Position: Advocating for using Motion Figures in your ML Submissions

Zhenyuan Chen [1] **Feng Zhang** [1]

Abstract

The rapid advancement of machine learning research—particularly in areas producing dynamic results such as video generation, 3D/4D scene synthesis, robotics, and world modeling—has created a fundamental disconnect between the dynamic nature of results and the static medium of traditional paper submissions. While researchers often resort to external websites or supplementary materials to display video results, this practice introduces friction for reviewers, risks archival decay through link rot, and raises concerns about double-blind review integrity. In this position paper, we propose a minimal specification for embedded dynamic figures (DFIG v1.0) that leverages the LaTeX `animate` package to present frame-sequence animations directly within PDF documents. We discuss technical feasibility, viewer compatibility constraints, and concrete guidelines for resolution, frame rate, file size, accessibility, and security. Rather than advocating immediate standardization, we present this as a call for community discussion and propose a pilot program for future venues. We argue the benefits of self-contained, archivally robust dynamic figures merit serious consideration as modern machine learning research increasingly depends on temporal evaluation.

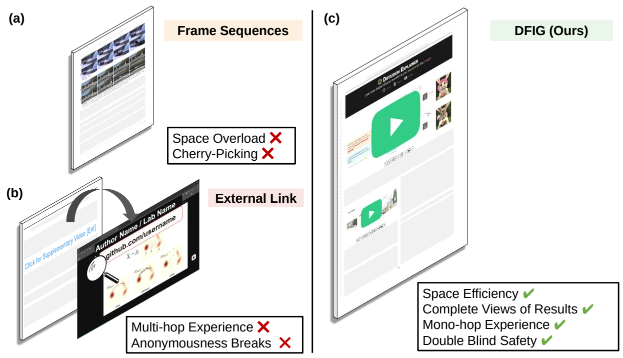

Figure 1. A comparison of between previous ML submission methods on motion figures and our proposal (DFIG v1.0).

1. Introduction

Generative models and dynamic learning systems have surged to the forefront of machine learning research, transforming how we approach visual understanding, synthesis, and multimodal interaction (He et al., 2025; Ma et al., 2025). Recent years have seen remarkable progress in video generation (Wang et al., 2025c; Waseem & Shahzad, 2025; Xue et al., 2025; Lin et al., 2025), 3D and 4D content generation (Kong et al., 2025; Miao et al., 2025; Wang et al., 2025a; Wen et al., 2025), speech synthesis (Cui et al., 2025; Xie et al., 2025b), world modeling (Ding et al., 2025; Xie et al., 2025a; Zhao et al., 2025a;b; Zhu et al., 2025), multimodal generation (Foo et al., 2025; Vilaça et al., 2025), robotics (Bai et al., 2025; Liu et al., 2025; Long et al., 2025; Shao et al., 2025; Xiang et al., 2025; Yu et al., 2025; Zhang et al., 2025a; Zheng et al., 2025; Zhong et al., 2025), and autonomous driving (Feng et al., 2025; Guan et al., 2024; Wang et al., 2025d; Xu et al., 2025).

As these inherently temporal and spatial domains continue to advance, the research community faces a fundamental challenge: ***how can we effectively present dynamic results in static PDF submissions?*** The current practice of using static "filmstrip" figures fails to capture temporal coherence, motion quality, and visual artifacts that are central to evaluating dynamic and temporal results. Researchers increasingly resort to external project websites or supplementary video files, but this workflow introduces significant problems: reviewers must context-switch between documents and browsers, external links suffer from archival decay ("link rot"), and hosting videos on identifiable servers can compromise double-blind review integrity.

We argue that embedded dynamic figures—frame-sequence animations that play directly within PDF documents—offer a promising solution to this disconnect. By leveraging the LaTeX `animate` package, authors can embed video-like content that remains self-contained within the submission PDF, eliminating external dependencies while preserving the temporal information essential for fair evaluation.

[1]Zhejiang University, Hangzhou, China. Correspondence to: Feng Zhang <zfcarnation@zju.edu.cn>.

Preprint. April 2, 2026.

Rather than advocating for immediate mandatory adoption, this position paper proposes a **minimal specification for Dynamic Figures (DFIG v1.0)** as a starting point for community discussion. We present concrete technical guidelines, acknowledge limitations honestly, and suggest a path toward gradual adoption through pilot programs and iterative refinement based on community feedback.

Roadmap. Section 2 surveys related work on dynamic visualization in scientific publishing and relevant standards. Section 3 motivates the need for dynamic figures in modern machine learning research. Section 4 presents our technical proposal and the DFIG v1.0 specification. Section 5 discusses benefits, limitations and mitigations. Section 6 demonstrates alternative views against our proposal. Section 7 outlines a path to adoption.

2. Related Work

Our proposal builds upon prior work in dynamic visualization for scientific publishing, human-computer interaction research on video interfaces, and evolving standards for scholarly document formats.

2.1. Dynamic Visualization in Scientific Publishing

The challenge of representing multi-dimensional and temporal data in static documents is not unique to machine learning. The astronomy community recognized this limitation early, with Barnes & Fluke (2008) pioneering techniques for embedding interactive 3D visualizations of astronomical datasets directly within PDF research papers. Their approach, using Adobe's PDF extensions and the S2PLOT library, enabled readers to rotate and explore 3D models of redshift catalogues, cosmological simulations, and instrument designs without leaving the document. Similarly, the biomedical sciences have embraced interactive 3D PDF technology, with Newe (2016) developing comprehensive toolboxes for creating ready-to-publish figures with embedded 3D anatomical models.

Notably, major journals have institutionalized these practices. *Astronomy & Astrophysics* (A&A) provides detailed author guidelines[1] for embedding multimedia content including videos (in .mov, .avi, .mpg, .mp4 formats up to 10MB) and 3D models (U3D/PRC formats) using the `media9` LaTeX package. These precedents demonstrate both the technical feasibility and the scientific value of embedding dynamic content directly within scholarly documents—a capability that remains largely unexplored in the machine learning community despite the inherently temporal nature of modern generative research.

[1] https://www.aanda.org/for-authors/latex-issues/multi-media

2.2. PDF Standards and Digital Library Policies

Any proposal for embedded dynamic content must contend with document standards and archival policies. PDF/A, the ISO-standardized archival format, prohibits JavaScript and embedded multimedia to ensure long-term accessibility and security (Abrams & Levenson, 2004). Major digital libraries, including those maintained by ACM and IEEE, typically sanitize uploaded PDFs, removing active content that could pose security risks.

These constraints are real and inform our specification design. The `animate` package, which embeds frame sequences rendered as static images with JavaScript-controlled playback, represents a minimal footprint approach: the frames themselves are archivable as static content, and the JavaScript merely controls display timing. We discuss how dual-format archival (PDF with static fallbacks plus sidecar frame sequences) can satisfy both dynamic viewing and archival requirements in Section 5.

2.3. Temporal Evaluation in Generative Models

The importance of temporal evaluation in generative research has been increasingly recognized. Benchmarks such as RBench (Deng et al., 2026) and world model evaluations (Fan et al., 2026) emphasize spatiotemporal consistency and physical plausibility as key metrics that cannot be assessed from static frames alone. Video quality metrics increasingly focus on temporal artifacts like flickering, motion blur, and frame-to-frame consistency (Wang et al., 2025c). This evaluation challenge motivates our proposal: if temporal coherence is central to the scientific contribution, the publication format should enable reviewers and readers to directly observe temporal phenomena rather than relying on verbal descriptions or cherry-picked frame sequences.

3. Motivation: Why Dynamic Figures Matter

Motivation

Current static submission formats and external supplements are insufficient for modern machine learning research.

In subfields such as video generation, 4D reconstruction, and robotics, the core contribution often lies in the **temporal consistency** and **realistic motion** of the content. Static "filmstrip" figures fail to capture these nuances, hiding artifacts like flickering, temporal incoherence, or unnatural motion that are immediately apparent when viewing the actual video output. Several critical limitations are shown as below.

