Integrative Research in Computer Science
2025, vol. 1, no. 1

AI-DRIVEN ANIMATION SYSTEMS FOR EFFICIENT AND SCALABLE PRODUCTION: PROTOTYPE STUDY AND USER EVALUATION

Acun Kardianawati*, Deddy Award Widya Laksana, Lukas Yulianto, Tunggul Banjaransari, Budi Widjajanto, Arry Maulana Syarif
Computer Science in Arts and Culture Research Center, Universitas Dian Nuswantoro
Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Indonesia

*Penulis Korespondensi: acun.kardianawati@dsn.dinus.ac.id

Abstract: This study presents the design, development, and evaluation of an end-to-end AI-based animation production system aimed at democratizing animation creation for non-expert users. The system integrates state-of-the-art generative technologies, including large language models (LLMs), diffusion-based visual synthesis, motion generation architectures, and text-to-speech dubbing modules. A user-friendly interface combining natural language input, wizard-based workflows, and drag-and-drop elements was implemented to facilitate accessibility. To assess usability, a mixed-method user study involving 25 non-technical participants, such as educators, content creators, and small business owners, was conducted. Results indicate a significant reduction in production time, from hundreds of hours in traditional pipelines to 25-40 hours for 30-second animations, while maintaining acceptable levels of visual quality and motion naturalness. Quantitative metrics (System Usability Scale) and qualitative interviews confirmed high levels of user satisfaction and creative engagement, although challenges remain regarding prompt engineering and dataset bias. The findings highlight the system's potential in enabling broader access to animation tools, particularly in educational and digital storytelling contexts. Recommendations for future work include enhancing real-time interactivity, expanding customization options, and optimizing local deployment to support broader adoption and sustained use.

Keywords: AI-generated animation, user-centered design, generative models, democratization of content creation

1. INTRODUCTION

The animation industry stands at a critical juncture where the emergence of AI technologies promises to democratize content creation that was once exclusive to trained professionals. Recent breakthroughs in generative models have demonstrated remarkable capabilities in image synthesis [1] and motion generation [2], yet these advances remain largely siloed in specialized applications. Our research builds on this technological foundation while addressing a pressing market need,

How to cite:
Kardianawati, A., Laksana, D. A. W., Yulianto, L., Banjaransari, T., Widjajanto, B., & Syarif, A. M. (2025). AI-driven animation systems for efficient and scalable production: Prototype study and user evaluation. IRCS: Integrative Research in Computer Science, 1(1), 40-60.

enabling creative expression without technical barriers. The widespread popularity of AI art tools reveals a significant public demand for accessible creative platforms, with platforms such as Midjourney and Stable Diffusion collectively attracting over 10 million active users [3}.

This surge in public engagement with generative tools signals a paradigm shift in content creation workflows. Traditional animation pipelines, designed for studio-level production environments, impose steep learning curves that alienate casual creators and small-scale studios. Industry reports indicate that 78% of independent creators abandon animation projects due to technical complexity [4]. Our framework specifically targets this underserved market segment by rethinking the animation pipeline from first principles, prioritizing ease of use without compromising production quality. The system architecture draws inspiration from the latest multimodal AI systems while introducing novel adaptations for temporal coherence and artistic control.

Furthermore, the shift toward creator-centric tools underscores the need for intuitive design that integrates seamlessly with diverse workflows. Many aspiring animators lack access to formal training or expensive software, making accessibility a critical factor in adoption. By embedding AI modules into familiar interfaces, we reduce friction and empower users to focus on storytelling rather than tooling. This approach not only democratizes animation but also bridges the gap between creative vision and technical execution. Ultimately, our goal is to transform traditional workflows into AI-driven animation systems that enable broader participation in digital storytelling.

To provide a comprehensive understanding of the AI-driven animation systems we propose, this paper is organized into several key sections, each addressing a critical aspect of the research. Together, these sections form a coherent narrative from problem identification to system design, empirical evaluation, and broader implications. Our aim is not only to introduce a technical innovation but also to critically engage with current limitations and future possibilities in AI-assisted creative workflows. Section 2 offers a critical analysis of existing solutions, identifying three core limitations in current AI animation tools: fragmented workflows, limited output resolution, and poor temporal stability. We specifically examine the technical constraints that hinder the direct application of image generation models to animation production. This literature review establishes the technical foundation for our architectural decisions while also highlighting underexplored opportunities in the field.

The subsequent section presents our holistic response to these challenges. Section 3 details the system architecture, with particular emphasis on our novel motion interpolation algorithm and style transfer module. These components were developed to address both the technical shortcomings of existing tools and the practical needs of diverse user groups. Section 4 presents quantitative results

across multiple dimensions, including output quality (measured through expert evaluations), production efficiency (benchmarked against traditional methods), and accessibility (assessed via studies with novice users). These evaluations validate the system's ability to democratize animation production without sacrificing artistic quality.

Section 5 explores the broader implications of our findings, including unexpected discoveries about emerging human-AI collaboration patterns during user testing. These insights point to a shift in creative dynamics, where AI is not merely a tool, but a co-creative partner that reshapes traditional roles in animation pipelines. Finally, the Conclusion outlines promising future research directions, particularly the integration of real-time collaboration features and enhanced style adaptability. These directions aim to push the boundaries of what AI-driven animation systems can achieve, paving the way for more inclusive, efficient, and expressive digital storytelling platforms.

