

AI ART

From generative images
to cinematic worlds

The history.
The technology.
The image makers.
The video revolution.
The connected studio.

**A GLOBAL
CREATIVE HISTORY
1950s-2026**

Researched through
11 October 2026



Inside this article

One connected history of AI images, moving scenes and creative production.

Follow the chapters in order, or use the linked contents to explore a model family or a technical idea. Numbered source links lead to the reference library at the end.

The new creative medium 3

Before the boom: the long history of machine-made images 4

How an AI image is made 6

Three years of acceleration: 2023 to October 2026 9

The image makers: different routes to the same canvas 11

How AI video is generated 17

Dreamina and ByteDance: from a prompt to a directed sequence 19

Kling: native 4K, audiovisual control and the 4.0 transition 22

The wider video landscape 24

Creative agents and the connected studio 27

A practical production journey 30

What has improved, and what still needs judgment 32

A compact glossary 34

Sources and research notes 35

01 / BUNCOSMOS AI ART

The new creative medium

A few years ago, asking a computer for a picture often meant accepting whatever the computer happened to produce. A beautiful face might arrive with the wrong number of fingers. A sign might resemble writing without containing readable words. A short animation might turn a character's sleeve into a second arm halfway through the shot. The result could be astonishing, but directing it was another matter.

By October 2026, the ambition has changed. Creators can bring reference photographs, sketches, character sheets, previous clips and sound into the same process. They can ask for an edit rather than a replacement, define the beginning and end of a shot, generate dialogue alongside motion, or ask an agent to organize a sequence of creative tasks. None of these capabilities is perfect. Together, however, they explain why AI art now belongs to a much larger story than typing a sentence and receiving a picture. [1, 2, 3]

Imagine a BunCosmos project: Bun is painting in her studio, Mini Sol enters with a glowing idea, and the scene becomes a short animated story. Making one attractive frame is only the first requirement. Bun must remain recognizably Bun. Her clothes should not change without a reason. The easel must occupy a consistent place. Sol's movement must match the camera. A line of dialogue needs the right voice, expression and timing. The finished film needs cuts, sound, typography and an export suitable for its destination. Each requirement exposes a different part of the technology.

This article follows that whole creative chain. It begins before the current boom, explains the machinery in everyday English, traces the rapid developments of 2023 through 2026, and examines the American, Asian and European developers that shaped them. Dreamina and Kling receive close attention because they make the transition from isolated generation to controllable video especially visible.

The research cutoff is **11 October 2026**. A launch date is not necessarily the date everyone gained access. A research demonstration is not a consumer product. A studio that offers many models is not automatically the developer of those models. Throughout this article, those distinctions matter more than a dramatic promotional claim. Specifications are dated snapshots; access can differ by region, subscription and API.

02 / BUNCOSMOS AI ART

Before the boom: the long history of machine-made images

Computers were making art before they learned from photographs

Computer art did not begin with Midjourney or ChatGPT. Early digital and electronic artists explored how rules, signals and machines could become artistic materials. The V&A's collection includes Ben Laposky's *Oscillon 40* from 1952, made by photographing electronic waveforms. Later computer artists worked with programmed geometry, repetition, randomness and plotters. These works belong to the ancestry of generative art, although calling every early electronic artwork modern machine-learning AI would blur an important distinction. [4]

Harold Cohen's AARON is a more direct landmark in AI art. Cohen began exploring computers in the late 1960s and developed AARON through his work in the 1970s, including a period at Stanford's AI laboratory in 1973. He encoded knowledge about drawing into a system that could produce different compositions. Its early abstract drawings were sometimes hand-coloured by Cohen. AARON did not learn a visual style by absorbing millions of online pictures; Cohen developed and refined its rules. [5, 6]

That difference explains two distinct creative traditions. In rule-based generative art, the artist may explicitly design the procedure that produces the work. In contemporary learned generation, the developer designs a learning system, and the system acquires statistical representations from data. The artist then works through prompts, references, edits, selection and other controls. Both traditions can involve human intention, but the mechanism and the relationship to training material differ.

The 2010s: learning becomes a creative engine

In 2014, Ian Goodfellow and colleagues introduced generative adversarial networks, or GANs. A generator learned to produce samples while a discriminator learned to distinguish generated samples from training examples. Their competing objectives encouraged the generator to become more convincing. GANs helped establish that a neural network could learn a distribution of visual appearances rather than simply follow a list of drawing instructions. [7]

The distinction between learning an appearance and retrieving a picture is central. A trained generator can synthesize a new sample from its learned representation. That does not mean every output is unique, or that memorization never occurs. It means the basic

operation is generation from learned parameters rather than a search engine returning the nearest photograph.

Google's 2015 Inceptionism experiments, associated with DeepDream, offered another route into neural imagery. Researchers amplified features an image-recognition network detected, producing strange recurring eyes, animal forms and architectural patterns. The results revealed something about what the network had learned to recognize. They also gave artists an accessible glimpse of a machine's visual associations. [8]

Neural style transfer, introduced by Leon Gatys, Alexander Ecker and Matthias Bethge in 2015, separated representations of an image's content from representations of its style. It could preserve the recognizable structure of one image while drawing on the visual texture of another. This was not yet the broad, conversational creation system people use today, but it made a powerful idea tangible: learned representations could become tools for recombining visual qualities. [9]

These developments were not steps in a single inevitable product roadmap. GANs, recognition-based visualizations and style-transfer systems solved different problems. Their combined significance was cultural as well as technical: machine learning stopped being associated only with identifying what was in a picture and became associated with making and transforming pictures.

Consumer tools helped carry those ideas beyond laboratories. Artbreeder, originally Ganbreeder, let people explore and combine visual traits using GAN-based systems such as BigGAN and StyleGAN. NightCafe, founded in 2019, developed an art-making community around accessible generation. WOMBO Dream brought text-driven artwork to a large mobile audience in 2021. Their significance was accessibility: a person did not need to train a network or write a research implementation to experiment with machine-made imagery. These applications helped establish the social and creative habits that the later boom expanded. [10, 11, 12]

2020 to 2022: language enters the canvas

Diffusion research provided a particularly influential foundation. The 2020 paper *Denoising Diffusion Probabilistic Models*, by Jonathan Ho, Ajay Jain and Pieter Abbeel, showed strong image synthesis by learning to reverse a gradual corruption process. Instead of relying on an adversarial contest, a model could learn how noisy visual data should become less noisy. This approach would become central to a large part of the subsequent image-generation ecosystem. [13]

OpenAI's CLIP, introduced in 2021, linked visual representations with natural-language descriptions. It was trained to associate images with text, making language a useful bridge into visual concepts. CLIP was not itself a painting application, and later generators do not all use CLIP. Its importance was the wider demonstration that text and images could be organized within compatible learned representations. [14]

The first DALL-E, announced in January 2021, used a transformer-based approach to generate images from descriptions. DALL-E 2 followed in 2022 with a different system and much stronger visual quality. These releases helped make the phrase *text-to-image* familiar outside specialist research communities. The public experience was no longer only "apply an effect to this photograph"; it was "create a scene I can describe." [15, 16, 17]

Latent diffusion made that possibility more practical. Robin Rombach and colleagues' research moved much of the generative work into compressed representations rather than full-resolution pixels, reducing computational demands while retaining useful visual detail. Stable Diffusion's public release on 22 August 2022 brought this research lineage into a widely adaptable model ecosystem, developed through collaboration involving Stability AI, academic researchers and Runway. It was not the work of one company in isolation. [18, 19, 20]

By the end of 2022, a new creative culture was taking shape around commercial services, downloadable weights, communities, custom models and prompt experimentation. Midjourney became one of the defining names of that culture. The stage was set for the next three years: stronger images, more control, moving scenes, sound and finally creative orchestration.

03 / BUNCOSMOS AI ART

How an AI image is made

Training and generation are different activities

Training is the process in which a model's numerical parameters are adjusted using examples and an objective. Generation, often called inference, uses an already-trained model to produce an output for a new request. The distinction is easy to miss because a conversational application can look as though it is learning the character permanently every time an image is uploaded.

An uploaded reference can influence the current generation without changing the underlying model weights. A service may retain it as a project asset, but retaining a file is different from training on it. Fine-tuning is another process again: selected weights or adaptation parameters are deliberately trained to specialize the system. DreamBooth demonstrated subject-specific personalization using a small reference set, while LoRA introduced a parameter-efficient adaptation method that later became widely useful in generative workflows. [21, 22]

These distinctions help explain a common disappointment. If a creator uploads Bun once and later opens an unrelated project, the service may not automatically know which Bun to reproduce. Consistency may depend on bringing the reference back, using a saved

character feature, loading a trained adaptation, or continuing within a supported project context. "The AI has seen her" is not a complete description of the technical setup.

From words to numerical conditioning

A prompt is converted into numerical representations that the generator can use. The exact mechanism varies by architecture, but the purpose is to condition the output on the requested scene. A phrase such as "a red hoodie embroidered with gold stars" supplies information about an object, colour, material and decoration. A longer description may add spatial relationships, lighting, camera position and the intended mood.

The model must do more than recognize a collection of nouns. "Bun holds a brush beside an easel" is different from "a brush rests on the easel beside Bun." Producing the right relationship requires the visual generator to associate language with structure. Stronger language conditioning, better captions and architectures that let visual and text representations interact have helped improve this ability. Stable Diffusion 3's rectified-flow transformer research explicitly explored separate text and image processing with bidirectional information exchange. [23]

The same problem explains why a prompt can contain the correct words but still produce the wrong image. A system may draw all the named objects yet place them incorrectly, assign a colour to the wrong object, or omit a less prominent instruction. Better prompt adherence means that these relationships are followed more reliably; it does not mean the model has become an infallible interpreter.