Page Constraints Limit Comprehensive Frame Display. Current papers displaying video generation results through discrete frames are constrained by page limitations, forcing authors to show only a sparse subset of frames—typically 4–8 frames from videos containing hundreds. This fundamentally fails to convey the full dynamic nature of the content. Consider a 5-second video at 24 fps: showing 6 representative frames means displaying only 5% of the temporal information. Critical artifacts like mid-sequence flickering, motion blur inconsistencies, or temporal aliasing can easily fall between the shown frames. Dynamic figures, in contrast, can demonstrate all frames within a compact spatial footprint, enabling comprehensive evaluation of temporal dynamics.

Fragmented Review Experience and Cognitive Load. To bridge the gap left by static figures, the community has relied on supplementary materials hosted on external platforms. While modern platforms like OpenReview now support hosting anonymized video supplements—mitigating the immediate risk of identity exposure—this workflow remains suboptimal. It forces a fragmented reading experience where reviewers must constantly context-switch between the manuscript PDF and browser tabs, disrupting narrative flow and increasing cognitive load. Reviewers may need to manually synchronize textual descriptions with corresponding video segments, a tedious process that can lead to incomplete evaluation. Anecdotally, reviewers report sometimes skipping supplementary videos entirely due to this friction, potentially missing important evidence of temporal quality.

Cherry-Picking and Comparison Fairness. Static frame selection introduces opportunities for cherry-picking. Authors naturally select frames that best showcase their method's strengths, potentially avoiding frames where artifacts are visible. This asymmetry undermines fair comparison: when comparing against baselines, authors may (consciously or not) select less favorable frames for competing methods. Dynamic figures level the playing field by showing complete temporal sequences, making it harder to hide weaknesses and easier for reviewers to make fair assessments across methods.

Archival Decay and Link Rot. External supplements suffer from archival fragility. Project websites go offline, cloud storage links expire, and institutional servers are decommissioned. A survey of computer vision papers from 5+ years ago reveals that a substantial fraction of project page links are now broken, leaving readers with only the static figures in the archived PDF. This "link rot" problem means that the most important evidence for temporal research—the actual video results—may be permanently lost. Embedded dynamic figures, by contrast, remain part of the archival PDF, ensuring that temporal evidence persists as long as the paper itself.

The Case for Embedded Dynamics. These limitations collectively argue for a shift toward embedded dynamic figures. By including frame-sequence animations directly within PDF documents, we can: (1) show complete temporal information without page constraints; (2) eliminate context-switching friction for reviewers; (3) reduce opportunities for cherry-picking; and (4) ensure archival robustness. The technical feasibility of this approach has been demonstrated in other scientific disciplines and in recent machine learning submissions, as we discuss in the following section.

4. Proposal: The DFIG v1.0 Specification

We propose a minimal specification for Dynamic Figures in scientific publications, which we call DFIG v1.0. This specification is designed to be technically feasible with existing tools, compatible with double-blind review requirements, and conservative in its security footprint. We present it as a starting point for community discussion rather than a final standard.

4.1. Technical Foundation

Technical Foundation

The `animate` LaTeX package provides a technically feasible method to embed dynamic content directly in PDFs by embedding frame sequences as static images with JavaScript-controlled playback.

The technical feasibility of embedded dynamic figures has been demonstrated in recent machine learning submissions (Zhang et al., 2025b; Hyung et al., 2025; Yun & Choo, 2025; Kim et al., 2025). The core solution uses the `animate` LaTeX package[2], which embeds frame sequences directly into the PDF as a series of static images with JavaScript-controlled playback. This approach offers several advantages over alternatives: (1) **No external dependencies**: all frames are embedded within the PDF file itself; (2) **Minimal JavaScript**: only frame-stepping logic is required, with no network access or file system operations; (3) **Graceful degradation**: unsupported viewers display the first frame as a static fallback; and (4) **Wide toolchain support**: it works with standard LaTeX distributions (TeX Live, MiKTeX) without additional software.

4.2. DFIG v1.0 Specification

We propose the following constraints for dynamic figures intended for conference submissions. These constraints

[2] https://gitlab.com/agrahn/media9

balance expressiveness with security, accessibility, and practical file size limits.

4.3. Implementation Workflow

The workflow for creating DFIG-compliant dynamic figures involves three steps:

Step 1: Generate Frame Sequence. Extract frames from video results in-line using a command such as: `ffmpeg -i video.mp4 -vf "fps=10,scale=720:-1" frames/frame_%04d.jpg` This command extracts frames at 10 fps, scaled to 720p width. For file size optimization, using JPEG output (`.jpg`) is preferable for photographic content.

Step 2: Embed with `animate`. Use the `\animategraphics` command in LaTeX (see Figure 2):

Step 3: Verify and Optimize. Test the PDF in Adobe Acrobat Reader (free) to verify playback. Check file size and optimize frame compression if needed. Verify that the first frame provides a meaningful static fallback.

4.4. Case Studies

Video Generation. Figure 3 demonstrates temporal evolution from video generation research (Zhang et al., 2025b), enabling direct observation of motion smoothness and temporal consistency that would require dozens of static frames to partially convey.

Pedagogical Visualization. Figure 4 shows an animated Stable Diffusion denoising process (Lee et al., 2024), explaining complex processes more effectively than static diagrams.

3D Reconstruction. Figure 5 demonstrates turntable visualization of 3D reconstruction (Wang et al., 2025b), revealing geometric consistency from multiple angles without requiring interactive 3D viewers.

4.5. Sidecar Artifact Proposal

To address viewer compatibility limitations (see Section 5), we propose an optional **sidecar artifact**: a supplementary ZIP containing all frame sequences as numbered images, a `manifest.json` (storing metadata such as fps, descriptions, checksums), and pre-rendered GIF or WebP previews for browser viewing. This artifact complements the animated PDF by enabling browser-based preview (e.g., for PDF.js), archiving raw frames, and allowing content verification, while ensuring that embedded animations in the PDF remain the primary submission.

This dual-artifact approach mirrors practices in astronomy journals, where both embedded multimedia PDFs and separate media files are archived together.

5. Discussion: Benefits, Limitations, and Mitigations

Adopting embedded dynamic figures would offer significant benefits for communicating temporal research, but faces real limitations that must be acknowledged honestly. We discuss both sides and propose mitigations.

5.1. Benefits

Embedded dynamic figures offer several advantages. They allow reviewers to assess temporal quality without leaving the document, reducing context-switching friction compared to opening supplementary videos in separate browser tabs. Unlike external project websites that frequently go offline, embedded figures remain part of the archival PDF indefinitely, eliminating the “link rot” problem. A single dynamic figure can convey information that would require many static frames arranged in grids—a 5-second animation at 10 fps contains 50 frames of information in the space of a single figure, valuable given strict page limits. By showing complete temporal sequences rather than selected frames, dynamic figures reduce opportunities for cherry-picking and allow reviewers to observe the full output, including any artifacts or failures that might be hidden in carefully selected static frames.

5.2. Viewer Compatibility Limitations

The primary challenge is **incomplete viewer support**. The `animate` package requires JavaScript execution, which is not supported by all PDF viewers. Table 2 summarizes current compatibility. Several mitigations address this limitation: conference organizers can provide explicit instructions advising reviewers to use Adobe Acrobat Reader (free) or Foxit Reader, analogous to existing instructions for viewing supplementary videos; the DFIG v1.0 specification requires that the first frame be representative and that captions describe temporal content, ensuring that even reviewers using unsupported viewers can evaluate submissions with reduced fidelity; and the proposed sidecar ZIP (see Section 4) provides GIF/WebP previews that can be viewed in any browser, bridging the gap for PDF.js users.

5.3. Security Considerations

PDF documents can execute JavaScript, raising legitimate security concerns. Enterprise environments and security-conscious institutions often disable PDF JavaScript by default. The `animate` package uses JavaScript solely for frame-stepping—cycling through embedded static images

Table 1. DFIG v1.0 specification constraints for embedded dynamic figures.