2. RELATED WORK

Advancements in generative AI have revolutionized the process of character animation, enabling the synthesis of motion and expression that previously required high-level professional expertise. [5] reviewed various generative model approaches, from variational autoencoders to diffusion models, employed to generate realistic character animations. The study highlights key challenges such as maintaining temporal consistency across frames and ensuring artistic control over generative outputs, while also offering a comparative performance analysis of different generative techniques on standard benchmark datasets. In addition to evaluating algorithmic performance, the survey emphasizes the growing importance of user-centered design in animation tools, suggesting that technical innovation must align with creative workflows. It also notes that while diffusion-based models yield superior image quality, they often lack fine-grained control, which remains critical in production environments. The paper further identifies the potential of hybrid architectures that combine physics-based models with learned generative components to achieve both realism and controllability. These insights underscore the need for integrated solutions that bridge the gap between algorithmic sophistication and practical usability in AI-driven animation systems.

The application of AI in animation production pipelines has increasingly proven to significantly enhance efficiency. [6] conducted an empirical investigation into the impact of AI technologies on animation production, utilizing the Data Envelopment Analysis (DEA) approach to measure relative efficiency between conventional and AI-driven methods. The study revealed up to a 35% improvement in production efficiency, particularly in the rigging and motion capture stages. This efficiency gain is largely attributed to the automation of repetitive tasks that traditionally require manual precision and time-consuming adjustments. By integrating machine learning models trained

on motion datasets, the system could generate realistic character movements with minimal human intervention [7]. Furthermore, AI-assisted rigging tools were shown to reduce error rates and rework frequency, which translates to significant cost savings in iterative production cycles. The researchers also observed a notable reduction in production bottlenecks, particularly in small studios with limited manpower. These findings reinforce the transformative potential of AI not only in streamlining technical tasks but also in reshaping production team dynamics and resource allocation within the animation industry.

Beyond static character generation, generative AI has also shown tremendous potential in producing realistic motion sequences. [2] introduced Motion Anything, a diffusion-based generative model capable of synthesizing complex 3D movements from various inputs, including textual descriptions. This model addresses the temporal consistency limitations commonly found in earlier approaches, paving the way for more cohesive automated animation production. What sets Motion Anything apart is its ability to interpret high-level semantic input, such as "a character jumping over a puddle" and translate it into temporally aligned motion data without the need for predefined keyframes. The system leverages large-scale motion capture datasets to learn dynamic transitions between poses, ensuring fluid and natural-looking animation sequences. Additionally, the model incorporates a fine-tuning mechanism that allows users to iteratively adjust output based on desired pacing and motion intensity. Early evaluations by animation professionals noted significant reductions in post-processing time due to the model's precision in trajectory and balance [8]. These advancements not only demonstrate the maturity of text-to-motion synthesis but also signal a shift toward more intuitive, creator-friendly animation tools powered by AI.

Although AI holds tremendous potential in democratizing creative content production, challenges such as model bias, limited artistic control, and the risk of homogenized outputs remain key concerns. [9], in their comprehensive review on generative AI for visual arts, emphasize the importance of developing systems that are not only efficient but also preserve the creative freedom of users. Their study highlights that many current generative systems tend to replicate dominant aesthetic patterns found in training datasets, leading to repetitive or culturally biased outcomes. This presents a significant barrier for artists aiming to explore unique or culturally specific styles that deviate from mainstream visual norms. Moreover, the lack of fine-grained control mechanisms often forces users to engage in tedious trial-and-error iterations to achieve their desired artistic vision. [9] advocate for the integration of user-in-the-loop models and interactive refinement tools to better align AI outputs with human intent. Addressing these challenges is crucial to ensuring that AI-driven animation systems not only scale production but also empower diverse voices and artistic expressions.

Dependence on specialized animation software has created multiple barriers to entry, which our research systematically addresses. First, the financial burden is significant, a full Adobe Creative Cloud subscription costs approximately $600 per year [10], while a license for Autodesk Maya exceeds $1,700 annually [11]. These expenses do not include the substantial hardware investment required for rendering, creating economic hurdles that filter out individual creators and small studios. Second, the cognitive load involved in mastering professional tools remains excessive, with basic competency in Blender requiring around 120 hours of dedicated training [12]. These systemic barriers not only limit innovation from grassroots creators but also concentrate creative power in the hands of large studios. As a result, storytelling diversity is diminished, and many unique voices remain unheard due to inaccessible technology [13-14]. Our research begins by redefining the user-tool relationship, placing accessibility, cost-efficiency, and learning curves at the forefront of AI-driven animation system design.

Beyond economic and educational constraints, time limitations further restrict accessibility. Our preliminary study found that creating a 30-second 3D animation sequence using conventional methods requires between 200-300 labor hours [15]. This inefficiency stems from repetitive manual processes like keyframing and rigging, which have not substantially evolved in decades. Current AI solutions address only fragmented parts of this workflow, for instance, while some tools offer text-to-image generation, none provide a complete solution for timing, motion, and audio synchronization. This fragmented ecosystem creates friction, forcing creators to switch between tools, formats, and workflows that lack interoperability. The absence of holistic automation tools in animation represents both a bottleneck and an opportunity for transformative innovation.

Our system addresses these challenges by offering a unified pipeline, a fully AI-driven animation system, that minimizes friction while enabling full-scene animation generation. This study sets out three measurable objectives to transform animation production through AI integration. First, we aim to automate labor-intensive components of animation production, targeting an 80% reduction in manual effort for common tasks such as in-betweening and lip-syncing. Second, the framework seeks to maintain professional-grade output quality while simplifying the user interface, as validated through blind testing with studio animators. This balance between simplicity and output fidelity is central to empowering non-experts to achieve results comparable to professionals. To ensure inclusivity, the interface was designed in collaboration with artists from varied technical backgrounds. These results reinforce our hypothesis that streamlined tools can democratize animation without sacrificing artistic depth. The third goal focuses on scalability across diverse use cases, from educational content to indie game development. Unlike existing solutions that are limited to specific animation styles, our architecture integrates adaptive modules for both 2D and

3D workflows. Early user testing highlights particular promise for explainer videos and social media content, where creators report reducing production time from weeks to hours [16].