Diffusion: a learned route out of noise

A useful simplified picture of diffusion is a sculptor working from visual static. During training, the system sees corrupted examples and learns how to move toward a cleaner representation. At generation time, it starts from noise and repeatedly updates the sample, guided by the prompt or other conditions. The process is mathematical, not an artist consciously imagining each brushstroke. [13]

In latent diffusion, those updates occur in an encoded, compressed space. An encoder maps an image into a smaller numerical representation; a decoder can map the representation back into pixels. The generative model works on the compressed version, and the decoder produces the visible image. The latent space is a representation system, not a hidden folder of complete photographs. [18]

Diffusion Transformers, or DiTs, replace the commonly used U-Net backbone with a transformer operating on latent patches. This changes the network that performs the generative computation; it does not eliminate the need for data, conditioning or a sampling procedure. The architecture helped connect visual generation with the scaling methods already important in transformer research. [24]

Flow-based formulations offer a related way to learn a path between noise and data. Rectified-flow methods aim for more direct transformations. In practical products, the

architecture, training objective, sampling schedule and acceleration techniques interact. Calling every modern image system simply "a diffusion model" conceals meaningful differences, even when the user experience looks similar. [23, 25]

Some models build images differently

Autoregressive generation predicts a sequence, with later elements conditioned on earlier ones. The first DALL-E used a transformer-based sequence approach. OpenAI's 2025 system card described native GPT-4o image generation as autoregressive rather than the diffusion approach used by later DALL-E systems. That historical statement should not be stretched into an undocumented description of every subsequent GPT Image release. [15, 26]

Hybrid systems also exist. Z.ai's GLM-Image combines an autoregressive generator with a diffusion decoder, separating parts of semantic construction from detailed image rendering. Its published architecture illustrates why the field cannot be explained by one universal recipe. A model can use more than one generative mechanism within the same system. [27]

For a creator, the most useful question is usually not which mathematical family has the best slogan. It is whether the model follows the brief, preserves the needed details, produces an appropriate visual style, edits reliably and fits the workflow. Architecture helps explain progress, but it does not replace output evaluation.

Seeds, variation and the limits of control

Generative systems commonly use randomness to explore different plausible outputs. A seed can make a sampling run more reproducible within a particular setup. It is not a universal identity code for a character. Changing the model, resolution, implementation or other settings can change the result even when a familiar seed is reused.

Reference images provide another kind of control. They can guide a character's appearance, the composition, a material, a visual style or a particular object. These are different goals. A style reference may transfer the atmosphere of Bun's studio without preserving Bun's face. A subject reference may preserve her identity while changing the lighting and pose. The creator needs to know which role a tool assigns to each input.

ControlNet made structural conditioning especially visible in 2023. Its controls included edges, depth, segmentation and human pose. The model could be guided by a description of where things belonged rather than text alone. This matters whenever the layout or posture is part of the brief. Structure-guided generation is still synthesis, however; it does not guarantee that every source pixel survives. [28]

Inpainting changes an area within an image; outpainting expands the image beyond its existing boundary. Conversational editing can add an instruction such as "keep the character and replace only the background." Midjourney's current Edit Model combines instruction-based changes, reference-led generation and regional editing, while OpenAI's

Image API and Responses API support generation and editing within broader application workflows. [29, 30]

For an exact requirement such as "place Bun and her desk scene without changing them," conventional compositing remains valuable. The original image can be placed as an intact asset while other design elements are built around it. Generating a new version that looks similar is a different task. Greater generative control expands the choices; it does not make preservation and reconstruction identical.

04 / BUNCOSMOS AI ART

Three years of acceleration: 2023 to October 2026

2023: the image becomes a mainstream creative tool

By 2023, the public had moved beyond the question of whether text-to-image generation could work. The questions became how good it could look, how accurately it could follow a brief, and whether it could fit into professional production. SDXL brought a larger, more capable open image system into the Stable Diffusion family. ControlNet gave creators structural control. Adobe introduced Firefly on 21 March, initially focused on images and text effects, connecting generative AI with the established creative-software world. [31, 28, 32]

DALL-E 3 made the conversational relationship between language and image creation more prominent. ChatGPT could help develop a brief and formulate a prompt for the image model, reducing the burden of learning a specialist prompting dialect. This was conversational orchestration of a distinct generator, a useful historical distinction from the native GPT-4o image capability introduced later. [33]

Midjourney's V6, released in December 2023, marked another stage in the pursuit of richer imagery and stronger prompt interpretation. The competitive pressure was no longer only about striking style. Creators increasingly wanted the system to respect a specific description and to produce coherent details. [34]

Video remained much less dependable than still images. A short moving scene had to maintain appearance over time, something an individual attractive frame did not prove. That gap between visual beauty and sustained control would define much of the following year's competition.

2024: moving scenes become credible

OpenAI's February 2024 Sora research preview attracted attention with longer, more coherent examples. Its technical report described a diffusion transformer working with spacetime patches and emphasized that video models still had significant weaknesses as simulators, including physical interactions and cause-and-effect behaviour. The demonstration was influential, but a preview should not be described as unrestricted public availability. [35]

Runway introduced Gen-3 Alpha in June 2024, emphasizing improvements in fidelity, consistency, motion and temporally detailed captioning. Kuaishou began Kling testing in June and expanded its beta internationally in July. The emergence of Kling made the global nature of the competition unmistakable: advanced consumer-accessible video generation was not confined to American laboratories. [36, 37, 38]

Open video research also gained momentum. CogVideoX released models in August 2024 and added more capabilities later that year. Tencent's HunyuanVideo became part of the expanding open video ecosystem. Developers could increasingly examine code, adapt workflows and run models outside a single hosted application. [39, 40]

Meta's Movie Gen research explored video generation, editing, personalization and sound, while Amazon introduced Nova Canvas and Nova Reel in December. These projects reflected different distribution strategies: research programs, creator-facing tools and enterprise cloud services all contributed to progress. They should not be treated as interchangeable subscription apps. [41, 42]

2025: consistency, editing and sound move to the centre

The defining shift of 2025 was toward using generation inside a sustained creative process. Native GPT-4o image generation arrived in March, emphasizing text rendering, contextual understanding and editing. OpenAI released the corresponding GPT Image API capability in April. Google's Nano Banana, introduced in August, emphasized likeness preservation and natural-language changes to supplied pictures. [43, 44, 45]

Google introduced Veo 3 and Imagen 4 in May, together with Flow. Veo 3 made generated dialogue, ambient sound and effects part of the video story. Midjourney introduced its first video model in June, allowing users to animate images. Seedance 1.0, documented publicly in June, emphasized native multi-shot video and stable motion. The expanding feature set made the distinction between image generation and filmmaking harder to ignore. [46, 47, 48]

Image editing continued to mature through systems such as FLUX.1 Kontext and later FLUX.2. Nano Banana Pro followed in November. OpenAI's GPT Image 1.5 release in December emphasized more precise changes and stronger continuity across edits. In video, Kling's Omni research represented the push to unify creation, editing and multimodal instructions rather than keep each task isolated. [49, 50, 51, 52, 53]

2026: the workflow becomes part of the product

By October 2026, the field includes conversation-led video editing, longer audiovisual clips, model-specific character controls, multi-reference inputs and agents that coordinate creative tools. These changes are not just improvements in image sharpness. They expand how a person can direct a project.

The current image landscape includes ChatGPT Images 2.5, Nano Banana 2.1 on Google's advertised consumer surface, Midjourney V8.2, Seedream 5.0 Pro and Ideogram 4.5. The chronology of their earlier releases is discussed below. In video, Seedance 2.5, MiniMax H3, Wan 3.0, FLUX 3 Video and Google Omni illustrate the move toward richer multimodal control. Kling 4.0 is in a staged early-access rollout; its announcement is not evidence that every user already has the complete model. [1, 54, 34, 55, 56, 57, 58, 2, 59, 60, 61]

There is also a caution against treating technical progress as an uninterrupted list of available services. OpenAI's official notice records the discontinuation of the Sora web and app experiences on 26 April 2026 and the API discontinuation date as 24 September 2026. Sora remains an important chapter in AI video history, but it is not presented here as a currently available OpenAI video service. [62]

The most reliable account of this period therefore combines chronology with practical distinctions: what was announced, what was demonstrated, what actually shipped, and what a creator can control in the current implementation.

05 / BUNCOSMOS AI ART

The image makers: different routes to the same canvas

OpenAI: from DALL-E to the current ChatGPT Images

OpenAI's image history is a sequence of changes in both the generator and the way people communicate with it. DALL-E established an early transformer-based text-to-image demonstration in 2021. DALL-E 2 broadened the visual quality and editing possibilities in 2022. DALL-E 3 made detailed instruction following and ChatGPT-assisted prompting central to the experience in 2023. The common thread was making language a more effective creative interface. [15, 16, 33]

The March 2025 release of native GPT-4o image generation changed that relationship again. OpenAI emphasized the use of conversation context, supplied images, world knowledge and accurate text rendering. The practical appeal was not only a nicer illustration: a creator could develop the idea and revise the resulting visual within the

same conversation. The system card explicitly distinguished its autoregressive architecture from DALL-E's diffusion approach. [43, 26]

The API release on 23 April 2025 made GPT Image available inside other products. This is an important part of the industry's history: the same underlying model can appear in many interfaces. A design app that offers GPT Image is not thereby the inventor of GPT Image. The host app may add valuable layout, asset management and export features while drawing on OpenAI's generative engine. [44]

GPT Image 1.5 arrived in December 2025 with stronger editing and instruction following. ChatGPT Images 2.0 was announced on 21 April 2026. The current release at this article's cutoff is **ChatGPT Images 2.5**, introduced on **8 September 2026**. OpenAI describes improvements in subject fidelity, detail, lighting, targeted editing and continuity over multiple revisions, together with sketch references and comments on images. [52, 63, 1]

The current API distinguishes **GPT Image 2.5 Flare**, oriented toward everyday quality and speed, from **GPT Image 2.5 Sunburst**, oriented toward more detailed precision with longer generation time. These are image model names, not simply alternate names for the general conversational model a user selects. The Image API supports generation and editing; the Responses API can integrate image generation into an ongoing tool-enabled interaction. [1, 30]

The creative significance of this progression is continuity. A conversation can carry the brief, reference material and successive decisions, reducing the need to begin from an empty prompt box at every stage. The limitation is equally important: a claim of improved preservation is not a guarantee that every unedited pixel will remain identical. An exact brand asset still deserves close comparison against the original.