Constraint	Specification
Allowed Technology	`animate` package only (frame sequences). No `media9` video embedding, no Flash-based players, no external network access.
Resolution	Maximum 1280×720 pixels (720p). Lower resolutions recommended for file size efficiency.
Frame Rate	5–15 fps recommended; 30 fps maximum. Higher frame rates rarely needed for scientific demonstration.
Per-Figure Size	Maximum 10 MB per dynamic figure (embedded frames).
Total PDF Size	Should respect venue submission limits (typically 50–100 MB).
Mandatory Fallback	First frame must be representative; caption must describe temporal content for readers using unsupported viewers.
User Controls	Recommend `controls` option to allow pause/play/scrub, respecting user agency.
Accessibility	Caption must include text description of motion content. Avoid strobing effects (>3 Hz).
Anonymization	Strip EXIF metadata from frames. No watermarks, URLs, or identifiable logos.
Licensing	Authors must have rights to embed all visual content.

Template: DFIG-Compliant Frame Sequence Animation

```
    % Preamble requirement:
    % \usepackage[method=ocg]{animate}

    \begin{figure*}[t]
        \centering
        \animategraphics[
            type=png,            % or jpg for photographic content
            width=\linewidth,    % max 720p (1280x720)
            autoplay,            % starts on page view
            loop,                % continuous playback
            controls,            % user pause/play/scrub
            every=5              % every 5th frame (~20\%) to limit PDF size
        ]{2}{figures/frames/frame_}{0001}{0100}
        % {fps}{path/prefix}{first_frame}{last_frame}

        \caption{Description of temporal content visible in animation.
        The animation shows [describe motion/process].
        \textbf{Accessibility:} [Text description for screen readers].
        \textbf{Viewing:} Use Adobe Acrobat Reader or Foxit Reader
        for animation playback; other viewers display the first frame.}
        \label{fig:example_dfig}
\end{figure}
```

Figure 2. Template code for creating a DFIG-compliant frame sequence animation using the `animate` package.

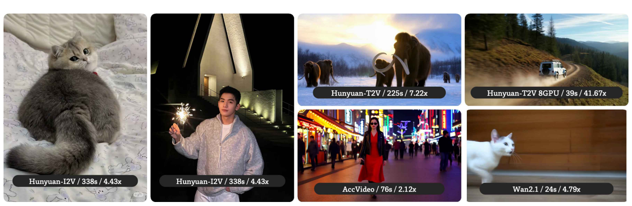

Figure 3. Dynamic figure example derived from (Zhang et al., 2025b). The source assets are from the paper's project website. Please use **Adobe Acrobat Reader** for a live video visualization.

at specified intervals. The JavaScript code is minimal, does not access the network, does not read or write files, and cannot execute arbitrary code. Compared to legacy approaches (e.g., `media9` with Flash-based players), frame-sequence animation represents a minimal-risk approach. PDF/A, the archival standard, prohibits JavaScript and dynamic content; however, `animate`-based PDFs degrade gracefully as the embedded frames are standard images that remain visible and archivable even when JavaScript is disabled. For reviewers in institutions that block PDF JavaScript entirely, the sidecar artifact and mandatory fallbacks ensure that review remains possible, if less convenient.

5.4. Accessibility Considerations

Dynamic content introduces accessibility challenges that the DFIG specification addresses. Screen readers cannot in-

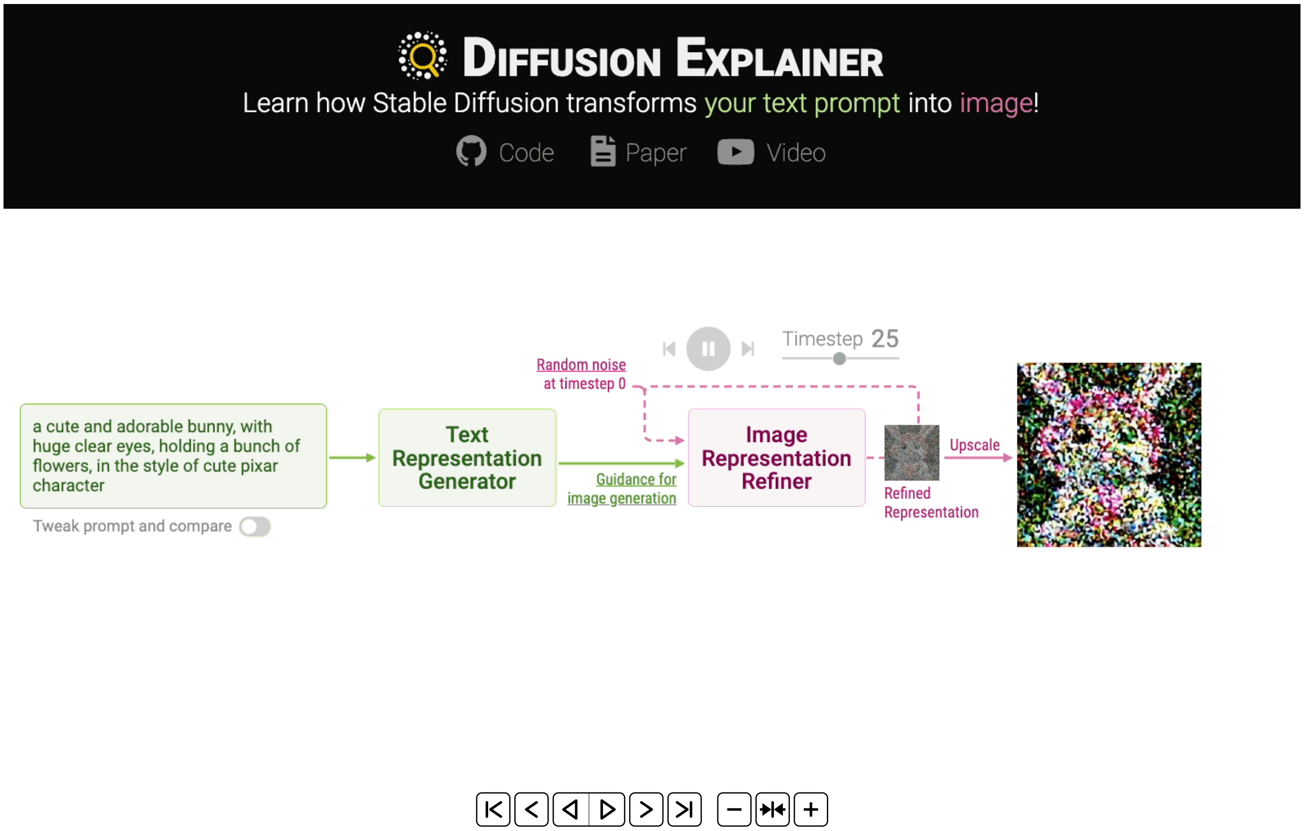

Figure 4. Visual explanation of the text-to-image Stable Diffusion denoising process, derived from Diffusion Explainer (Lee et al., 2024). The animation illustrates how noise is progressively removed to generate coherent images. **Accessibility note:** The animation shows iterative refinement from noise to a clear image over 100 steps. **Viewing:** Use Adobe Acrobat Reader or Foxit Reader for animation playback; other viewers will display the first frame.

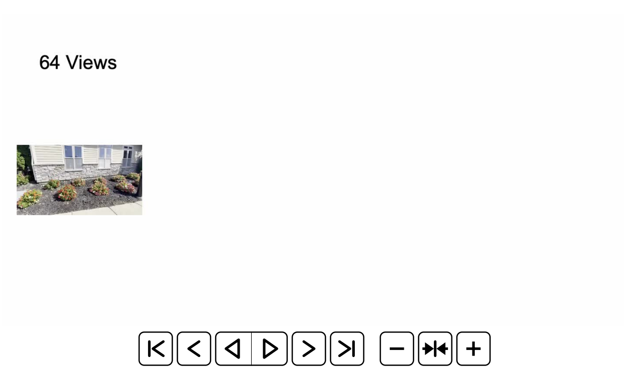

Figure 5. 3D reconstruction visualization (turntable animation), derived from VGGT (Wang et al., 2025b). The rotating view reveals reconstruction quality from multiple angles, demonstrating geometric consistency that is difficult to assess from static views. **Viewing:** Use Adobe Acrobat Reader or Foxit Reader for animation playback.