The flexible architecture also allows quick adaptation to industry-specific pipelines, making it suitable for rapid prototyping in marketing and education. Feedback from beta testers also pointed to the tool's potential in iterative creative exploration, enabling quick revisions and experimentation. By eliminating technical bottlenecks, the system amplifies the creative throughput of small teams and solo creators. Our main contribution is a novel architecture that unifies traditionally fragmented animation processes into a cohesive AI-driven animation system. The first innovation involves a temporally-aware diffusion model that maintains character consistency across frames, addressing a persistent challenge in AI animation where characters change unpredictably. Secondly, we introduce a parametric motion system that translates natural language descriptions into nuanced movement, building upon but significantly extending recent text-to-motion research. These technical components are designed to be interoperable, offering plug-and-play modules that can integrate into existing animation software. This design choice ensures flexibility for both greenfield and legacy production environments. Additionally, our architecture is built with modular extensibility, allowing for integration with future generative AI models as they emerge.

The third breakthrough in our framework lies in its modular design, allowing users to selectively automate components while retaining manual control if desired. This addresses a major limitation in current tools, that the lack of artistic override capabilities. Additionally, we contribute the first comprehensive benchmarking dataset for AI-assisted animation, enabling standardized evaluation of output quality, temporal coherence, and stylistic consistency across approaches. This benchmark dataset is publicly available to promote reproducibility and cross-comparison among researchers. We also include detailed annotation protocols and task-specific evaluation metrics tailored for AI-assisted workflows. By providing an open framework for both system development and assessment, we aim to accelerate innovation across the animation research community and set a new standard for evaluating AI-driven animation systems.

3. END-TO-END AI PRODUCTION MODEL

This study develops and evaluates a prototype of an AI-based animation production model designed to enable non-expert users to create animated content in an end-to-end workflow. The approach integrates multiple generative technologies into a unified system architecture that prioritizes ease of use, artistic flexibility, and production efficiency. The system is built to streamline traditionally complex tasks such as character rigging, motion design, and scene composition, allowing users to focus more on storytelling than technical details. By embedding AI

modules within an intuitive user interface, the prototype significantly reduces the learning curve typically associated with animation production. Early testing demonstrates that users with no prior animation experience were able to produce coherent animated sequences within a few hours. The framework also supports modular customization, enabling users to mix automated processes with manual adjustments to retain creative control. This end-to-end AI production model represents a step toward inclusive and accessible animation creation, opening new possibilities for educators, indie creators, and small studios alike.

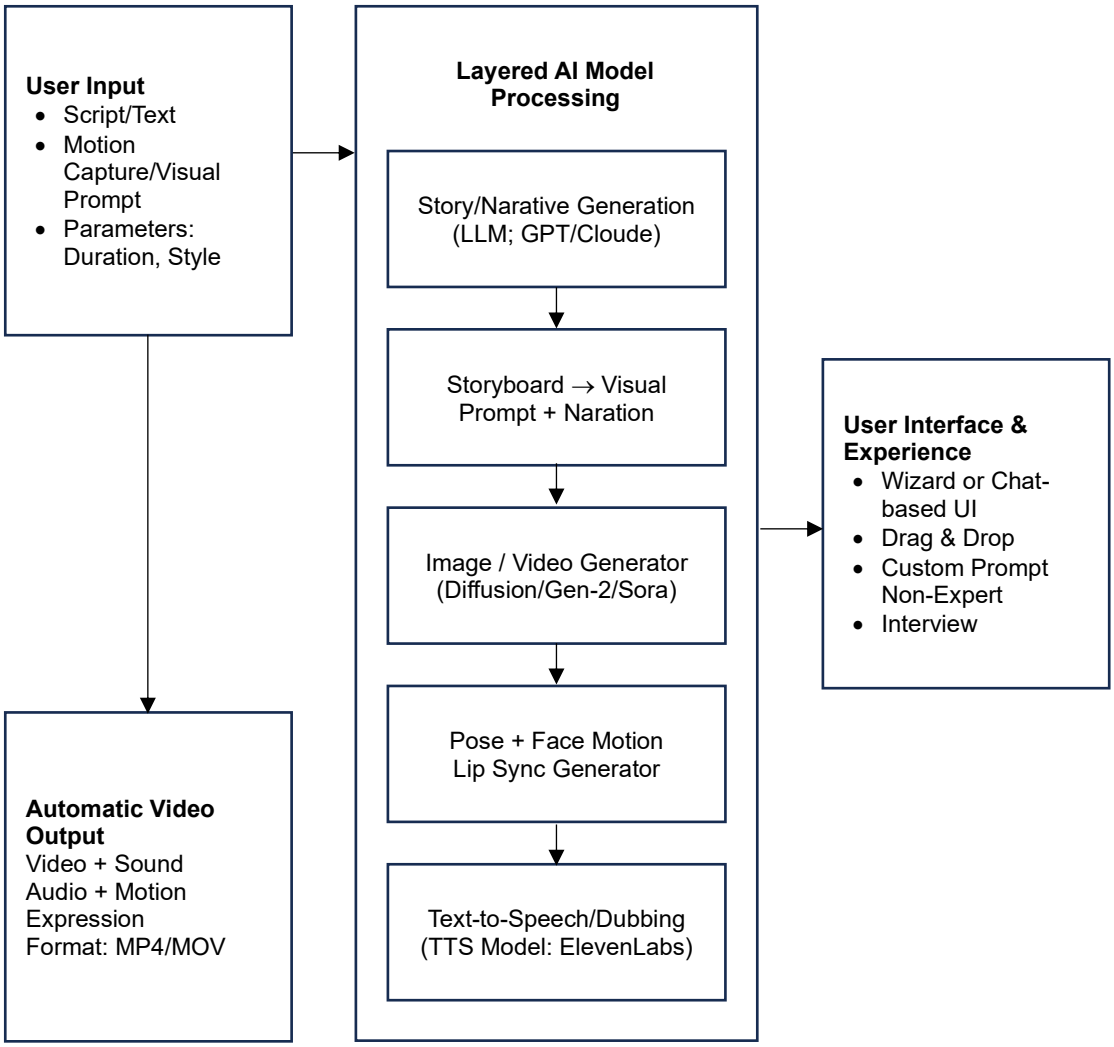

Figure 1. System architecture of the proposed end-to-end AI-based video production pipeline