Midjourney: aesthetic exploration becomes a broader workflow

Midjourney helped define the public image-generation boom because its outputs made visual imagination feel immediately accessible. Its later development shows the movement from a striking aesthetic toward more dependable interpretation and editing. The official version history records V6 in December 2023, V6.1 in July 2024 and V7 in April 2025. V7 introduced Draft Mode and Omni Reference. [34]

In 2026, V8.0 appeared as an early-access alpha in March, V8.1 followed in April, and **V8.2 became the default on 24 July**. V8.1 added native 2K HD generation. V8.2 emphasized aesthetics and personalization, with a new Edit Model replacing several earlier reference and retexture features. Niji, developed with Spellbrush, remains a distinct branch focused on anime and related illustration aesthetics. [34]

Midjourney's current Edit Model can modify an existing image through written instructions, create with up to four reference images, and support inpainting and outpainting. That is a meaningful change from treating each prompt as a fresh independent composition. Creators can bring a visual direction forward instead of repeatedly describing it from scratch. [29]

The first Midjourney video model arrived on 18 June 2025. It animated a still image into short clips, with further extension options described in the product documentation. The initial release produced five-second videos, illustrating a natural path from an image-first community into motion. It did not mean that every still-image control immediately became an equivalent video control. [47, 64]

Personalization and identity should be distinguished. A personalization profile helps the system reflect a user's aesthetic preferences. A reference-led edit helps specify a subject or scene. A preferred style does not automatically preserve Bun's exact face, and a subject reference does not automatically reproduce the desired palette. Understanding that separation leads to more deliberate creative choices.

Google: Imagen, Nano Banana and visual reasoning

Google's generative-media work spans research models, consumer products and developer services. Imagen was an important text-to-image research family before the Nano Banana brand existed. By May 2025, Imagen 4 was presented alongside Veo 3 and Flow, emphasizing improved imagery and typography within a wider media ecosystem. The distinction between Imagen and Gemini Image matters: they are related parts of Google's offerings, not names for one unchanging model. [46, 65]

Nano Banana entered the Gemini app on 26 August 2025 as the public-facing name associated with Gemini 2.5 Flash Image. Its appeal centred on editing: maintain a recognizable person or pet while changing clothing, surroundings or style, or combine elements from different pictures. Instead of describing an imagined subject in detail, the creator could supply the actual visual starting point. [45, 66]

Nano Banana Pro, introduced in November 2025, extended that approach with the Gemini 3 Pro Image model, emphasizing visual reasoning, more controllable designs and stronger text rendering. **Nano Banana 2**, announced on 26 February 2026, was identified as Gemini 3.1 Flash Image, combining faster operation with capabilities previously associated with the Pro experience. Google's release also discussed provenance through SynthID and C2PA. [51, 67, 68]

Google added **Nano Banana 2 Lite** for developers in June 2026, emphasizing throughput and cost-efficient iteration. At the October cutoff, Google's image-generation overview advertises **Nano Banana 2.1** as an advanced option available through a Google AI plan. The public page verifies the current product designation; it should not be used to invent an exact launch date or assume that every API has the same name and access conditions. [69, 54]

The broader change is toward an image model that can work with the logic of a request: how a poster should be arranged, which element an instruction refers to, and how a reference should influence the outcome. That remains different from guaranteed factual accuracy. A model can render a beautiful scientific infographic while introducing an incorrect label. Visual reasoning must be paired with editorial checking.

Adobe Firefly: generation inside established creative production

Adobe entered this period with an advantage that a stand-alone generator does not automatically have: existing applications for illustration, photography, layout and filmmaking. Firefly's March 2023 introduction focused on images and text effects. Subsequent developments brought generative tools into an ecosystem where creators already worked with selections, layers, timelines and export requirements. [32, 70]

The Firefly Video Model entered public beta in February 2025. By October 2025, Adobe had introduced Firefly Image Model 5 in public beta and described Firefly as a creative studio combining its own models with partner models. That distinction should remain visible. The Adobe-developed engine and an external model available inside the Firefly app are not the same thing. [71, 72]

Adobe states that its Firefly models are trained on licensed content, including Adobe Stock, and public-domain content whose copyright has expired. It also offers custom-model approaches for brand assets. These are relevant differences in training and commercial positioning, but "commercially safe" is a provider's product claim, not a universal legal guarantee covering every output, jurisdiction or third-party model. [73, 74]

For many professional users, the decisive question is how a generated result becomes an editable part of a larger piece. A local change to a photograph, an extension of a shot, a vector element or a layout-ready asset may be more valuable than another attractive but inflexible flat image. Firefly helps illustrate why integration can be as important as the generator's headline score.

Stable Diffusion and the importance of adaptable models

Stable Diffusion's 2022 public release helped establish an ecosystem in which people could run, customize and build around visual models. SDXL in 2023 expanded the family, and Stable Diffusion 3 and 3.5 in 2024 reflected changes in architecture and capability. Stability AI's published work on rectified-flow transformers helped explain the technical direction rather than leaving the mechanism entirely behind a hosted interface. [19, 31, 75, 23]

Adaptability created a different kind of creative practice. Instead of choosing only among a vendor's buttons, a user could combine a model with structural controls, adapters, subject training, restoration or scaling tools. This made the workflow more demanding, but also more configurable. It supported experimentation by researchers, independent artists and developers who wanted to build products of their own.

"Open" needs careful reading. Available weights allow a user to inspect or run a model, but the training data, full training procedure and all rights may not be open. Licenses can contain use restrictions or commercial conditions. Those conditions vary by release and variant. The original Stable Diffusion license is part of the historical record; it should not be assumed to govern every later model. [19]

The significance of this branch is therefore not that everyone should install a complex local workflow. It is that the field has been shaped by both accessible hosted tools and inspectable, adaptable model releases. Progress did not come only from a small number of closed consumer apps.

Black Forest Labs: FLUX expands from images into multimodal media

Black Forest Labs emerged in 2024 from a research lineage connected to influential visual-generation work. The FLUX family became important for image generation and later contextual editing. FLUX.1 Kontext, introduced in May 2025, combined text and image inputs for generation and changes within context. FLUX.2 followed in November, emphasizing multiple references, consistency, typography and editing at higher resolutions. [49, 50]

In July 2026, BFL announced **FLUX 3**, a unified model trained across images, video and audio. Its design explicitly aimed to learn connections among appearance, motion and sound. This was a move beyond an image-only family, although the first announcement described early access and capabilities under continuing development. [76]

FLUX 3 Video became generally available through BFL's API and selected partners on 4 August 2026. The initial release supported clips up to 20 seconds with native audio, keyframe controls, continuation and multi-shot creation. BFL explicitly distinguished native HD generation from Full HD output obtained through upscaling. That is the kind of specification detail that a simple "1080p AI video" label can conceal. [59]

By October, BFL's current offerings also include **FLUX 3 Image**; its pricing page records an October launch promotion. The company's catalog continues to include FLUX.2 variants and tools. As with other providers, an evolving website can contain both new products and earlier documentation, so the current catalog and release notes need to be read together. [77, 78]

FLUX illustrates the international nature of this field. A history framed only as America versus Asia would miss important European research and model development. It also illustrates that a developer's identity can change over time: an image specialist may become a video and audio developer rather than remain permanently in the category where it first became famous.

Ideogram, Recraft and Leonardo: specialized creative strengths

Not every successful image model aims at the same balance. Ideogram has emphasized readable type and design. Its June 2026 **Ideogram 4.0** release provided open weights with a commercial license and described a 9.3-billion-parameter diffusion transformer using a vision-language text encoder and structured prompts. Its training approach connected layout information with descriptions so the system could learn where visual and textual elements belonged. [79, 80, 81]

At the research cutoff, **Ideogram 4.5** is also advertised, emphasizing targeted edits and reduced deterioration over successive revisions. This is a useful example of the industry's change in priorities: designers need an image to survive the fourth or fifth correction, not merely look attractive on the first attempt. Vendor demonstrations of superiority over competitors remain vendor evidence rather than a universal ranking. [56]

Recraft's **V4**, introduced on 17 February 2026, emphasizes design aesthetics and includes both raster and vector models. Its vector output consists of actual editable SVG files, not only a bitmap that resembles vector art. This distinction matters for logos, icons and assets that need to scale cleanly or be edited as shapes. V4's Standard, Pro and vector variants serve different production needs. [82]

Leonardo develops its own models as well as providing access to external ones. Its Phoenix family established a proprietary foundation-model direction, while **Lucid Origin**, launched in August 2025, emphasizes aesthetic range, prompt adherence and text rendering. Leonardo's published model documentation makes the distinction between its own systems and partner models visible. The existence of an all-in-one interface does not automatically make a company merely a reseller. [83, 84, 85]

These examples also broaden the geography: Ideogram is associated with Toronto, Recraft contributes another international design-focused branch, and Leonardo originated in Australia. A useful map follows the actual developers and their capabilities, rather than assuming all English-language creative tools are American.

Baidu, DeepSeek and Meitu: additional Asian foundations

Baidu's ERNIE-ViLG research predates the recent consumer boom. Its 2.0 system explored knowledge-enhanced text-to-image diffusion, using information about scene elements and different denoising experts across stages. The later ERNIE-Image repository describes an open-weight image generator built around a single-stream diffusion transformer. This is a research and model lineage, rather than simply an app adopting another provider's engine. [86, 87]

DeepSeek's Janus program explores unified multimodal understanding and generation, while separating the visual encoding used for those different tasks. Janus-Pro was released on 27 January 2025. Its contribution belongs to the architecture and open-model story; it should not be confused with a claim that a general chat interface automatically supplies a complete professional art studio. [88]

Meitu introduced its self-developed MiracleVision model in June 2023 alongside creative products for images, design and video. Later announcements described integration into its wider product range. Meitu illustrates another path into AI art: an established consumer imaging company develops a model and connects it with familiar creative applications. Together, these additional programs show why Asian image development extends well beyond the video brands that receive the most international attention. [89, 90]

06 / BUNCOSMOS AI ART

How AI video is generated

Time changes the problem

An AI video model has to produce a scene that develops. If Bun turns her head, the model must retain the relationship among her face, curls, beret and body. If she lifts a brush, the brush should remain in her hand. If the camera moves behind the easel, the system must infer appearances that were not visible in the original frame.