Table 2. PDF viewer compatibility with embedded dynamic figures (`animate` package). Fallback behavior indicates what users see in unsupported viewers.

Software	Platform	Animation	Fallback Behavior
Adobe Acrobat Reader	Win, Mac	Full support	N/A
Adobe Acrobat Pro	Win, Mac	Full support	N/A
Foxit Reader	Win, Mac, Linux	Full support	N/A
PDF-XChange Editor	Windows	Full support	N/A
Okular	Linux, Win, Mac	Partial	Shows first frame
PDF.js (OpenReview)	Web browsers	No	Shows first frame
Apple Preview	macOS	No	Shows first frame
Sumatra PDF	Windows	No	Shows first frame
Mobile viewers	iOS, Android	Varies	Usually first frame

terpret animated visual content; the DFIG v1.0 requirement for descriptive captions ensures that readers using assistive technology receive a textual description of the temporal content. Some readers experience discomfort from animation (e.g., vestibular disorders), so the `controls` option allows users to pause playback, and the DFIG specification prohibits strobing effects above 3 Hz to avoid triggering photosensitive reactions. Operating systems increasingly support "reduced motion" preferences; while future PDF viewers may respect these settings, currently the controls option provides user agency. The current DFIG v1.0 specification deliberately excludes audio embedding due to technical limitations; for research involving audio (e.g., text-to-speech, audio-visual generation), external supplements with transcripts remain the recommended approach until better in-PDF audio solutions emerge.

5.5. Archival and Digital Library Compatibility

Major digital libraries (ACM DL, IEEE Xplore, arXiv) have varying policies on PDF content. Many libraries sanitize uploaded PDFs, potentially stripping JavaScript or active content, which could remove animation functionality while preserving the static frame fallbacks. We propose that venues adopting DFIG archive both: (1) the original PDF with embedded animations; and (2) the sidecar artifact containing raw frames and metadata. This ensures that temporal content is preserved even if PDF sanitization removes animation capability. Astronomy & Astrophysics and other journals already archive multimedia alongside PDFs, demonstrating that dual-format archival is operationally feasible for scholarly publishers.

6. Alternative Views

While we advocate for embedded dynamic figures as a solution to the temporal communication gap, a credible alternative position merits serious consideration.

Alternative View

External video supplements hosted on platforms like OpenReview or project websites are superior to embedded animations. This view challenges the fundamental premise of our proposal and deserves a thorough response.

Proponents of external video supplements argue that this approach offers several decisive advantages over embedded animations. First, **modern video codecs (H.264, VP9, AV1) provide superior compression** compared to frame sequences, enabling higher resolution and frame rates within reasonable file sizes. A 1080p video at 30 fps can be compressed to 2–5 MB using modern codecs, whereas embedding 90 frames as JPEGs might require 10–15 MB. Second, **universal compatibility**: any device with a web browser can play videos without requiring specific PDF viewers, eliminating the viewer compatibility issues that plague embedded animations. Third, **easier authoring workflow**: authors can directly upload video files without frame extraction, conversion, and optimization steps. Fourth, **better accessibility features**: video players offer standard controls (playback speed adjustment, captions, transcripts) that PDF animations lack. Finally, **scalable infrastructure**: video hosting platforms can optimize delivery via CDNs and adaptive streaming, whereas embedded animations bloat individual PDF files.

We acknowledge these advantages as legitimate and substantial. Modern video codecs do achieve better compression; browser-based playback is indeed more universal than PDF viewer support; and video authoring workflows are simpler than frame extraction pipelines. However, embedded animations address critical limitations that external videos cannot fully resolve.

The most fundamental limitation of external videos is archival persistence: external links decay over time as project websites go offline, cloud storage expires, and institutional servers are decommissioned. Additional limitations include **reviewer friction**: context-switching between PDF and browser tabs disrupts narrative flow and increases cognitive load, potentially leading reviewers to skip supplementary videos entirely. Embedded animations keep temporal content within the document flow, reducing this friction. Third, **blind review integrity**: even with anonymized hosting platforms, external links introduce potential identity leakage vectors (e.g., DNS lookups, server logs, timing analysis). Embedded content eliminates these vectors entirely.

We view embedded animations and external supplements as **complementary rather than mutually exclusive**. Authors can provide both: embedded animations serve as the archival baseline that persists indefinitely, while external videos offer enhanced quality and universal compatibility for reviewers who prefer them. The DFIG v1.0 specification deliberately constrains quality (720p, 30 fps max) to balance expressiveness with file size, accepting that external videos can achieve higher fidelity. Our proposal is not to replace external supplements, but to add embedded animations as a persistent, friction-reducing option that addresses archival and review workflow concerns that external videos alone cannot solve.

7. Path to Adoption

Rather than advocating for immediate mandatory adoption, we propose a gradual path that allows the community to evaluate embedded dynamic figures in practice, gather empirical evidence, and iteratively refine the specification.

7.1. Proposed Pilot Program

We suggest that a major ML venue (e.g., ICML, NeurIPS, CVPR) consider a pilot program with the following structure:

Optional Track. For one submission cycle, authors submitting to relevant tracks (video generation, 3D/4D, world models, robotics) are *invited but not required* to include DFIG-compliant dynamic figures. Clear guidelines and templates are provided.

Reviewer Preparation. Reviewers assigned to papers with dynamic figures receive instructions to install Adobe Acrobat Reader (free) or use Foxit Reader, a test PDF to verify their viewer supports animations, and assurance that static fallbacks are available if they cannot view animations.

Post-Review Survey. After the review cycle, both authors and reviewers complete a survey measuring perceived utility of dynamic figures for evaluation, time spent on submissions with vs. without dynamic figures, technical difficulties encountered, and suggestions for improving the DFIG specification.

Publication of Results. Survey results and lessons learned are published, informing whether and how to expand DFIG adoption in future cycles.

7.2. OpenReview Integration

OpenReview, the primary review platform for many ML venues, currently displays PDFs using PDF.js, which does not support `animate` playback. We propose the following integration path:

Short-term: Download Prompt. When a PDF contains dynamic figures (detectable via metadata or JavaScript presence), OpenReview displays a notice: "This PDF contains animated figures. For full viewing, download and open in Adobe Acrobat Reader." A download button is prominently displayed.

Medium-term: Sidecar Preview. Authors upload both the PDF and the sidecar artifact. OpenReview renders GIF/WebP previews from the sidecar alongside the PDF, allowing browser-based preview of animations without requiring specialized software.

Long-term: Native Support. PDF.js could potentially add support for `animate`-style frame sequences, which are simpler than full video playback. This would require collaboration with the PDF.js open-source community.

7.3. Digital Library Coordination

For archival robustness, we recommend coordination with digital libraries:

ACM Digital Library. The ACM already supports supplementary materials. Dynamic figure PDFs could be archived alongside sidecar frame sequences, with appropriate metadata indicating the relationship.

arXiv. arXiv currently processes PDFs but does not explicitly support dynamic content. The sidecar artifact approach (frames + manifest) provides a fallback that arXiv can archive without modification.

Venue-Specific Archives. Venues can host complete submission packages (PDF + sidecar) on their own infrastructure, ensuring long-term availability independent of external platforms.

7.4. Community Feedback Mechanism

To enable iterative improvement of the DFIG specification, we propose an open specification repository on GitHub hosting the DFIG specification, LaTeX templates, and example code, where the community can submit issues for compatibility problems, feature requests, and clarifications. The specification uses semantic versioning (v1.0, v1.1, v2.0, etc.), where minor versions address clarifications and compatibility updates, and major versions indicate significant changes to constraints or requirements. Aligned with major venue deadlines, the specification is reviewed annually, with community feedback, pilot program results, and technological changes (e.g., improved PDF.js support) informing updates.