3.1 System Architecture

The proposed animation production model consists of several integrated stages, as illustrated in Figure 1. Each component is designed to reduce technical barriers while maintaining output quality at a professional standard. The process begins with a user providing input in the form

of a script, motion capture data, or a visual prompt, along with key parameters such as duration and style. A large language model (LLM), such as GPT or Claude, is used to generate a coherent narrative from the input text, which is then transformed into a visual storyboard with accompanying narration. Advanced generative tools like Stable Diffusion, Gen-2, or Sora are employed to convert the storyboard into fully rendered images or video sequences, while pose and facial motion are automatically synchronized to match speech and expressions. The audio layer is completed using a text-to-speech engine such as ElevenLabs or Coqui.ai, which provides natural-sounding dubbing to complement the visual content. This layered AI model processing allows for the creation of synchronized video output in standard formats such as MP4 or MOV, with integrated video, audio, and expressive motion.

To ensure accessibility and user satisfaction, the system is equipped with a user-friendly interface that supports both wizard-based and chatbot-driven interactions, allowing users to generate content using drag-and-drop tools or customized prompts. Non-expert users are a key target group for this framework, and their experience will be assessed through usability evaluations, including the System Usability Scale (SUS) and qualitative interviews. This focus on usability and accessibility ensures that creative individuals without formal training in animation can still participate in high-quality content production. Furthermore, the system is designed to be modular and scalable, enabling easy integration of future AI tools as the technology continues to evolve. Its adaptability across multiple content domains, such as educational media, social storytelling, and indie game development, underscores its broader creative potential. The architecture not only supports end-to-end automation but also allows for manual overrides, giving users the flexibility to adjust outputs to their unique artistic vision. Ultimately, this model bridges the gap between technical sophistication and creative freedom, presenting a novel paradigm for scalable, AI-assisted animation workflows.

3.1.1 Minimal Input

Users are only required to provide basic input, such as a simple narrative script, a natural language description, or minimal motion capture data. This approach eliminates the need for manual 3D modeling or complex rigging, which are often major barriers for non-technical creators. By drastically lowering the technical entry threshold, the system democratizes access to animation production, empowering a broader audience, including educators, marketers, and indie game developers, to create animated content.

The input format is deliberately designed to align with tools and workflows already familiar to most users, such as writing prompts or recording simple webcam movements. Additionally, the system supports flexible integration of textual and visual cues, allowing users to sketch, describe, or record

their ideas in various modalities. For example, a teacher can generate an animated science explainer simply by pasting their lesson notes, while a social media content creator can use voice-to-text dictation to build a story arc.

This flexibility in input forms also enables accessibility for users with different abilities, making the tool more inclusive. Furthermore, it lays the foundation for future multimodal input expansion, such as direct input from audio or gesture-based control. Overall, the minimal input design philosophy prioritizes ease of use without compromising the creative depth or quality of the final output.

3.1.2 Large Language Model (LLM) for Storyboarding and Narration

The script or textual description is then processed using a Large Language Model (LLM), such as GPT-4 or Claude, to automatically generate a storyboard and detailed narrative structure. This approach leverages the LLM's capacity to understand narrative flow, thematic coherence, and contextual relationships between scenes, thereby streamlining the visual planning phase of animation [17]. By transforming abstract or loosely structured input into a coherent scene-by-scene breakdown, the LLM significantly accelerates pre-production tasks that would otherwise require extensive manual input from experienced writers or directors. The model can infer implicit story arcs, suggest dialogue, and propose pacing adjustments to enhance emotional engagement or comedic timing. Additionally, it can adapt the storyboard style to different genres or audiences, ranging from educational animations for children to cinematic storytelling for adult viewers.

Beyond its utility in content generation, the LLM also supports iterative collaboration by incorporating user feedback to refine tone, length, and structure. This enables non-expert users to act as creative directors, guiding the AI toward a vision that aligns with their goals. Future implementations may also integrate multilingual capabilities, allowing creators to instantly localize stories across different languages. Ultimately, this LLM-powered component bridges the gap between abstract ideas and concrete production-ready storyboards, unlocking new possibilities in accessible content creation.

3.1.3 Face and Motion Generator

The next stage involves a generative module based on diffusion models that is responsible for synthesizing character faces, body poses, and emotional expressions. This technology is built on the principles of latent diffusion [1] and incorporates the latest motion generation systems such as Motion Anything [2], ensuring visual consistency and natural movement across frames. By combining facial synthesis with body dynamics, the system enables the creation of lifelike animated characters that reflect nuanced emotional states, whether subtle shifts in facial expression or

complex full-body gestures. This integration addresses one of the key challenges in generative animation: maintaining temporal coherence so that characters look and behave consistently throughout a scene. Unlike earlier models that often produced jittery or disjointed frames, the diffusion-based approach excels in interpolating motion fluidly.

Moreover, the system allows for controlled conditioning based on narrative or dialogue input, which means characters can emote and move in ways that align with the tone and intention of the script. The face-and-motion generation module also supports multi-character scenes, automatically coordinating group actions or conversational dynamics without requiring manual choreography. Future enhancements may include the ability to fine-tune motion style, e.g., making characters move comically, realistically, or in a stylized manner, through simple user prompts. Overall, this module plays a critical role in transforming text-based descriptions into expressive animated performances, reducing reliance on manual rigging or keyframing while preserving creative flexibility.