Generating each frame independently would usually produce unstable results because each image could be a different plausible interpretation. Video systems learn representations that connect visual information across time. Sora's 2024 technical report described compressed video representations and spacetime patches processed by a diffusion transformer. Other systems use their own temporal representations and architectures, so that report is an explanation of one important approach, not a blueprint for every product. [35]

Video training also requires meaningful descriptions of action. A caption that only says "a woman in a studio" leaves out the sequence. A temporally informative description might explain that she reaches for a brush, pauses, then makes a mark. Runway's Gen-3 Alpha announcement emphasized dense temporal captioning as part of its approach to controlling how scenes develop. [36]

The main creation modes

Text-to-video begins with a written scene. It gives the system substantial freedom over the look and action. **Image-to-video** begins with a supplied image and adds motion. The starting frame supplies appearance and composition, but unseen views still require inference. **Reference-to-video** uses images or other assets as guides to subjects, locations or style without necessarily treating the first reference as the opening frame.

First-and-last-frame generation defines two visual states and asks the system to construct the transition. **Multi-keyframe generation** adds intermediate states. **Video-to-video** transforms an existing clip, often preserving some motion or structure while changing a character, environment or visual treatment. Current providers expose different subsets of these modes, and similar names do not imply identical behaviour. [91, 92]

Motion transfer uses an existing performance as guidance. It can be more directed than asking a model to invent the performance from a sentence. A reference actor, camera path or rough 3D animation supplies timing and movement that the system can reinterpret

visually. The resulting clip still needs review for body structure, interactions and identity continuity.

Sound becomes part of generation

Early public AI video was often silent. Creators added music, effects and speech afterward. Native audiovisual generation learns to produce visual motion and sound together, so a door slam, footstep or line of dialogue can be aligned with what appears on screen. It is different from simply attaching a soundtrack to an unrelated video.

Seedance 2.0's technical report describes a unified multimodal audiovisual framework, while LTX-2's published report describes coupled video and audio streams with cross-attention. These architectures show how sound and vision can inform each other during generation. They do not imply that every service uses the same network or provides every sound-related feature through every endpoint. [93, 94]

Lip synchronization is only part of the task. A convincing spoken scene also needs the right speaker, vocal delivery, timing, facial performance and response from other characters. Multi-speaker scenes are especially demanding: an attractive face with synchronized mouth movement is not enough if the wrong character says the line.

Coherence is not the same as accurate physics

A video can look photographic while containing an impossible interaction. A cup may change shape as it is lifted. A character can walk convincingly yet fail to make proper contact with the floor. A reflection may ignore a moving object. These are temporal and causal errors, not merely problems of texture.

Longer clips make the issue harder because mistakes can accumulate. A model has more time to drift away from the original identity or lose track of an object's state. Keyframes and references can constrain the sequence, but they do not turn a statistical generator into a guaranteed physical simulator. OpenAI's original Sora report explicitly documented simulation limitations despite impressive visual demonstrations. [35]

For filmmaking, there is a useful practical response: design shots around what the system can reliably sustain, inspect the motion, and use ordinary editing and compositing where they offer better control. A successful finished film may combine many generated shots and substantial human work. Its total running time does not prove that one model generated the entire film in one uninterrupted pass.

Resolution, duration and HDR need precise language

A 4K file can be generated at a lower resolution and enlarged afterward. An upscaler may infer extra detail, but it cannot guarantee recovery of a texture that was never present. Similarly, frame interpolation can create intermediate frames and smoother playback without proving that the original generator produced all those frames natively.

FLUX 3 Video's launch documentation is unusually explicit: native HD generation, with Full HD output through upscaling. LTX's current support matrix differentiates models and resolution tiers. Kling already introduced native 4K generation through its 3.0 series in April 2026; the additional 4.0 announcement describes 10-bit HDR output at 4K and 1080p as coming soon. Existing native 4K and a forthcoming HDR upgrade are distinct milestones. [59, 95, 96, 61]

HDR concerns the representation of light and colour, not just the number of pixels. EXR output can preserve data for professional compositing and grading workflows. Luma's Ray3.2 announcement highlights native HDR and 16-bit EXR export because a production team may need to integrate the result with live-action footage. A bright-looking social-media clip is not, by appearance alone, evidence of a particular HDR pipeline. [97]

The useful specification is a complete one: generation mode, native or enhanced resolution, duration, frame rate, sound support, colour format, reference controls and access conditions. This is how a creator can compare tools without mistaking export labels for equivalent capabilities.

07 / BUNCOSMOS AI ART

Dreamina and ByteDance: from a prompt to a directed sequence

The app and the model are different parts of the system

Dreamina is a creative application in the CapCut/ByteDance ecosystem. **Seedream** is ByteDance Seed's image-model family; **Seedance** is its video-model family. Keeping the names separate helps a reader understand what is being improved. A change to Dreamina's workspace is an application change. A new Seedance release changes the underlying generation capabilities. Dreamina can also offer models from other developers. [98, 99]

This relationship is especially relevant for beginners because the app presents tasks in familiar creative language. A user can begin with a picture, describe an action, work with references and select a format without directly managing the model's numerical representation. The underlying model still determines much of what the system can synthesize; the surrounding app determines how easy it is to direct and organize that synthesis.

The image foundation: Seedream

Seedream matters to video even when the final deliverable moves. Strong character sheets, backgrounds and opening frames provide the visual material a video system can use. ByteDance's **Seedream 5.0 Pro**, introduced on 8 July 2026, emphasizes information visualization, precision editing, realistic imagery and multilingual generation. Its release describes spatial editing controls such as point or lasso selection, material replacement, layer separation and multi-image fusion. [55]

Those features reflect a broader production requirement. The creator needs to fix a specific prop or adapt an existing design, not merely ask for another loosely related picture. Multilingual typography also expands usefulness beyond English-language campaigns. A multilingual model can still misspell or misstate information, so the text remains an editorial responsibility.

Seedance 1.0: more than an animated still

ByteDance's June 2025 Seedance 1.0 technical announcement described 1080p generation from text and images, stable motion and native multi-shot storytelling. Its examples included changes among wide, medium and close-up shots within a ten-second clip. That is a different ambition from only adding a gentle camera drift to a still image. The generator begins to take responsibility for a sequence of views. [48]

The practical difference is narrative compression. In a short advertising clip, the system may establish a product, show a person using it and finish on a detail. The shots need to belong to the same event. Their arrangement matters as much as the individual frames, even when the sequence is only a few seconds long.

Seedance 2.0: references become a visual language

Seedance 2.0's published technical report describes native audiovisual generation and a unified framework for multimodal input. Reference images, video and audio can carry different kinds of direction. An image may specify a character; a clip may specify camera movement or performance; audio may specify a voice or rhythmic structure. Combining these signals is more expressive than trying to translate every detail into a long text prompt. [93]

This is also where a model's ability to interpret relationships becomes important. "Use the character from this image, the camera movement from this video, and the pacing of this audio" asks the system to assign a role to each reference. Simply mixing all the inputs indiscriminately would not satisfy the brief.

Seedance 2.5: longer creation and finer revision

On 31 July 2026, ByteDance introduced **Seedance 2.5**, built on its unified audiovisual direction. The release supports up to 30 seconds in one generation, longer sequences through extension, and expanded multimodal references. It lists as many as 30 images, 10 videos and 10 audio clips in a single request. Timestamped audiovisual editing lets a creator specify changes at particular moments instead of regenerating everything without direction. [100]

The product page also describes reference-led control and up to two extension operations. Longer generation is useful because dialogue, gestures and transitions have more room to develop. It also raises the difficulty: a thirty-second scene requires consistency across a much larger sequence of states than a five-second animation. Reference control and temporal editing help address that problem. [57]

Dreamina's current Seedance 2.5 page presents 30-second audiovisual creation, multimodal referencing and 4K output. The page's export and subscription features should not be treated as proof that every mode shares the same native resolution or that every account receives identical access. The more solid historical milestone is the documented thirty-second model capability, together with broader referencing and revision. [101]

What this changes for a BunCosmos-style project

Consider a short scene where Bun is painting and Sol brings her an idea. A reference image can establish Bun's appearance. A second asset can define Sol. A studio image can specify the location. A rough performance clip can guide an action, and an audio reference can influence the timing. The creator can then ask for a particular sequence of events rather than merely "make a cute animation."

The control becomes more specific when the request states what each asset is for. Bun's image defines her identity; the studio defines the surroundings; the motion reference defines the gesture. If Bun's original posture and desk scene must remain literally unchanged, a generative transformation is not automatically the right method. Preserve the original asset through compositing or use generation only where change is intended.

The real advance is a larger creative vocabulary. Text, still images, motion and sound can all become forms of direction. Dreamina demonstrates how an application can place that vocabulary within reach of ordinary creators. The creator still supplies the idea, chooses the material, judges the result and decides when the sequence is ready.

08 / BUNCOSMOS AI ART

Kling: native 4K, audiovisual control and the 4.0 transition

A model developer with a creator-facing ecosystem

Kling is developed by Kuaishou, which also developed the Kolos image family. Initial Kling testing began in June 2024, with international beta access expanding in July. Kuaishou's announcements documented a self-developed video model and a wider image-and-video model program. Kling should therefore be understood as a model family and creative product ecosystem, not only a front end for unrelated providers. [37, 38, 102]

Kling 2.0's April 2025 announcement emphasized multimodal visual language and improved creative control. Its subsequent Omni research described a framework connecting generation, editing and reasoning over visual-language inputs. That progression is important because creators rarely need only one isolated operation. A film project moves repeatedly among creating, correcting and reusing assets. [103, 53]

Kling 3.0: consistency, sound and multi-shot direction

Kuaishou introduced the Kling 3.0 family on 5 February 2026, including video and image variants and Omni versions. Its video announcement described stronger element consistency, image and video references, native speech across languages and accents, and clips up to 15 seconds. Multi-character dialogue and custom multi-shot storyboarding extended the creative task beyond attractive motion. [104]

For a creator, element consistency can concern more than a face. A costume, product, room or repeated prop may need to persist. These are all parts of continuity. A character who remains recognizable but changes clothing between shots may still fail the brief; a product whose logo drifts may be unusable for an advertisement.