7.5. Template and Tooling Support

To lower the barrier to adoption, we commit to providing **LaTeX templates** (ready-to-use templates with DFIG-compliant figure environments and accessibility annotations), **frame extraction scripts** (Python/bash scripts for converting videos to optimized frame sequences with appropriate compression), **validation tools** (scripts to check PDF file size, frame dimensions, and metadata compliance with DFIG v1.0), and **tutorial documentation** (step-by-step guides for authors unfamiliar with the `animate` package). These resources will be made available in the specification repository, with community contributions welcome.

8. Conclusion

As the machine learning community advances increasingly dynamic frontiers—video generation, 3D/4D synthesis, world modeling, and embodied AI—our modes of scientific communication face growing tension with the static PDF format. This position paper has argued that embedded dynamic figures offer a promising path forward, enabling temporal results to be evaluated directly within the archival document rather than through fragmented external supplements.

A. Resources

Table 3 summarizes which embedding approaches are recommended under DFIG v1.0 and which are deprecated due to compatibility or security concerns.

B. LLM Usage Disclaim

We used large language models (LLMs) to polish English grammar and improve readability throughout this manuscript. All scientific content, experimental design, and technical writing—including all claims, figure captions, and appendix sections—were authored and fact-checked by the listed authors. No text, data, or code was generated by LLMs without human review and, where relevant, original authorial input.

C. Additional Case Studies

The following case studies demonstrate DFIG-compliant dynamic figures across different domains. These supplement the main text examples and provide additional evidence of technical feasibility.

Case Study: Text-to-Video Generation. Figure 6 shows a video generation example from CogVideoX (Yang et al., 2025), demonstrating the ability to evaluate temporal consistency across generated video frames.

Case Study: Diffusion Process Illustration. Figure 7 illustrates the reverse SDE solving process from score-based generative modeling (Yang Song, 2021; Song et al., 2021), demonstrating pedagogical use of dynamic figures.

Case Study: High-Fidelity 3D Asset Generation. Figure 8 shows a turntable visualization of 3D assets generated by Hunyuan3D 2.5 (Lai et al., 2025).

Case Study: Robotics. Figure 9 demonstrates robotics visualization with a quadruped robot performing agile maneuvers (Pan et al., 2025).

Case Study: Autonomous Driving Simulation. Figure 10 shows synthetic driving data generation from Cosmos-Drive-Dreams (Ren et al., 2025).

Case Study: Framework Illustration. Figure 11 demonstrates dynamic visualization of a reasoning framework from Multiplex Thinking (Tang et al., 2026).

Table 3. Recommended and deprecated approaches for embedding dynamic content in PDFs.

Approach	Status	Notes
Frame sequences (`animate`)	**Recommended**	Best compatibility, minimal JavaScript, graceful fallback to static frames.
3D PRC files (`media9`)	Deprecated	Requires Adobe Acrobat only; inconsistent support in other viewers.
3D U3D files (`media9`)	Deprecated	Limited viewer support; better to render turntable animations as frame sequences.
MP4 video (`media9`)	Deprecated	Relies on Flash-based players deprecated since 2020; non-functional in modern viewers.
MP3 audio (`media9`)	Deprecated	Same Flash dependency; conflicts with `animate` package when used together.
Animated GIF (direct)	Not supported	LaTeX compilers do not natively support animated GIFs; extract frames instead.

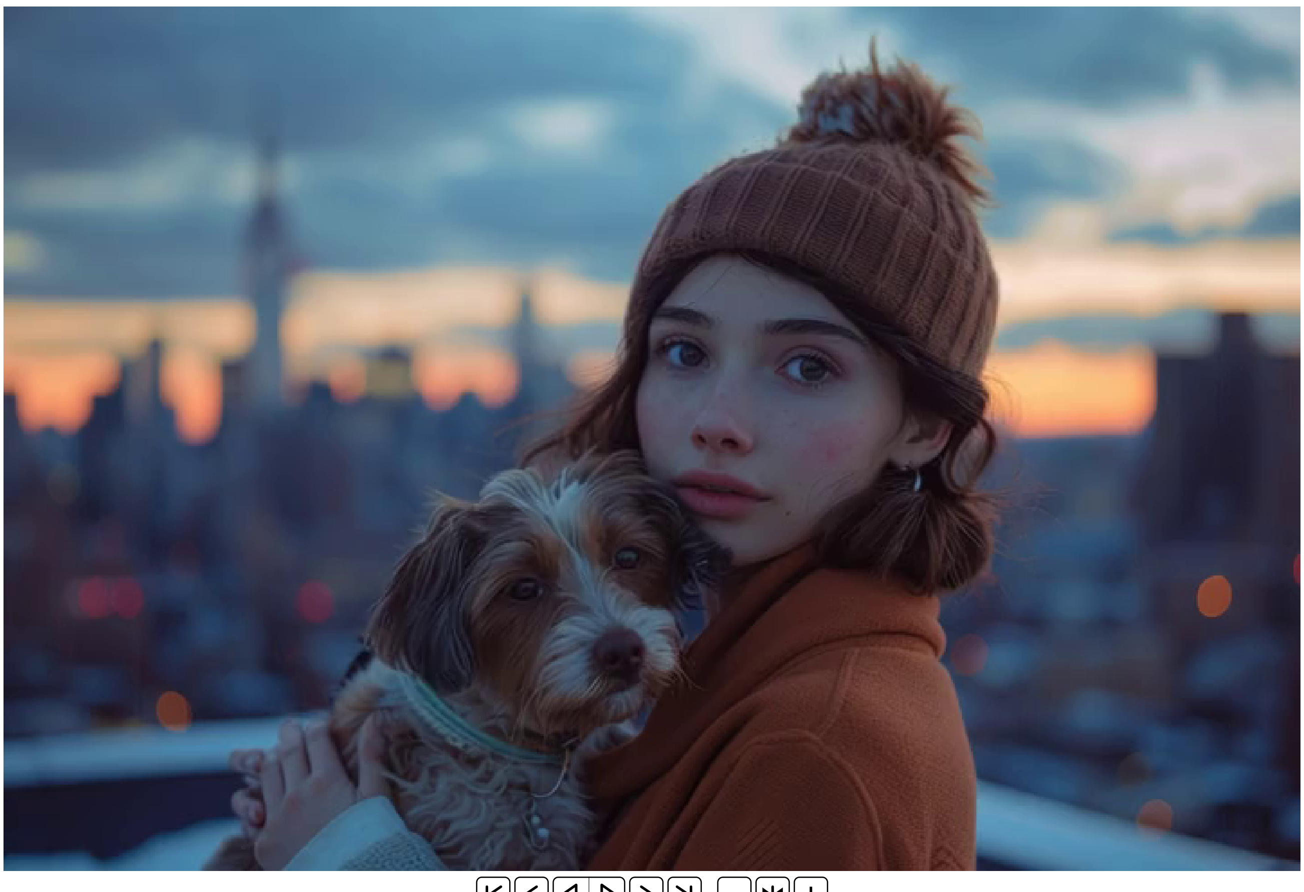

Figure 6. Text-to-video generation example from CogVideoX (Yang et al., 2025). **Accessibility:** The animation shows a generated video sequence demonstrating temporal coherence in synthesized content. **Viewing:** Use Adobe Acrobat Reader or Foxit Reader for animation playback.