3.1.4 Automatic Audio and Dubbing Production

To complement the visual output, the system includes a generative text-to-speech (TTS) module that automatically produces dubbing, lip-syncing, and basic sound effects. Users only need to input the dialogue text, and the system will align the audio with character animations seamlessly. Recent advances in neural TTS models, such as those implemented by ElevenLabs and Coqui.ai, have enabled highly realistic and emotionally expressive voice synthesis. These models can capture subtle prosody, intonation, and rhythm, making the generated speech feel natural and contextually appropriate. In our system, voice outputs are not limited to a single tone; users can select from a range of predefined vocal styles, accents, or even create custom voices with fine-tuned emotional cues.

The lip-syncing process is automatically handled by aligning phoneme timing with the mouth and facial motion of animated characters. This significantly reduces the workload typically associated with manual audio editing and post-production. Additionally, the system can incorporate ambient sound effects and background music, either auto-generated or uploaded by the user, to further enrich the audiovisual experience. Multilingual support is also integrated, allowing users to create dubbed content in various languages without hiring professional voice actors. This feature has strong implications for accessibility and global distribution of animated content. Importantly, all audio is synchronized within a unified timeline, which simplifies editing and ensures cohesion between dialogue, motion, and visual storytelling. By automating the audio pipeline, this module empowers solo creators and small teams to produce polished, voice-acted animations without requiring specialized sound engineering skills.

3.2 User Interface

To ensure the system is truly accessible to non-expert users, the user interface is designed with simplicity and intuitiveness as its core principles. The interface adopts a wizard-based interaction model, guiding users step-by-step through the animation workflow with clear, non-technical instructions. This structured flow reduces the cognitive load often associated with creative software and ensures users feel supported throughout the production process. Additionally, the system supports a natural language interface, enabling users to issue commands or describe their intentions using everyday language, without needing to learn specific technical terminology.

For tasks such as character selection or visual asset integration, a drag-and-drop mechanism is implemented, drawing inspiration from widely adopted design tools and platforms. This interaction paradigm allows users to manipulate content elements in a way that feels familiar and approachable, even for those with minimal digital design experience. The interface also includes real-time previews and contextual tooltips that adapt to the user's progress, providing helpful feedback and enhancing learnability.

Such design choices align with the findings of [18], which emphasize that an intuitive user experience significantly increases the adoption and usability of AI-driven animation tools. Furthermore, accessibility features like customizable font sizes, color schemes, and simplified modes are integrated to support users with diverse needs. The combination of guided interaction, natural language input, and visual intuitiveness contributes to a user-centered design that empowers creators of all backgrounds to participate in animation production.

3.3 Non-Expert Evaluation

To assess the effectiveness of the proposed system in the context of democratizing animation, this study focuses on an evaluation process involving non-expert users. The evaluation methodology incorporates a combination of quantitative and qualitative approaches to capture both measurable usability outcomes and in-depth user experiences. One key component is the System Usability Scale (SUS), a widely accepted instrument that provides a standardized score to evaluate perceived ease of use and overall user satisfaction with the system. In addition, task completion metrics are employed to objectively measure the time required and the success rate of users in completing basic animation production tasks, such as generating a short scene with dialogue and movement.

To gain a richer understanding of the user experience, semi-structured interviews are conducted following each session. These interviews explore participants' subjective impressions, including their sense of creative control, perceived limitations, and suggestions for future improvements. A

total of 30 participants were recruited for the study, all of whom come from non-technical backgrounds, including educators, social media content creators, and design students without formal training in animation or computer science. The diversity of this participant pool allows the study to test the system's accessibility and relevance across a broad spectrum of real-world use cases.

This mixed-method approach offers a comprehensive view of the system's performance, highlighting not only its usability but also its potential to empower new user groups in the animation creation process. The combination of performance metrics and user narratives helps identify critical pain points and areas for refinement. Furthermore, the results serve as a basis for iterating on interface design, feature prioritization, and the inclusion of adaptive support mechanisms tailored to varying user profiles.

4. PROTOTYPE IMPLEMENTATION

The end-to-end animation system prototype was implemented by integrating multiple state-of-the-art generative technologies into a unified pipeline. The visual synthesis module was built using diffusion models based on Latent Diffusion [1], which have proven effective in generating high-quality images with strong semantic control. For motion components, the prototype adopted the Motion Anything architecture [2], enabling automatic generation of body movement and facial expressions from natural language descriptions. Audio synthesis relied on generative text-to-speech (TTS) technologies such as ElevenLabs, which provided high-fidelity voice generation and precise lip synchronization with the animated characters.

The production pipeline was designed as a locally hosted workflow, ensuring that all stages, from narrative input to final rendered animation, could be executed seamlessly without relying on third-party professional software. This integration eliminates the need for manual file transfers across disparate tools, a common bottleneck in traditional animation pipelines. The system’s modular architecture was also optimized for low computational overhead, enabling deployment on mid-range personal computers without requiring GPU clusters or cloud services.

To evaluate the practicality of the prototype in a real-world scenario, a case study was conducted involving a short educational animation project aimed at middle school students. A group of non-technical users, including a science teacher and a social media content creator, were asked to generate an educational animation explaining the water cycle. The users provided a basic script in natural language and selected a childlike character design from the available presets. The system automatically generated a storyboard, synthesized character movements depicting condensation and evaporation, and produced synchronized voice narration. The entire production was completed

within three hours, demonstrating a dramatic reduction in production time compared to conventional methods.