Native 4K was already a separate milestone

Kling's official May 2026 article records native 4K generation through the Kling 3.0 series as launched on **23 April 2026**. It explicitly presents this as generation at 4K, rather than only enlargement after a lower-resolution render. This predates the full Kling 4.0 rollout and is an important correction to any account that treats "Kling 4K" and "Kling 4.0" as the same event. [96]

Native resolution matters because fine detail is part of the original synthesis. It can reduce dependence on an extra enhancement stage, although it does not automatically guarantee

identity stability or correct physics. High resolution gives the viewer more detail to inspect, including more detail in any mistakes.

Kling 4 Flash is live; the full 4.0 release is staged

Kling's 30 September 2026 announcement says **Kling 4.0 Flash** became available to **Ultra Yearly subscribers from 28 September**, while the broader Kling 4.0 rollout was planned for October. This is early access to a real product, not merely a research-only video demonstration. It is also distinct from universal availability of the complete forthcoming model across all plans and APIs. [61]

The announced 4.0 upgrade includes native generation up to 30 seconds, stronger dynamic motion, improved stereo audio and lip synchronization, and up to 15 multimodal references. The 10-bit HDR capability at 4K and 1080p is described in the launch article as forthcoming. These specifications describe the full-release direction; available Flash controls may differ. [61]

The accompanying keyframe guide describes up to 10 keyframe images for defining the sequence's visual progression. That gives a creator several planned states within one generation rather than only a starting picture. [92]

References define identity; keyframes define development

Kling's keyframe guide makes a useful distinction: reference assets help define what should remain consistent, while keyframes define important visual states the sequence should reach. A thirty-second scene can have a beginning, a reveal, a reaction and a final composition. The model constructs motion and transitions between those anchors. [92]

For example, a product film might begin with Bun holding a wrapped object, show the reveal, move to a close-up, and finish with Sol reacting. A reference sheet can establish the characters. Keyframes can specify the stages of the action. A prompt can explain the relationship among those states. This gives the creator more influence over the arc than a single unconstrained instruction.

The controls still need to be coherent. If keyframes contradict each other, ask for impossible transitions or change a character's proportions, the model must resolve the conflict somehow. More anchors are not automatically better; clear anchors that agree on identity, scene geometry and timing can be more effective than an overcrowded set.

The workspace becomes part of the advance

Kling's 4.0 announcement also describes a revised creation interface, timeline, node-based canvas and agent experience. These features organize assets and creative operations around a project. They should be distinguished from the underlying video's visual capability, but they affect whether a complex workflow is manageable. [61]

Kling's development illustrates the main transition in AI filmmaking: greater quality is being paired with greater direction. Existing native 4K, live Flash access and the forthcoming full thirty-second 4.0 experience belong in the same history, but each represents a different milestone.

09 / BUNCOSMOS AI ART

The wider video landscape

Google: Veo and the arrival of Omni

Google's Veo family became prominent in 2024. Veo 3 arrived in May 2025 with native sound, and Veo 3.1 added more creative control later that year. Flow served as the filmmaking environment connecting models, assets and scene-building. The advance was audiovisual production and editing, not only a higher score for a silent clip. [46, 105]

In May 2026, Google introduced **Gemini Omni**, beginning with Omni Flash for video creation and conversational editing. The developer release in June exposed Omni Flash through Google AI Studio and the Gemini API in preview. That release explicitly documented limitations, including ten-second generations and differences between the API's reference support and the wider product vision. [106, 69]

An August update brought **Omni 1.1 Flash** features into Flow, including start and end frames, draft generation and higher-resolution exports. The history therefore needs both the model story and the product-specific implementation. A feature in Flow is not proof that an earlier API preview supports the same workflow. [60]

Google's direction connects multimodal knowledge with generation: a model can draw on a description, existing media and broader semantic context when revising a scene. It remains necessary to distinguish visual plausibility from historical or scientific truth. A generated "documentary" scene can be illustrative without being evidence of an actual event.

Runway: generation, performance and video editing

Runway's Gen-3 Alpha announcement in June 2024 emphasized motion, consistency and temporal control. The current **Gen-4.5** documentation describes text-to-video and image-to-video, sequenced instructions, camera choreography, two-to-ten-second duration and 720p output. A cinematic appearance does not mean native 4K, and a tool's title should not be used as a substitute for its specification. [36, 107]

Runway also built performance-transfer and editing tools, including Act-One and later-generation offerings shown in its catalog. These complement generation because the creator may want to transfer a recorded performance or transform an existing clip instead

of inventing every gesture. At the cutoff, older Gen-2 and Gen-3 tools have retirement notices, while current documentation points to Gen-4.5 and newer editing products. [108, 109, 110]

The general lesson is that a creator-facing system can contain several specialized operations. A performance model, a video-editing model and a text-to-video generator may all contribute to one final shot. They should not be described as if a single button necessarily performs every task with equal precision.

Luma: control and professional post-production

Luma's Dream Machine and Ray model line contributed to the rapid rise of accessible AI video. The current video model at this cutoff is **Ray3.2**, released in June 2026. Its announcement emphasizes frame-level control, performance preservation, reframing and native HDR with 16-bit EXR export. This positions generation within professional finishing workflows, where the output must be composited and graded alongside other material. [97, 111]

Luma also develops the Uni image-model line and offers creative agents. Its Uni-1.1 API announcement in May 2026 reflects the expanding relationship between image reasoning, editing and production. The agents coordinate work across modalities and shared project assets, a separate capability from the video generator itself. [112, 113]

Pika and PixVerse: accessible motion and interactive directions

Pika's current product surfaces list **Pika 2.5**, creative effects and other transformation tools. Pikaformance focuses on expressive motion synchronized to sound. These offerings illustrate a creator-friendly branch of the market, where a specific effect or performance can be more immediately useful than an unrestricted scene generator. [114, 115, 116]

PixVerse's **V6**, launched in March 2026, supports several video-generation modes and multi-shot parameters. Its separate **R1** direction explores real-time interactive video, and the later **R2** announcement describes movement toward persistent, playable worlds. Offline generation of a finished clip and real-time response to interaction are different technical goals, even when both produce moving pictures. [117, 118]

An interactive system needs low latency and continuity while a user changes the situation. A traditional clip generator can spend longer producing a fixed result. Comparing them only by the prettiest still frame would miss what each system is designed to do.

MiniMax and Hailuo: multimodal video with open weights

MiniMax's Hailuo line developed through the 01 and 02 generations and the 2.3 release in 2025. Its current **MiniMax H3** was introduced on 31 July 2026, followed by an open-weight release on 3 August under a community license. H3 accepts multimodal context and generates video with native stereo sound, up to 15 seconds and 2K resolution in its published specification. [119, 120, 58]

H3's examples assign different roles to a camera reference, a character image and an audio reference. The model is designed to unify tasks that previously required separate specialists. Open weights also enable deployment and adaptation beyond a single hosted interface, subject to license and technical requirements. That does not make local deployment free of hardware costs or licensing obligations. [120, 58]

Alibaba: Wan, HappyHorse and a substantial image program

Alibaba's visual work includes Wan video models, Wan image models and Qwen-Image. Open and hosted releases occupy different roles in that ecosystem. Current Model Studio documentation lists **Wan 3.0** as an all-in-one model with native sound, two-to-thirty-second generation, first/last-frame control, references, editing and extension. It describes up to 20 multimodal reference materials, with access and endpoints varying by region. [2]

The catalog also includes **Wan 2.7** and **HappyHorse 1.1**, together with editing and motion-transfer modes. Alibaba introduced HappyHorse 1.0 in a limited beta in April 2026. These are related developer offerings, not grounds for assuming that every family uses identical architecture or has the same distribution model. [91, 121]

In images, **Wan2.7-Image** was announced in April 2026. **Qwen-Image 2.0**, documented in February 2026, emphasizes typography, semantic adherence and native 2K output. Alibaba's **Z-Image** family provides another efficient image-generation branch, with models intended for generation, adaptation and editing. These families make Alibaba's contribution wider than one video app. [122, 123, 124]

Vidu, Tencent and Z.ai: Asian development extends beyond two names

Vidu, developed by Shengshu, emphasizes reference-led video creation. Its current Q4 product page lists up to 16-second clips, as many as 15 image references and three audio references, with 4K maximum output. The API documentation separately identifies a Q4 preview, so the consumer product and a specific developer endpoint should not be treated as identical access surfaces. [125, 126]

Tencent's Hunyuan program includes video generation, image generation and 3D/world-model research. HunyuanImage 3.0 weights and code were released in September 2025, with instruction-focused variants in January 2026. A September 2026 **Hy Image 3.5 preview** announcement describes multi-image references, conversational editing and output up to 2K. A preview designation is part of the fact, not a detail to omit. [40, 127, 128]

Z.ai's CogVideo and CogVideoX contributed to open video development, while GLM-Image illustrates a hybrid approach to image creation with autoregressive semantic generation and a diffusion decoder. Their significance is not only another website for users to visit.