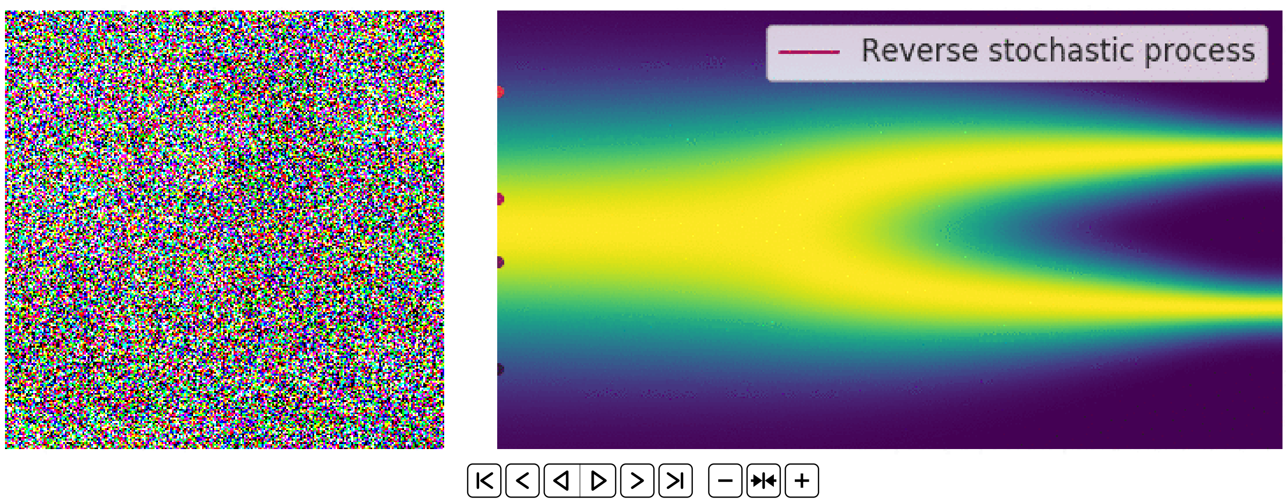

Figure 7. Reverse SDE solving process for score-based generative modeling (Yang Song, 2021). **Accessibility:** The animation shows noise being progressively removed from random samples to generate structured data. **Viewing:** Use Adobe Acrobat Reader or Foxit Reader for animation playback.

Figure 8. High-fidelity 3D asset generation turntable from Hunyuan3D 2.5 (Lai et al., 2025). **Accessibility:** The animation shows a 360-degree rotation around a generated 3D model, revealing geometric detail and texture quality from all angles. **Viewing:** Use Adobe Acrobat Reader or Foxit Reader for animation playback.

Figure 9. Quadruped robot performing agile basketball maneuvers (Pan et al., 2025). **Accessibility:** The animation shows a four-legged robot demonstrating dynamic locomotion and ball manipulation. **Viewing:** Use Adobe Acrobat Reader or Foxit Reader for animation playback.

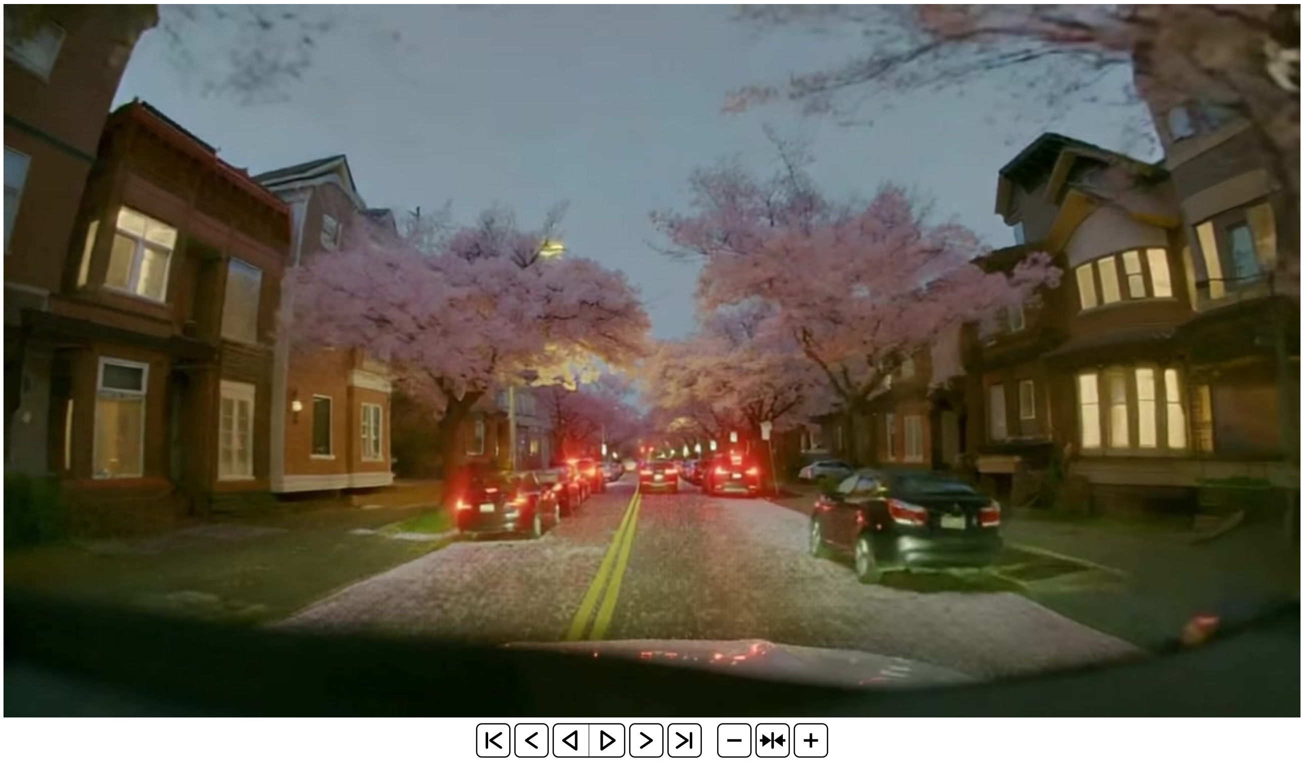

Figure 10. Synthetic driving data generation from Cosmos-Drive-Dreams (Ren et al., 2025). **Accessibility:** The animation shows a simulated driving scenario with realistic scene generation and temporal consistency. **Viewing:** Use Adobe Acrobat Reader or Foxit Reader for animation playback.