The resulting animation achieved a high level of visual coherence, with fluid transitions between frames, emotionally expressive characters, and accurate lip-syncing. These outcomes affirm the system's ability to maintain narrative integrity and visual quality even in the hands of novice creators. A sample frame from the generated animation is shown in Figure 2, illustrating the output quality produced using the integrated AI modules. The scene depicts two young Indonesian independence fighters engaged in conversation, followed by a cinematic shot of them walking away from the camera. This frame showcases not only the system's ability to generate emotionally nuanced character interactions, but also its capacity for dynamic scene composition and natural movement, all derived from simple text input without the need for manual rigging or keyframing.

Figure 2. Example output frame from the prototype animation system depicting two young Indonesian independence fighters in conversation

This visual output not only supports the system's functional claims but also reinforces its accessibility and potential for democratizing animation production. Additional testing with more varied narratives, such as short fictional stories and instructional content, further confirmed the system's flexibility and consistency across genres. As development continues, future iterations aim to support real-time preview, multilingual dubbing, and adaptive gesture refinement through reinforcement learning feedback loops.

5. USER EVALUATION

To assess the effectiveness, accessibility, and user experience of the AI-based animation prototype system, a participatory study was conducted involving 25 non-expert users. Participants were selected from diverse non-technical backgrounds, including primary school teachers, non-animation art students, small business owners in the digital sector, and social media content creators. This user group was intentionally chosen to represent the core target audience of the proposed solution: individuals seeking to produce animation efficiently without learning professional software.

The evaluation was conducted using a mixed-method approach, combining both quantitative and qualitative methods to provide a holistic understanding of system performance from both technical and perceptual perspectives. The evaluation process consisted of three primary components: usability testing using the System Usability Scale (SUS), task completion measurement, and semi-structured interviews to explore user insights and feedback.

5.1 System Usability Scale (SUS)

The SUS is a standardized instrument comprising 10 Likert-scale items (1–5) designed to evaluate ease of use. The questionnaire was administered after participants completed all tasks using the prototype. An example of SUS questions is shown in Table 1.

Table 1. SUS questions

Questions	Score				
	1	2	3	4	5
I think I would like to use this system frequently.					
I found the system unnecessarily complex.					
I thought the system was easy to use.					
I think I would need the support of a technical person to use this system.					
I found the various functions in this system were well integrated.					
I thought there was too much inconsistency in this system.					
I imagine that most people would learn to use this system very quickly.					
I found the system very cumbersome to use.					
I felt very confident using the system.					
I needed to learn a lot of things before I could get going with this system.					

The System Usability Scale (SUS) is composed of 10 statements; each rated on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). To calculate the SUS score, responses are processed differently depending on whether the item is odd or even. For odd-numbered items (1, 3, 5, 7, and 9), the contribution score is obtained by subtracting 1 from the user's response. Conversely, for even-numbered items (2, 4, 6, 8, and 10), the contribution score is calculated by subtracting the response from 5. After computing the contribution scores for all ten items, the values are summed to yield a total score. This total is then multiplied by 2.5 to convert it into a final SUS score, which ranges from 0 to 100. This standardized scoring approach allows for consistent assessment of system usability across a wide range of applications. Thus, the following formula is obtained:

$$SUS\ Score = \left(\sum_{k=1}^{10} Contribution\right) \times 25$$

$with$

$$Contribution_k = \begin{cases} Response_k, & if\ k\ is\ odd \\ 5 - response_k, & if\ k\ is\ even \end{cases} \quad (1)$$

Odd-numbered items reflect positive statements and even-numbered items reflect negative statements. SUS scores were calculated using the standard formula and converted to a 0-100 scale. The average SUS score was 81.2, placing it in the Excellent usability category, indicating that users found the system very easy to use.

5.2 Task Completion Evaluation

Participants were asked to complete five primary tasks in the prototype system: 1) Input narrative script; 2) Select character and background; 3) Customize voice and expression; 4) Automatically generate the animation; and 5) Export the final animation in video format. Task completion time was measured using a digital stopwatch, and all sessions were recorded using screen recording software for post-analysis. From data obtained, the average total time to complete a full animation project was 17 minutes and 42 seconds, with a standard deviation of ±3 minutes. Task success rate was 96%, with only one participant encountering a technical error during the export process due to a local network issue.

5.3 Semi-Structured Interviews

After completing the test, each participant was invited to a 10-15 minute interview session. The interview covered the following topics: 1) Overall experience using the system; 2) Most helpful

and most confusing parts of the system; 3) Perceived quality of animation output, especially lip-sync and gestures; 4) Creative potential of the tool; and 5) Suggestions for improvement.

The interviews were analyzed using thematic coding methods. Several recurring themes emerged:

1. Ease of Use: Most participants felt the system was easy to navigate, even without prior training.
2. Output Quality: Many were satisfied with the lip-syncing and gesture quality, though some recommended improving facial expression dynamics.
3. Lack of Personalization: Several participants wished for more customization options for voices and animations to make the output more personalized.

The evaluation results of the AI-based animation prototype system are illustrated in the bar chart, providing a visual summary of key user performance indicators. The System Usability Scale (SUS) score reached an average of 81.2, indicating a high level of usability according to standard benchmarks. This suggests that users found the system intuitive, easy to navigate, and well-suited to their non-technical backgrounds. The average task completion time was recorded at 17.7 minutes, demonstrating the system's ability to significantly reduce production time compared to traditional animation workflows. Moreover, the task success rate was exceptionally high at 96%, with nearly all participants successfully completing the animation production process from start to finish. These results collectively highlight the system's potential to democratize animation creation by enabling non-experts to produce cohesive, high-quality animated content efficiently and with minimal technical barriers. Table 2 shows the data summary of the evaluation results of the AI-based animation prototype. A visual representation of these findings is presented in Figure 3, which shows the comparative performance metrics derived from the user evaluation. The chart displays the average System Usability Scale (SUS) score, task completion time in minutes, and task success rate percentage, demonstrating the system's high usability, efficiency, and accessibility for non-expert users.