They supply alternative architectures and inspectable implementations that researchers and developers can build around. [39, 27]

Lightricks, Amazon, Meta and xAI broaden the map

Lightricks' **LTX** family connects audiovisual generation with an open model ecosystem and its studio products. LTX-2.3 became an important 2026 release, followed by **LTX-2.5** in August. The API documentation distinguishes Fast and Pro variants and their resolution limits; support is not identical across modes. Audio-to-video and targeted retakes are part of its production direction. [129, 130, 94]

Amazon's **Nova Canvas** and **Nova Reel** entered Bedrock in December 2024. Nova Reel 1.1 added multi-shot generation of videos up to two minutes in 2025. That is sequence-level output, and it should not be described as an uninterrupted two-minute shot. Amazon represents an enterprise deployment path as much as a consumer creativity story. [42, 131]

Meta's Movie Gen research explored personalized video, editing and sound, helping establish the broader research agenda. xAI's **Grok Imagine** represents another developer with image and audiovisual products; its June 2026 Video 1.5 release emphasized generation speed and practical video output. Current catalogs and release notes remain the right place to verify specific modes. [41, 132]

The result is a global field with many centres of development. American, Chinese, European, Canadian, Australian and Israeli contributors participate through different combinations of research, models, APIs and applications. A history of successful AI art needs that wider map, even when a creator uses only a small selection of tools.

10 / BUNCOSMOS AI ART

Creative agents and the connected studio

The model makes media; the agent organizes the work

A model is the engine that generates or transforms content. An application gives people a way to use that engine. An agent adds a layer that can interpret a goal, choose tools, carry out steps and use feedback to decide what to do next. These layers can exist in the same product, but they do different jobs.

Consider the instruction: "Create a short launch film for a new BunCosmos collection." A conventional generator might return one clip. A creative agent might first propose a concept, develop a shot list, identify the required reference assets, generate stills, animate

selected shots and prepare revisions. The difference is coordinated activity over time, rather than one media request.

This is already visible in actual products. Google's May 2026 Flow update introduced agents intended to support brainstorming, plot and dialogue development, and creation and editing within the filmmaking environment. Luma's March 2026 agent documentation describes a shared board connecting image, video, audio, voiceover and research tools. Its separate brainstorm and create modes distinguish planning from rendering. These are provider descriptions of product capabilities, rather than proof that every complex brief will succeed without intervention. [3, 133]

The agent's contribution can be organizational as much as visual. It can remember which candidate was approved, attach the correct reference, prepare alternate aspect ratios and keep related assets together. The underlying generator still determines much of the quality of each image or clip. An excellent planner cannot guarantee perfect fingers from an unsuitable image model or flawless motion from an inadequate video model.

All-in-one studios change the workflow

Another major development is the all-in-one creative studio: one interface offering several image and video engines, motion presets, reference management, editing, speech, music and export tools. Creators can move among models without rebuilding the entire project in separate applications. Some studios also develop proprietary components; others emphasize integration. A storefront's name therefore does not necessarily identify the developer of the model generating a particular shot. [134]

These environments increasingly accept a complete creative task rather than only a prompt. A music-video workflow, for example, can take a track, a character and a visual direction and assemble a sequence around them. The controls and degree of automation vary, but the product category is clearly moving toward production systems. [135]

This article names the principal model developers and discusses this integration category collectively. That distinction keeps the history understandable: the laboratory advances the engine, while the creative platform determines how people combine engines, preserve assets and finish work. Both are important, but a platform carrying the latest models has not necessarily trained those models itself.

What MCP actually contributes

Model Context Protocol, or **MCP**, was introduced by Anthropic in November 2024 as an open way to connect AI applications with external data and tools. It provides a common interface between an AI host and services that expose capabilities. In a creative setting, those capabilities might include submitting a generation, inspecting its status or retrieving a completed asset. [136]

The protocol's tool specification describes named operations with defined input schemas. A client can discover available tools and call them with structured arguments. MCP is

therefore a connection standard, not an image model, video codec or technique that automatically improves a render. It can help an assistant operate a creative service; the service's model still performs the generation. [137]

Vidu's own documentation provides a concrete example: its MCP integration exposes video workflows including text-to-video, image-to-video, reference-based generation and start/end-frame control. An assistant connected to this service can invoke the supported operations from a compatible environment. Separately, an all-in-one studio's agent integration documents model access, asynchronous job handling and reuse of previously generated assets. [138, 139]

An ordinary API can also support an agent. MCP is useful when a shared discovery-and-invocation layer makes connections easier; it is not a requirement for all automation. Likewise, a node graph is not automatically an agent. A graph can execute a fixed sequence. An agent can choose among steps based on the task and results. A production system may combine both.

From prompt to reviewable production

A useful creative loop starts with a clear brief, continues through generation and inspection, and returns to revision. The human defines the purpose and important constraints. The agent helps translate those into operations. A selected still can become an image-to-video input; a finished shot can become a reference for the next shot; a reviewer can reject a continuity error before it spreads across the sequence.

Long-running generation also creates practical requirements. The system needs to track jobs, wait for completion, associate outputs with the correct scene and handle failures. Repeatedly submitting the same job can waste credits. Passing the wrong reference can contaminate a whole batch. Production reliability includes these mundane details alongside visual quality.

Permissions matter too. An assistant allowed to generate a draft should not assume it may publish the campaign or send private reference files to any service it encounters. MCP's tools specification recommends visible tool activity and human control over invocations. For a creator, the relevant boundary is simple: approve what the system may use, what it may spend and what it may release. [137]

The near-term achievement is substantial: a conversation can increasingly coordinate several creative systems. The stronger claim that an agent can independently make a flawless film of any length is not established by the existence of these interfaces. Direction, review and conventional editing remain part of reliable work.

11 / BUNCOSMOS AI ART

A practical production journey

Begin with decisions that the model cannot make for you

Imagine a thirty-second BunCosmos film introducing Bun and Sol to a new audience. Before choosing a model, decide what the audience should feel and understand. Is this a warm welcome, a product announcement or an educational introduction? Which character details are essential? Where will the film appear? These decisions shape the script, framing, pace and final format.

Create a small approved asset set: the original character images, logo, palette, room reference and any indispensable props. Keep originals separate from generated derivatives. Name the approved versions clearly. A model can follow references more effectively when those references agree with one another.

For an image that must remain exactly unchanged, use conventional placement or compositing. Asking a generator to preserve Bun is an instruction about similarity; placing the original Bun image is preservation of the actual asset. This difference is especially important for a website illustration, a poster or a brand character whose posture and scene must remain intact.

Design the sequence before animating it

Write a short script and divide it into a few purposeful shots. An opening establishes the world. A middle shot introduces an action or idea. A final shot resolves it. Thirty seconds does not require thirty separate scenes. Fewer well-designed shots can give the viewer more time to notice character and meaning.

Develop still compositions for those shots. Check face, costume, proportions, products, lighting and background geometry. Correct a still before spending time animating it. If a reference image already has the wrong hand, video generation may carry that problem forward and add motion to it.

Choose motion appropriate to the shot. A gentle camera move and a small gesture may communicate more than simultaneous walking, spinning, object exchange and background transformation. The purpose of control is to express the idea, not to exercise every available parameter.

Generate, inspect and revise selectively

Use image-to-video when the opening composition is important, reference-based generation when identity must persist, and keyframes when the action has specific visual stages. Use video-to-video or performance transfer when an existing motion is the strongest foundation. These are complementary strategies rather than a single universal workflow.

Inspect the entire clip, not only its thumbnail. Look at hands during contact, props as they move behind other objects, mouth movement during speech and the background during camera changes. A clip can begin beautifully and fail halfway through. Listen separately for pronunciation, unwanted sounds, timing and emotional tone.

Repair the smallest useful unit. If one shot fails, replace that shot. If a supported editor can correct a brief region or action, test that before regenerating everything. Preserve approved assets so that each revision starts from a stable foundation. This is a practical editorial strategy, not a guarantee of pixel-perfect generative edits.

Finish as a film, not a folder of clips

Assemble the selected shots in an editor. Adjust pacing, transitions, music levels, titles and captions. Native generated sound may be useful, but it still needs a final mix. Create deliberate end cards rather than hoping the model will render every brand letter correctly throughout a moving scene.

Check the delivery specification: actual pixel dimensions, frame rate, colour space, HDR compatibility, sound levels and aspect ratio. A 4K file can still have a lower-detail source; an HDR master may need a separate standard-dynamic-range version for distribution. Check the exported file itself, not merely the settings selected before export.

An agent can assist at several points, especially with organizing variants and preparing revisions. The creator still makes the choices that give the sequence a coherent voice. The workflow becomes faster when generation, asset management and editing support one another.

12 / BUNCOSMOS AI ART

What has improved, and what still needs judgment

The advance is larger than photorealism

The last three years have improved several dimensions at once: visual quality, instruction following, typography, reference consistency, editing, motion, sound and workflow integration. A 2023 image could be impressive while resisting precise revision. By October 2026, creators can often direct an existing composition, carry references into motion and coordinate several operations inside a project.

The largest practical gain is often control. A beautiful picture that cannot be revised may be less useful than a slightly less spectacular one that can be adapted to the actual brief. Similarly, a video model's ability to preserve a product, reach a planned final frame and synchronize dialogue can matter more than its ability to make a dramatic camera movement.

Progress remains uneven. A model may excel at portraits and struggle with diagrams; generate persuasive texture while missing a requested count; preserve a face while altering a logo; or deliver credible motion until objects collide. Some of these failures become easier to see as resolution and duration increase.

How to read a benchmark or a model claim

A benchmark is a defined evaluation, not a verdict on every possible creative task. It may measure preference, text rendering, prompt adherence, motion or another dimension. Its prompts, scoring method and comparison set determine what the result means. A provider's showcase is also selected material. It demonstrates a possible result without telling the reader how frequently that result occurs.

For practical evaluation, use the same small set of representative briefs across candidate models. Include a demanding brand reference, a typography task, a composition requirement and a motion sequence that matters to the project. Record cost, time, usable-output rate and ease of repair. This provides a comparison aligned with the work rather than a search for a permanent universal winner.

Terms such as "world model" and "reasoning" describe research ambitions or mechanisms whose details differ by system. They should not be read as promises of complete physical understanding. Nor does "native" settle every quality question. Native audio can contain a poor performance; native 4K can depict an incorrect object in great detail.

Authorship, rights and trust

AI art raises several separate questions: who created the expressive work, what training material a model used, what a service's contract permits, and whether an output infringes another person's rights. A commercial-use subscription does not resolve all of those questions. A platform's permission to use an output and legal copyright protection for that output are different issues.