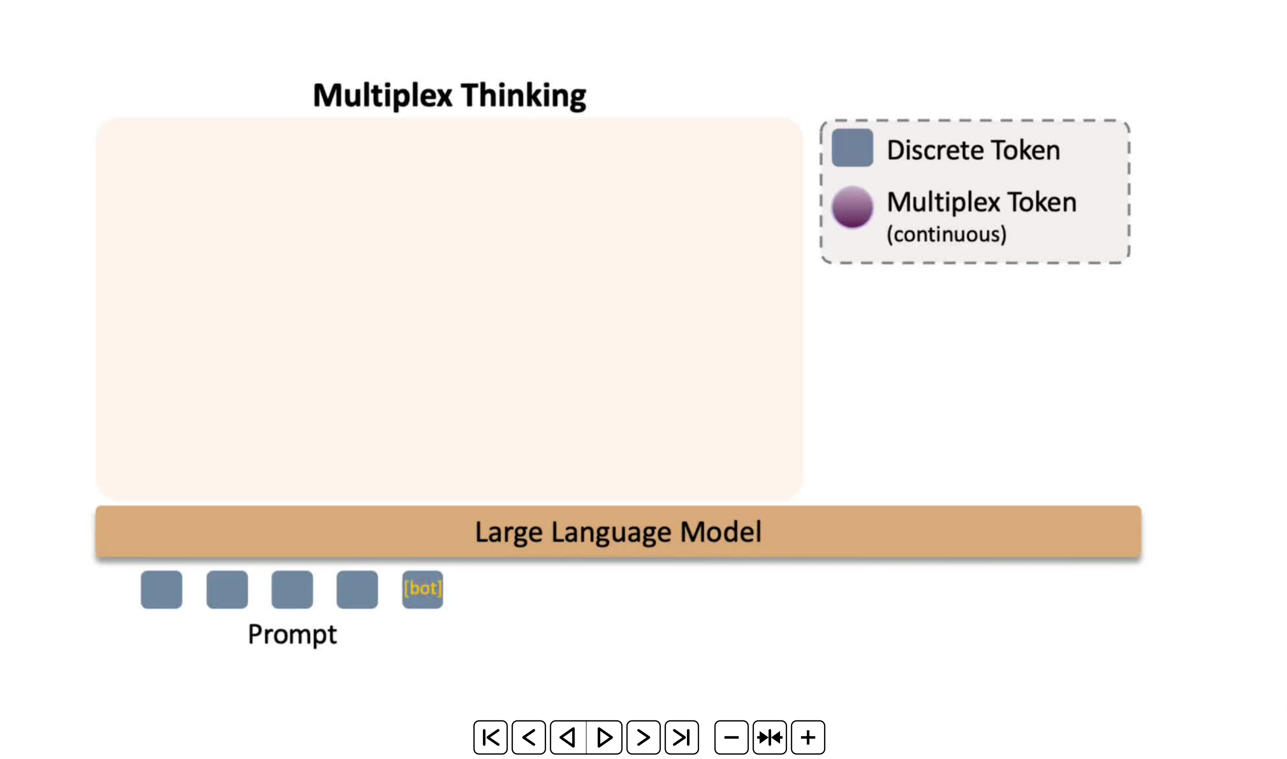

Figure 11. Multiplex Thinking reasoning framework visualization (Tang et al., 2026). **Accessibility:** The animation illustrates the token-wise branch-and-merge reasoning process, showing how multiple reasoning paths are explored and combined. **Viewing:** Use Adobe Acrobat Reader or Foxit Reader for animation playback.

References

Abrams, S. L. and Levenson, S. P. Pdf/a: An electronic document file format for long-term preservation. In *Archiving Conference*, volume 1, pp. 237–241. Society of Imaging Science and Technology, 2004. Archival format specification prohibiting JavaScript and multimedia.

Bai, S., Song, W., Chen, J., Ji, Y., Zhong, Z., Yang, J., Zhao, H., Zhou, W., Zhao, W., Li, Z., Ding, P., Chi, C., Li, H., Xu, C., Zheng, X., Wang, D., Zhang, S., and Chen, B. Towards a Unified Understanding of Robot Manipulation: A Comprehensive Survey, October 2025.

Barnes, D. G. and Fluke, C. J. Incorporating interactive three-dimensional graphics in astronomy research papers. *New Astronomy*, 13(8):599–605, November 2008. ISSN 1384-1076. doi: 10.1016/j.newast.2008.03.008.

Cui, W., Yu, D., Jiao, X., Meng, Z., Zhang, G., Wang, Q., Guo, S. Y., and King, I. Recent Advances in Speech Language Models: A Survey. In Che, W., Nabende, J., Shutova, E., and Pilehvar, M. T. (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 13943–13970, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.682.

Deng, Y., Pan, Z., Zhang, H., Li, X., Hu, R., Ding, Y., Zou, Y., Zeng, Y., and Zhou, D. Rethinking Video Generation Model for the Embodied World, January 2026.

Ding, J., Zhang, Y., Shang, Y., Zhang, Y., Zong, Z., Feng, J., Yuan, Y., Su, H., Li, N., Sukiennik, N., Xu, F., and Li, Y. Understanding World or Predicting Future? A Comprehensive Survey of World Models. *ACM Comput. Surv.*, 58(3):57:1–57:38, September 2025. ISSN 0360-0300. doi: 10.1145/3746449.

Fan, C.-K., Chi, X., Ju, X., Li, H., Bao, Y., Wang, Y.-K., Chen, L., Jiang, Z., Ge, K., Li, Y., et al. Wow, wo, val! a comprehensive embodied world model evaluation turing test. *arXiv preprint arXiv:2601.04137*, 2026.

Feng, T., Wang, W., and Yang, Y. A Survey of World Models for Autonomous Driving, September 2025.

Foo, L. G., Rahmani, H., and Liu, J. AI-Generated Content (AIGC) for Various Data Modalities: A Survey. *ACM Comput. Surv.*, 57(9):243:1–243:66, May 2025. ISSN 0360-0300. doi: 10.1145/3728633.

Guan, Y., Liao, H., Li, Z., Hu, J., Yuan, R., Zhang, G., and Xu, C. World Models for Autonomous Driving: An Initial Survey. *IEEE Transactions on Intelligent Vehicles*, pp. 1–17, 2024. ISSN 2379-8904. doi: 10.1109/TIV.2024.3398357.

He, C., Shen, Y., Fang, C., Xiao, F., Tang, L., Zhang, Y., Zuo, W., Guo, Z., and Li, X. Diffusion Models in Low-Level Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4630–4651, June 2025. ISSN 1939-3539. doi: 10.1109/TPAMI.2025.3545047.

Hyung, J., Kim, K., Hong, S., Kim, M.-J., and Choo, J. Spatiotemporal Skip Guidance for Enhanced Video Diffusion Sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11006–11015, 2025.

Kim, M.-J., Kim, D., Yun, S., and Choo, J. TV-LiVE: Training-Free, Text-Guided Video Editing via Layer Informed Vitality Exploitation, June 2025.

Kong, L., Yang, W., Mei, J., Liu, Y., Liang, A., Zhu, D., Lu, D., Yin, W., Hu, X., Jia, M., Deng, J., Zhang, K., Wu, Y., Yan, T., Gao, S., Wang, S., Li, L., Pan, L., Liu, Y., Zhu, J., Ooi, W. T., Hoi, S. C. H., and Liu, Z. 3D and 4D World Modeling: A Survey, September 2025.

Lai, Z., Zhao, Y., Liu, H., Zhao, Z., Lin, Q., Shi, H., Yang, X., Yang, M., Yang, S., Feng, Y., Zhang, S., Huang, X., Luo, D., Yang, F., Yang, F., Wang, L., Liu, S., Tang, Y., Cai, Y., He, Z., Liu, T., Liu, Y., Jiang, J., Linus, Huang, J., and Guo, C. Hunyuan3D 2.5: Towards High-Fidelity 3D Assets Generation with Ultimate Details, June 2025.

Lee, S., Hoover, B., Strobelt, H., Wang, Z. J., Peng, S., Wright, A., Li, K., Park, H., Yang, H., and Chau, D. H. P. Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion. In *2024 IEEE Visualization and Visual Analytics (VIS)*, pp. 96–100, October 2024. doi: 10.1109/VIS55277.2024.00027.

Lin, M., Wang, X., Wang, Y., Wang, S., Dai, F., Ding, P., Wang, C., Zuo, Z., Sang, N., Huang, S., and Wang, D. Exploring the Evolution of Physics Cognition in Video Generation: A Survey, March 2025.

Liu, Y., Chen, W., Bai, Y., Liang, X., Li, G., Gao, W., and Lin, L. Aligning Cyber Space With Physical World: A Comprehensive Survey on Embodied AI. *IEEE/ASME Transactions on Mechatronics*, 30(6):7253–7274, December 2025. ISSN 1941-014X. doi: 10.1109/TMECH.2025.3574943.

Long, X., Zhao, Q., Zhang, K., Zhang, Z., Wang, D., Liu, Y., Shu, Z., Lu, Y., Wang, S., Wei, X., Li, W., Yin, W., Yao, Y., Pan, J., Shen, Q., Yang, R., Cao, X., and Dai, Q. A Survey: Learning Embodied Intelligence from Physical Simulators and World Models, September 2025.

Ma, Z., Zhang, Y., Jia, G., Zhao, L., Ma, Y., Ma, M., Liu, G., Zhang, K., Ding, N., Li, J., and Zhou, B. Efficient Diffusion Models: A Comprehensive Survey From Principles to Practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(9):7506–7525, September 2025. ISSN 1939-3539. doi: 10.1109/TPAMI.2025.3569700.

Miao, Q., Li, K., Quan, J., Min, Z., Ma, S., Xu, Y., Yang, Y., Liu, P., and Luo, Y. Advances in 4D Generation: A Survey, July 2025.

Newe, A. Enriching scientific publications with interactive 3D PDF: An integrated toolbox for creating ready-to-publish figures. *PeerJ Computer Science*, 2:e64, June 2016. ISSN 2376-5992. doi: 10.7717/peerj-cs.64.

