does not advocate the replacement of the storyboard artist, but rather a reassignment of functions within the pre-production workflow. Generative AI, deployed through ComfyUI, emerges here as a sufficiently mature instrument for accelerating the phase of visual exploration and opening an intermediate zone between the screenplay and the finished storyboard. That zone, historically costly in terms of time and revisions, is precisely where the tool currently demonstrates its clearest professional suitability.

References

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

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). *Storyboard*ing: 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

AUTHOR CONTRIBUTIONS, FUNDING AND ACKNOWLEDGEMENTS

Author contributions:
Conceptualisation: Mejías Martínez, Guillermo and Blanco, Sonia
Software: Mejías Martínez, Guillermo
Validation: Mejías Martínez, Guillermo and Blanco, Sonia
Methodology: Mejías Martínez, Guillermo and Blanco, Sonia
Data curation: Mejías Martínez, Guillermo
Writing – original draft preparation: Mejías Martínez, Guillermo
Writing – review and editing: Mejías Martínez, Guillermo and Blanco, Sonia
Visualisation: Mejías Martínez, Guillermo
Supervision: Blanco, Sonia
Project administration: Mejías Martínez, Guillermo and Blanco, Sonia
All authors have read and agreed to the published version of the manuscript: Mejías Martínez, Guillermo and Blanco, Sonia.
Funding: DesinfoperIA. Journalistic applications of AI to reduce disinformation: trends, uses, and perceptions among professionals and audiences. Reference: PID2023-147486OB-I00, funded by the Spanish Ministry of Science and Innovation.
Acknowledgements: gen-IA: Library of artificial intelligence tools for content creation in media communication. Ref.: JA.B3-23. PPRO-B3-2023-03. Funded by the II Internal Plan for Research, Knowledge Transfer, and Scientific Dissemination. University of Málaga.
Conflict of interest: none.