The U.S. Copyright Office's January 2025 report concluded that copyright protection requires sufficient human-determined expressive elements. Human-authored material, creative arrangement and modifications may support protection; merely supplying prompts does not by itself establish it. Using AI as an aid does not automatically disqualify the human-authored aspects of a larger work. This is a U.S. position and should not be generalized as identical law everywhere. [140]

Training rights are a separate debate. The Office's AI initiative addresses training in a different report from output copyrightability. Models also have different training and licensing policies: Adobe, for example, describes Firefly's approach using licensed and public-domain material. Such a provider statement explains its stated policy; it is not a universal legal conclusion about all outputs or all models. [141, 73]

Provenance tools help communicate origin. Google describes SynthID and C2PA-related measures for its image products; OpenAI documents provenance mechanisms for its current image generation. These can support identification and disclosure. They do not establish that a depicted event really happened, and absence of a visible label is not evidence that an image is authentic. [68, 1]

Creators can make this field more trustworthy by respecting consent, avoiding deceptive impersonation, documenting important sources and explaining when realistic material is illustrative. Keeping project records also helps show the human decisions made during a mixed AI-and-conventional production.

The human contribution becomes more visible through choices

Generative tools have lowered some technical barriers while creating new editorial responsibilities. More people can explore visual ideas, but abundance makes selection more important. A creator's contribution appears in the concept, references, staging, timing, revisions and final arrangement, not only in the first prompt.

The story of AI art is therefore a story of expanding creative access and expanding control. It began long before the current applications. It accelerated through generative learning, diffusion and multimodal systems. It now reaches into production environments where images, motion, sound and agents work together.

The most convincing work still has something to say. Models can supply possibilities at extraordinary speed. The creator decides which possibilities belong together, which errors matter, what should remain unchanged and when the work is ready to meet an audience.

13 / BUNCOSMOS AI ART

A compact glossary

Agent. A system that pursues a goal through tool use and multiple steps, potentially adjusting its plan using results and feedback.

API. An application programming interface: a structured way for software to request operations from another service.

Autoregressive generation. Creation of a sequence one unit at a time, conditioned on preceding units and other inputs.

Checkpoint. A saved set of model parameters representing a particular trained version.

Conditioning. Information that guides generation, such as a prompt, image, pose, audio or keyframe.

Diffusion. A family of generative approaches that learn to reverse a corruption process, commonly producing content through iterative denoising.

Embedding. A numerical representation used to encode information and relationships for a model.

Fine-tuning. Additional training that adapts a model to a narrower task, style or domain.

Flow matching. Training a model to learn a transformation between distributions, such as noise and the distribution of images.

Frame rate. The number of video frames shown per second. It is distinct from pixel resolution and clip duration.

HDR. High dynamic range, involving a wider brightness range than standard-dynamic-range presentation and requiring a compatible workflow.

Inference. Running a trained model to produce an output from inputs.

Inpainting. Filling or changing a selected region of an image, commonly with a mask and contextual guidance.

Keyframe. An important visual state used as an anchor for a sequence or transition.

Latent space. An internal numerical representation in which a system encodes information, often in a compressed form.

LoRA. Low-rank adaptation, an efficient method of training a relatively small set of additional parameters to adapt a model.

MCP. Model Context Protocol, a standard for connecting AI applications to external tools and information.

Model weights. The learned numerical parameters that govern a model's computations.

Native resolution. The resolution at which the content is generated, as distinct from later enlargement.

Open weights. Released model parameters. The term alone does not guarantee open-source licensing, released training data or unrestricted use.

Outpainting. Generating content beyond an image's existing boundaries.

Reference. An input asset used to guide identity, appearance, composition, motion or another feature.

Seed. A value that initializes aspects of random sampling; its effect depends on the model and implementation.

Temporal consistency. Coherence across time, including identity, geometry, objects, lighting and motion.

Token. A unit of representation processed by a model; depending on the system, it can represent text or aspects of other modalities.

Upscaling. Increasing output dimensions after generation, with interpolation or an enhancement model.

14 / BUNCOSMOS AI ART

Sources and research notes

This article was researched through **11 October 2026**. Its strongest emphasis is the 2023-2026 acceleration, with earlier research included to explain how that acceleration became possible. It covers major influential model families and representative creative workflows, rather than claiming a complete inventory of every app, regional deployment or experimental release.

Numbered links identify the primary papers, official announcements, product documentation and authoritative policy sources used. Product specifications and access conditions are snapshots at the research cutoff. A preview, early-access model or announced feature is labelled accordingly. Vendor performance statements are presented as claims or descriptions, not independent proof of universal superiority.

Explanations and the BunCosmos production example are editorial synthesis. The hypothetical workflow is not a report of a commissioned film or a claim that every named

service offers every operation. The original Bun artwork on the cover is supplied by BunCosmos and placed unchanged.

Reference library

1. Introducing ChatGPT Images 2.5

openai.com • Accessed 11 October 2026

2. How to use Wanx wan3.0 video generation model - Alibaba Cloud Model Studio - Alibaba Cloud Documentation Center

alibabacloud.com • Accessed 11 October 2026

3. Introducing Gemini Omni for Google Flow and Flow Music

blog.google • Accessed 11 October 2026

4. Digital art · V&A

vam.ac.uk • Accessed 11 October 2026

5. Collection Selection Boxes - Artificial Intelligence · V&A

vam.ac.uk • Accessed 11 October 2026

6. A history of artificial intelligence in 10 objects · V&A

vam.ac.uk • Accessed 11 October 2026

7. Generative Adversarial Networks

arxiv.org • Accessed 11 October 2026

8. Inceptionism: Going Deeper into Neural Networks

research.google • Accessed 11 October 2026

9. A Neural Algorithm of Artistic Style

arxiv.org • Accessed 11 October 2026

10. Artbreeder

artbreeder.com • Accessed 11 October 2026

11. AI Art: Free AI Art Generator & Community

nightcafe.studio • Accessed 11 October 2026

12. How AWS supported WOMBO's wildly popular, AI-powered app | AWS Startups

startups.aws.com • Accessed 11 October 2026

13. Denoising Diffusion Probabilistic Models

arxiv.org • Accessed 11 October 2026

14. CLIP: Connecting text and images

openai.com • Accessed 11 October 2026

15. DALL-E: Creating images from text

openai.com • Accessed 11 October 2026

16. DALL·E 2

openai.com • Accessed 11 October 2026

17. DALL·E now available without waitlist

openai.com • Accessed 11 October 2026

18. High-Resolution Image Synthesis with Latent Diffusion Models

arxiv.org • Accessed 11 October 2026

19. Stable Diffusion Public Release - Stability AI

stability.ai • Accessed 11 October 2026

20. Stable Diffusion launch announcement - Stability AI

stability.ai • Accessed 11 October 2026

21. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation

arxiv.org • Accessed 11 October 2026

22. LoRA: Low-Rank Adaptation of Large Language Models

arxiv.org • Accessed 11 October 2026

23. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis

arxiv.org • Accessed 11 October 2026

24. Scalable Diffusion Models with Transformers

arxiv.org • Accessed 11 October 2026

25. Research | Black Forest Labs

bfl.ai • Accessed 11 October 2026

26. Addendum to GPT-4o System Card: Native image generation

cdn.openai.com • Accessed 11 October 2026

27. GLM-Image: official repository

github.com • Accessed 11 October 2026

28. Adding Conditional Control to Text-to-Image Diffusion Models

arxiv.org • Accessed 11 October 2026

29. Edit Model - Midjourney

docs.midjourney.com • Accessed 11 October 2026

30. Image generation

developers.openai.com • Accessed 11 October 2026

31. Announcing SDXL 1.0 - Stability AI

stability.ai • Accessed 11 October 2026

32. Adobe Unveils Firefly, a Family of new Creative Generative AI

news.adobe.com • Accessed 11 October 2026

33. DALL·E 3

openai.com • Accessed 11 October 2026

34. Version - Midjourney

docs.midjourney.com • Accessed 11 October 2026

35. Video generation models as world simulators

openai.com • Accessed 11 October 2026

36. Runway Gen-3 Alpha | AI Video Generation with Temporal Control

runwayml.com • Accessed 11 October 2026

37. Kuaishou Unveils Proprietary Video Generation Model 'Kling;' Testing Now Available | Kuaishou Technology

ir.kuaishou.com • Accessed 11 October 2026

38. Kuaishou Launches Full Beta Testing for 'Kling AI' to Global Users, Elevates Model Capabilities | Kuaishou Technology