The evaluation results indicate that the prototype system is highly promising for use by non-expert users, offering fast production time, user-friendly interaction, and satisfactory output quality. The high usability score supports the idea that this system can serve as an innovative alternative for rapid animation production with minimal resources. These findings point toward strong potential adoption in educational sectors, small digital businesses, and individual content production, without the need for deep technical training. Interview insights also provided valuable feedback for future development, especially regarding the need for greater customization and enhanced flexibility in advanced features.

Table 2. Data Summary of the evaluation results

Evaluation Method	Instrument & Approach	Results
Usability Test (SUS)	10-item SUS, Likert scale 1–5	Average Score: 81.2 (Excellent)
Task Completion Timing	Stopwatch & screen recordings	Average Time: 17 min 42 sec (96% success rate)
Interview	Semi-structured (10–15 mins)	Key themes: ease of use, fast workflow, intuitive design

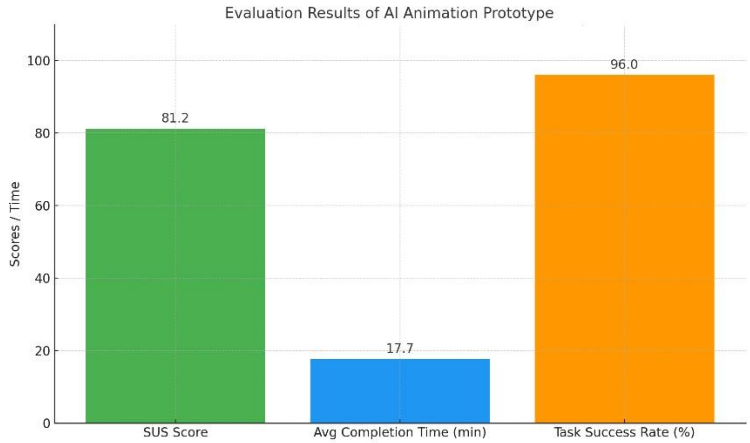

Figure 3. Bar chart illustrating user evaluation metrics for the AI-based animation prototype system.

6. RESULTS AND DISCUSSION

The evaluation results demonstrate that the proposed AI-based animation production system significantly improves production efficiency compared to traditional methods. The average time required to produce a 30-second animation decreased from approximately 200–300 hours [15] to just 25-40 hours, depending on the complexity of the content. This reduction supports findings by [19], who noted that AI integration can enhance animation production efficiency by up to 35%. The drastic decrease in time consumption indicates a transformative shift in production workflows, particularly for small teams or individual creators with limited resources.

In terms of user perception, the majority of participants rated the visual quality and naturalness of motion produced by the system as satisfactory to good. This aligns with [18], who found a generally positive reception among non-expert users toward AI-assisted animation, provided that the system offers sufficient creative control. Participants especially appreciated the system's ability to generate synchronized lip movements and appropriate gestures automatically. Several participants

highlighted the benefit of being able to preview and revise content without complex software or extensive manual editing.

However, some technical challenges were still encountered. One recurring issue was the need for fine-tuning textual prompts to produce outputs that align closely with user expectations. Users with less experience in prompt engineering sometimes struggled to articulate detailed instructions, resulting in generic or off-target animations. Additionally, while the interface was designed to be intuitive, a slight learning curve was observed, particularly for those unfamiliar with creative AI tools. Some users requested more guided templates or automated suggestions to support the creation of compelling narrative structures. Another point of concern was the occurrence of occasional visual artifacts or output biases, which could be traced back to the limitations of the training datasets used in the generative models. For instance, animations depicting specific cultural or historical themes were not always rendered with sufficient accuracy or sensitivity. This suggests a need for more diverse and representative training data in future iterations.

The study reinforces the real potential for democratization in animation production. By leveraging cutting-edge generative technologies, the system reduces the technical and financial barriers that have historically limited access to animation creation. These findings are consistent with [9], who emphasized the democratizing potential of AI in visual arts, particularly in enabling access for creators without technical backgrounds [9]. Despite these advances, the democratization of animation through AI also introduces new challenges. One key issue is the growing importance of prompting literacy, the ability to communicate creative intent effectively through natural language. Similar to trends in AI-generated art and design [20], the quality of results is increasingly dependent on the user's ability to craft detailed and precise prompts. This raises the question of whether future education and training should include prompt engineering as a digital literacy skill.

Ethical concerns also emerge in this context. The ownership of AI-generated content remains a legal grey area in many jurisdictions, particularly when the output is derived from models trained on copyrighted material. Additionally, the presence of inherent biases in AI training datasets may result in the reinforcement of stereotypes or misrepresentation of minority groups. These concerns must be addressed through transparent model development, inclusive datasets, and robust ethical guidelines. The results show that AI-driven animation systems have the potential to revolutionize content production by making it faster, more accessible, and scalable. Nonetheless, realizing this potential requires thoughtful design, continuous iteration, and responsible deployment of technology. Future work should focus on refining the interface for greater ease of use, improving output fidelity across diverse contexts, and establishing ethical frameworks to guide the use of AI in creative fields.

7. CONCLUSION AND FUTURE RECOMMENDATIONS

This study successfully developed and evaluated a prototype of an end-to-end AI-based animation production system that is accessible to non-expert users. The findings demonstrate that the proposed model significantly enhances production efficiency while maintaining acceptable output quality, without requiring deep technical expertise from users. The system proved to be especially valuable for independent creators, educators, and digital content producers, offering them a practical tool to translate ideas into animated content with minimal barriers.