Pan, Y., Qiao, R., Chen, L., Chitta, K., Pan, L., Mai, H., Bu, Q., Zhao, H., Zheng, C., Luo, P., and Li, H. Agility Meets Stability: Versatile Humanoid Control with Heterogeneous Data, November 2025.

Ren, X., Lu, Y., Cao, T., Gao, R., Huang, S., Sabour, A., Shen, T., Pfaff, T., Wu, J. Z., Chen, R., Kim, S. W., Gao, J., Leal-Taixe, L., Chen, M., Fidler, S., and Ling, H. Cosmos-Drive-Dreams: Scalable Synthetic Driving Data Generation with World Foundation Models, June 2025.

Shao, R., Li, W., Zhang, L., Zhang, R., Liu, Z., Chen, R., and Nie, L. Large VLM-based Vision-Language-Action Models for Robotic Manipulation: A Survey, September 2025.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021.

Tang, Y., Dong, L., Hao, Y., Dong, Q., Wei, F., and Gu, J. Multiplex Thinking: Reasoning via Token-wise Branch-and-Merge, January 2026.

Vilaça, L., Yu, Y., and Viana, P. A Survey of Recent Advances and Challenges in Deep Audio-Visual Correlation Learning. *ACM Comput. Surv.*, 57(12):299:1–299:46, July 2025. ISSN 0360-0300. doi: 10.1145/3696445.

Wang, C., Peng, H.-Y., Liu, Y.-T., Gu, J., and Hu, S.-M. Diffusion Models for 3D Generation: A Survey. *Computational Visual Media*, 11(1):1–28, February 2025a. ISSN 2096-0662. doi: 10.26599/CVM.2025.9450452.

Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., and Novotny, D. VGGT: Visual Geometry Grounded Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5294–5306, 2025b.

Wang, Y., Liu, X., Pang, W., Ma, L., Yuan, S., Debevec, P., and Yu, N. Survey of Video Diffusion Models: Foundations, Implementations, and Applications. *Transactions on Machine Learning Research*, April 2025c. ISSN 2835-8856.

Wang, Y., Xing, S., Can, C., Li, R., Hua, H., Tian, K., Mo, Z., Gao, X., Wu, K., Zhou, S., You, H., Peng, J., Zhang, J., Wang, Z., Song, R., Yan, M., Zimmer, W., Zhou, X., Li, P., Lu, Z., Chen, C.-J., Huang, Y., Rossi, R. A., Sun, L., Yu, H., Fan, Z., Yang, F. H., Kang, Y., Greer, R., Liu, C., Lee, E. H., Di, X., Ye, X., Ren, L., Knoll, A., Li, X., Ji, S., Tomizuka, M., Pavone, M., Yang, T., Du, J., Yang, M.-H., Wei, H., Wang, Z., Zhou, Y., Li, J., and Tu, Z. Generative AI for Autonomous Driving: Frontiers and Opportunities, May 2025d.

Waseem, F. and Shahzad, M. Video is Worth a Thousand Images: Exploring the Latest Trends in Long Video Generation. *ACM Comput. Surv.*, 58(6):154:1–154:35, December 2025. ISSN 0360-0300. doi: 10.1145/3771724.

Wen, B., Xie, H., Chen, Z., Hong, F., and Liu, Z. 3D Scene Generation: A Survey, May 2025.

Xiang, T.-Y., Jin, A.-Q., Zhou, X.-H., Gui, M.-J., Xie, X.-L., Liu, S.-Q., Wang, S.-Y., Duan, S.-B., Xie, F.-C., Wang, W.-K., Wang, S.-C., Li, L.-Y., Tu, T., and Hou, Z.-G. Parallels Between VLA Model Post-Training and Human Motor Learning: Progress, Challenges, and Trends, June 2025.

Xie, N., Tian, Z., Yang, L., Zhang, X.-P., Guo, M., and Li, J. From 2D to 3D Cognition: A Brief Survey of General World Models, June 2025a.

Xie, T., Rong, Y., Zhang, P., Wang, W., and Liu, L. Towards Controllable Speech Synthesis in the Era of Large Language Models: A Systematic Survey. In Christodoulopoulos, C., Chakraborty, T., Rose, C., and Peng, V. (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 764–791, Suzhou, China, November 2025b. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.40.

Xu, H., Chen, J., Meng, S., Wang, Y., and Chau, L.-P. A survey on occupancy perception for autonomous driving: The information fusion perspective. *Information Fusion*, 114:102671, February 2025. ISSN 15662535. doi: 10.1016/j.inffus.2024.102671.

Xue, H., Luo, X., Hu, Z., Zhang, X., Xiang, X., Dai, Y., Liu, J., Zhang, Z., Li, M., Yang, J., Ma, F., Wu, Z., Yang, C., Dai, Z., and Yu, F. R. Human Motion Video Generation: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11):10709–10730, November 2025. ISSN 1939-3539. doi: 10.1109/TPAMI.2025.3594034.

Yang, Z., Teng, J., Zheng, W., Ding, M., Huang, S., Xu, J., Yang, Y., Hong, W., Zhang, X., Feng, G., Yin, D., Yuxuan.Zhang, Wang, W., Cheng, Y., Xu, B., Gu, X., Dong, Y., and Tang, J. CogVideoX: Text-to-Video Diffusion Models with An Expert Transformer. In *The Thirteenth International Conference on Learning Representations*, 2025.

Yang Song. Generative Modeling by Estimating Gradients of the Data Distribution | Yang Song. https://yang-song.net/blog/2021/score/, 2021.

Yu, Z., Wang, B., Zeng, P., Zhang, H., Zhang, J., Gao, L., Song, J., Sebe, N., and Shen, H. T. A Survey on Efficient Vision-Language-Action Models, October 2025.

Yun, J. and Choo, J. Vector Prism: Animating Vector Graphics by Stratifying Semantic Structure, December 2025.

Zhang, D., Sun, J., Hu, C., Wu, X., Yuan, Z., Zhou, R., Shen, F., and Zhou, Q. Pure Vision Language Action (VLA) Models: A Comprehensive Survey, November 2025a.

Zhang, Y., Xing, J., Xia, B., Liu, S., Peng, B., Tao, X., Wan, P., Lo, E., and Jia, J. Training-Free Efficient Video Generation via Dynamic Token Carving. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, October 2025b.

Zhao, C., Liu, G., Zhang, R., Liu, Y., Wang, J., Kang, J., Niyato, D., Li, Z., Xuemin, Shen, Han, Z., Sun, S., Yuen, C., and Kim, D. I. Edge General Intelligence Through World Models and Agentic AI: Fundamentals, Solutions, and Challenges, August 2025a.

Zhao, C., Zhang, R., Wang, J., Zhao, G., Niyato, D., Sun, G., Mao, S., and Kim, D. I. World Models for Cognitive Agents: Transforming Edge Intelligence in Future Networks, May 2025b.

Zheng, Y., Yao, L., Su, Y., Zhang, Y., Wang, Y., Zhao, S., Zhang, Y., and Chau, L.-P. A Survey of Embodied Learning for Object-centric Robotic Manipulation. *Machine Intelligence Research*, 22(4):588–626, August 2025. ISSN 2731-5398. doi: 10.1007/s11633-025-1542-8.

Zhong, Y., Bai, F., Cai, S., Huang, X., Chen, Z., Zhang, X., Wang, Y., Guo, S., Guan, T., Lui, K. N., Qi, Z., Liang, Y., Chen, Y., and Yang, Y. A Survey on Vision-Language-Action Models: An Action Tokenization Perspective, July 2025.

Zhu, Z., Wang, X., Zhao, W., Min, C., Li, B., Deng, N., Dou, M., Wang, Y., Shi, B., Wang, K., Zhang, C., You, Y., Zhang, Z., Zhao, D., Xiao, L., Zhao, J., Lu, J., and Huang, G. Is Sora a World Simulator? A Comprehensive Survey on General World Models and Beyond, October 2025.