ir.kuaishou.com • Accessed 11 October 2026

39. CogVideo and CogVideoX: official repository

github.com • Accessed 11 October 2026

40. Tencent HunyuanVideo: official repository

github.com • Accessed 11 October 2026

41. Meta Movie Gen

ai.meta.com • Accessed 11 October 2026

42. Announcing Amazon Nova foundation models available today in Amazon Bedrock - AWS

aws.amazon.com • Accessed 11 October 2026

43. Introducing 4o Image Generation

openai.com • Accessed 11 October 2026

44. Introducing our latest image generation model in the API

openai.com • Accessed 11 October 2026

45. Nano Banana: Image editing in Google Gemini gets a major upgrade

blog.google • Accessed 11 October 2026

46. Fuel your creativity with new generative media models and tools

blog.google • Accessed 11 October 2026

47. Introducing Our V1 Video Model

updates.midjourney.com • Accessed 11 October 2026

48. ByteDance Seed: Seedance 1.0 technical announcement

seed.bytedance.com • Accessed 11 October 2026

49. FLUX.1 Kontext | Black Forest Labs

bfl.ai • Accessed 11 October 2026

50. FLUX.2: Frontier Visual Intelligence | Black Forest Labs

bfl.ai • Accessed 11 October 2026

51. Nano Banana Pro: Gemini 3 Pro Image model from Google DeepMind

blog.google • Accessed 11 October 2026

52. The new ChatGPT Images is here

openai.com • Accessed 11 October 2026

53. Kling-Omni Technical Report

arxiv.org • Accessed 11 October 2026

54. Nano Banana 2 - Gemini AI image generator & photo editor

gemini.google • Accessed 11 October 2026

55. ByteDance Seed: Introducing Seedream 5.0 Pro

seed.bytedance.com • Accessed 11 October 2026

56. Ideogram 4.5

ideogram.ai • Accessed 11 October 2026

57. Seedance 2.5

seed.bytedance.com • Accessed 11 October 2026

58. Open General Intelligence: MiniMax H3 Is Now Open Source - MiniMax News

minimax.io • Accessed 11 October 2026

59. FLUX 3 Video, Part 1: Generation | Black Forest Labs

bfl.ai • Accessed 11 October 2026

60. Google Flow brings new creative control features to enhance video editing.

blog.google • Accessed 11 October 2026

61. Introducing Kling 4.0: Visual Realism, Creative Control, and Next-Level Storytelling

klings.ai • Accessed 11 October 2026

62. What to know about the Sora discontinuation

help.openai.com • Accessed 11 October 2026

63. Introducing ChatGPT Images 2.0

openai.com • Accessed 11 October 2026

64. Video - Midjourney

docs.midjourney.com • Accessed 11 October 2026

65. Google Research at Google I/O 2024

research.google • Accessed 11 October 2026

66. Nano Banana AI image editing coming to Lens, Photos and more

blog.google • Accessed 11 October 2026

67. Developers can build with Nano Banana Pro (Gemini 3 Pro Image)

blog.google • Accessed 11 October 2026

68. Nano Banana 2: Google's latest AI image generation model

blog.google • Accessed 11 October 2026

69. Start building with Nano Banana 2 Lite and Gemini Omni Flash

blog.google • Accessed 11 October 2026

70. Bringing Generative AI into Creative Cloud with Adobe Firefly

blog.adobe.com • Accessed 11 October 2026

71. Adobe Expands Generative AI Offerings Delivering New Firefly App with Industry's First Commercially Safe Video Model

news.adobe.com • Accessed 11 October 2026

72. Explore the new Adobe Firefly, your all-in-one creative AI studio

blog.adobe.com • Accessed 11 October 2026

73. Adobe Firefly - Free Generative AI for Creatives

adobe.com • Accessed 11 October 2026

74. Adobe Firefly | Comprehensive & Commercially Safe AI Content Creation for Businesses

business.adobe.com • Accessed 11 October 2026

75. Introducing Stable Diffusion 3.5 - Stability AI

stability.ai • Accessed 11 October 2026

76. FLUX 3: Multimodal Video, Image & Audio | Black Forest Labs

bfl.ai • Accessed 11 October 2026

77. Pricing | Black Forest Labs

bfl.ai • Accessed 11 October 2026

78. Models | Black Forest Labs

bfl.ai • Accessed 11 October 2026

79. Ideogram 4.0 Press Release

ideogram.ai • Accessed 11 October 2026

80. Ideogram 4.0 Technical Details: Open model at the forefront of design

ideogram.ai • Accessed 11 October 2026

81. Ideogram 4.0

ideogram.ai • Accessed 11 October 2026

82. Introducing Recraft V4: Design Taste Meets AI

recraft.ai • Accessed 11 October 2026

83. Phoenix by Leonardo.Ai | Foundational AI Image Generation Model

leonardo.ai • Accessed 11 October 2026

84. Introducing Lucid Origin - Leo's Most Versatile AI Model | Leonardo.Ai

leonardo.ai • Accessed 11 October 2026

85. Generative AI Models for Image & Video Creation | Leonardo.Ai

leonardo.ai • Accessed 11 October 2026

86. ERNIE-ViLG 2.0: Improving Text-to-Image Diffusion Model with Knowledge-Enhanced Mixture-of-Denoising-Experts

arxiv.org • Accessed 11 October 2026

87. Baidu ERNIE-Image: official repository

github.com • Accessed 11 October 2026

88. DeepSeek Janus and Janus-Pro: official repository

github.com • Accessed 11 October 2026

89. Meitu releases seven image-production products and MiracleVision (19 June 2023)

preview-meitu.com • Accessed 11 October 2026

90. Meitu introduces MiracleVision 3.0 (9 October 2023)

meitu.com • Accessed 11 October 2026

91. Video generation and editing - Alibaba Cloud Model Studio - Alibaba Cloud Documentation Center

alibabacloud.com • Accessed 11 October 2026

92. Kling 4.0 multiple-keyframe guide

kling.ai • Accessed 11 October 2026

93. Seedance 2.0: Advancing Video Generation for World Complexity

arxiv.org • Accessed 11 October 2026

94. LTX-2: Efficient Joint Audio-Visual Foundation Model

videos.ltx.io • Accessed 11 October 2026

95. Supported Models | LTX Documentation

docs.ltx.io • Accessed 11 October 2026

96. Kling AI Native 4K Video Model: True 4K AI Generation

kling.ai • Accessed 11 October 2026

97. Luma Introduces Ray3.2 Model & API: Complete Creative Control for Video Generation

lumalabs.ai • Accessed 11 October 2026

98. Dreamina: Free AI Video & Image Generator Powered by Multiple Models

dreamina.capcut.com • Accessed 11 October 2026

99. ByteDance Seed

seed.bytedance.com • Accessed 11 October 2026

100. ByteDance Seed: Introducing Seedance 2.5

seed.bytedance.com • Accessed 11 October 2026

101. Official Seedance 2.5: 4K & 30s AI Video Generator with Audio

dreamina.capcut.com • Accessed 11 October 2026

102. Kuaishou Unveils Comprehensive AI Models, Reshaping Content and Commercial Ecosystem with AI | Kuaishou Technology

ir.kuaishou.com • Accessed 11 October 2026

103. Kling AI Advances to the 2.0 Era, Empowering Everyone to Tell Great Stories with AI | Kuaishou Technology

ir.kuaishou.com • Accessed 11 October 2026

104. Kling AI Launches 3.0 Model, Ushering in an Era Where Everyone Can Be a Director | Kuaishou Technology

ir.kuaishou.com • Accessed 11 October 2026

105. Bringing new Veo 3.1 updates into Flow to edit AI video

blog.google • Accessed 11 October 2026

106. 5 Takeaways from Google I/O 2026

blog.google • Accessed 11 October 2026

107. Creating with Gen-4.5 - Runway

help.runwayml.com • Accessed 11 October 2026

108. Creating with Act-One on Gen-3 Alpha and Turbo - Runway

help.runwayml.com • Accessed 11 October 2026

109. Gen-2 Deprecation - Runway

help.runwayml.com • Accessed 11 October 2026

110. Creating with Gen-3 Alpha and Gen-3 Alpha Turbo - Runway

help.runwayml.com • Accessed 11 October 2026

111. Luma - Official Information for AI Assistants

lumalabs.ai • Accessed 11 October 2026

112. Introducing the Uni-1.1 API: Intelligence You Can Direct. Aesthetic You Can Ship.

lumalabs.ai • Accessed 11 October 2026

113. The New Luma app: Your creative collaborator and the most capable multimodal Agent

lumalabs.ai • Accessed 11 October 2026

114. Pika

pika.art • Accessed 11 October 2026

115. Pika: product and Pikaformance overview

pika.art • Accessed 11 October 2026

116. Pika 2.5: image-to-video model documentation

mcp.pika.art • Accessed 11 October 2026

117. PixVerse V6 Model Introduction & Scenarios

docs.pixverse.ai • Accessed 11 October 2026

118. PixVerse Introduces R2, Advancing Its Real-Time World Model Toward Persistent, Playable Worlds | PixVerse

pixverse.ai • Accessed 11 October 2026

119. MiniMax Hailuo 2.3: A New Level of Complex Video ...

minimax.io • Accessed 11 October 2026

120. MiniMax H3: An Open Model Breaking the Boundaries Between Tasks and Modalities - MiniMax Research

minimax.io • Accessed 11 October 2026

121. Alibaba Rolls Out HappyHorse 1.0 in Limited Beta - Alibaba Cloud Community

alibabacloud.com • Accessed 11 October 2026

122. Alibaba Unveils Wan2.7 Redefining Personalized and Precision Image Creation - Alibaba Cloud

alibabacloud.com • Accessed 11 October 2026

123. Qwen-Image: official repository and model updates

github.com • Accessed 11 October 2026

124. Z-Image: official repository

github.com • Accessed 11 October 2026

125. Vidu Q4 AI Video Model with Native Audio | Vidu AI

vidu.com • Accessed 11 October 2026

126. Vidu Q4 Preview: reference-to-video API documentation

platform.vidu.com • Accessed 11 October 2026

127. Tencent HunyuanImage 3.0: official repository and updates

github.com • Accessed 11 October 2026

128. Tencent Hy Image 3.5 preview

hunyuan.tencent.com • Accessed 11 October 2026

129. March 5, 2026 | LTX Documentation

docs.ltx.io • Accessed 11 October 2026

130. API Changelog | LTX Documentation

docs.ltx.io • Accessed 11 October 2026

131. Amazon Nova Reel 1.1: Featuring up to 2-minutes multi-shot videos | AWS News Blog

aws.amazon.com • Accessed 11 October 2026

132. Grok Imagine Video 1.5

x.ai • Accessed 11 October 2026

133. The Luma Agent Explained

lumalabs.ai • Accessed 11 October 2026

134. Inside Higgsfield #1: How We Built Cinema Studio

higgsfield.ai • Accessed 11 October 2026

135. AI Music Video Creator - Generate Professional Music Videos Instantly | OpenArt

openart.ai • Accessed 11 October 2026

136. Introducing the Model Context Protocol \ Anthropic

anthropic.com • Accessed 11 October 2026

137. Tools - Model Context Protocol

modelcontextprotocol.io • Accessed 11 October 2026

138. Vidu MCP service guide

platform.vidu.com • Accessed 11 October 2026

139. Higgsfield Skills | AI Image & Video Generation for Any Agent

higgsfield.ai • Accessed 11 October 2026

140. U.S. Copyright Office: AI output copyrightability report announcement (29 January 2025)

copyright.gov • Accessed 11 October 2026

141. Copyright and Artificial Intelligence | U.S. Copyright Office

copyright.gov • Accessed 11 October 2026