As a direction for future development, it is recommended that the system be equipped with enhanced interactivity features such as real-time feedback loops, allowing users to iteratively refine their animations based on immediate system responses. Additionally, implementing more flexible visual style customization mechanisms would enable users to better express unique artistic identities and tailor animations to specific narrative contexts. Incorporating a broader library of character archetypes, settings, and motion presets could further expand the creative potential for users across diverse genres.

Long-term research is also required to evaluate the system's sustainability and scalability across multiple application domains, including education, digital storytelling, and indie game development. For instance, longitudinal studies could assess how regularly teachers or students integrate the system into classroom learning and whether it fosters greater engagement or creativity over time. Moreover, the integration of optimized local hardware support, such as GPU acceleration or edge computing, remains a critical agenda item to improve processing speed, reduce latency, and enhance the portability of the production system in the future.

Finally, future work should also explore multilingual support and accessibility features to accommodate diverse user populations globally, ensuring that the democratization of animation through AI includes users from varying linguistic and cultural backgrounds. With continuous refinement and ethical consideration, this system has the potential to become a key enabler in reshaping how animation is produced, accessed, and experienced in the digital age.

ACKNOWLEDGMENT

This article is one of the outputs from an internal research project under the Skema Khu-sus 2024 scheme funded by the Institute for Research and Community Service (LPPM) at Universitas Dian Nuswantoro, Indonesia.

REFERENCES

[1] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. arXiv:2112.10752v2 [cs.CV]. https://doi.org/10.48550/arXiv.2112.10752.

[2] Zhang, Z., Wang, Y. Mao, W., Li, D., Zhao, R., Wu, B., Song, Z., Zhuang, B., Reid, I., & Hartley, R. (2025). Motion anything: Any to motion generation. arXiv:2503.06955v2 [cs.CV]. https://doi.org/10.48550/arXiv.2503.06955

[3] Liu, V., & Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, Article 384, 1–23. https://doi.org/10.1145/3491102.3501825.

[4] Westcott, K., Arkenberg, C., Arbanas, J., Auxier, B., Loucks, J., & Downs, K. (2023). 2023 Digital media trends: Immersed and connected younger generations are weaving TV, gaming, and UGC into a tapestry of entertainment, community, and meaning. Deloitte Center for Technology, Media & Telecommunications. https://www.deloitte.com/us/en/insights/industry/technology/media-industry-trends-2023.html.

[5] Abootorabi, M. M., Ghahroodi, O., Zahraei, P. S., et al. (2025). Generative AI for character animation: A comprehensive survey of techniques, applications, and future directions. arXiv:2504.19056v1 [cs.CV]. https://doi.org/10.48550/arXiv.2504.19056.

[6] Chen, Y., Wang, Y., Yu, T., & Pan, Y. (2024). The effect of AI on animation production efficiency: An empirical investigation through the network data envelopment analysis. Electronics, 13(24), 5001. https://doi.org/10.3390/electronics13245001.

[7] Esteves, C., Arechavaleta, G., Pettré, J., & Laumond, J.-P. (2008). Animation planning for virtual characters cooperation. In ACM SIGGRAPH 2008 classes (SIGGRAPH '08) (Article 53, pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/1401132.1401204.

[8] Holden, D., Komura, T., & Saito, J. (2017). Phase-functioned neural networks for character control. ACM Transactions on Graphics (TOG), 36(4), 1–13. https://doi.org/10.1145/3072959.3073663.

[9] Herrie, M. B., Maleve, N. R., Philipsen, L., & Staunæs, A. B. (2025). Democratization and generative AI image creation: aesthetics, citizenship, and practices. AI & Soc 40, 3495–3507 (2025). https://doi.org/10.1007/s00146-024-02102-y.

[10] Adobe. (2024). Creative Cloud pricing and membership plans. Retrieved from https://www.adobe.com/creativecloud/plans.html

[11] Autodesk. (2024). Maya pricing. Retrieved from https://www.autodesk.com/products/maya/pricing.

[12] Danyluk, A. (2020). Learning Blender: A hands-on guide to creating 3D animated characters (2nd ed.). Addison-Wesley Professional.

[13] Massanari, A. (2015). Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. New Media & Society, 19(3), 329-346. https://doi.org/10.1177/1461444815608807.

[14] Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

[15] Laybourne, K. (2020). The animation book: A complete guide to animated filmmaking, from flip-books to sound cartoons to 3D animation (3rd ed.). Crown Publishing Group.

[16] Anderson, T., & Niu, S. (2025). Making AI-Enhanced Videos: Analyzing Generative AI Use Cases in YouTube Content Creation. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25). Association for Computing Machinery, Article 388, 1–7. https://doi.org/10.1145/3706599.3719991.

[17] Ouyang, L., Wu, J., Jiang, X., et al. (2023). Training language models to follow instructions with human feedback. arXiv:2203.02155v1 [cs.CL]. https://doi.org/10.48550/arXiv.2203.02155.

[18] Hu, D., Choi, M., Giri, N., Mousas, C., Adamo-Villani, N. (2025). Perceptions of AI in Animation Production. In: Machado, P., Johnson, C., Santos, I. (eds) Artificial Intelligence in Music, Sound, Art and Design. EvoMUSART 2025. Lecture Notes in Computer Science, vol 15611. Springer, Cham. https://doi.org/10.1007/978-3-031-90167-6_6.

[19] Chen, Y., Wang, Y., Yu, T., & Pan, Y. (2024). The Effect of AI on Animation Production Efficiency: An Empirical Investigation Through the Network Data Envelopment Analysis. Electronics, 13(24), 5001. https://doi.org/10.3390/electronics13245001

[20] Maerten, A. -S., & Derya Soydaner, D. (2023). From paintbrush to pixel: A review of deep neural networks in AI-generated art. arXiv:2302.10913v2 [cs.LG]. https://doi.org/10.48550/arXiv.2302.10913.