A Survey on Video Generation Technologies, Applications, and Ethical Considerations

Kaiqi Chen *

Posted Date: 19 November 2025

doi: 10.20944/preprints202511.1332.v1

Keywords: video generation; generative AI; diffusion models; autoregressive models; GANs; Interactive Generative Video; multimodal learning; ethics

 doi:10.20944/preprints202511.1332.v1

Article

A Survey on Video Generation Technologies, Applications, and Ethical Considerations

Kaiqi Chen

Independent Researcher, USA; kaiqichen115@gmail.com

Abstract

Video generation has rapidly advanced from early GAN-based systems to modern diffusion- and transformer-based models that deliver unprecedented photorealism and controllability. This survey synthesizes progress across foundational models (GAN, autoregressive, diffusion, masked modeling, and hybrids), information representations (spatiotemporal convolution, patch tokens, latent spaces), and generation schemes (decoupled, hierarchical, multi-staged, latent). We map applications in gaming, embodied AI, autonomous driving, education, filmmaking, and biomedicine, and analyze technical challenges in real-time generation, long-horizon consistency, physics fidelity, generalization, and multimodal reasoning. We also discuss governance and ethics, including misinformation, intellectual property, fairness, privacy, accountability, and environmental impact. Finally, we summarize evaluation methodologies (spatial, temporal, and human-centered metrics) and highlight future directions for efficient, controllable, and trustworthy video generation.

Keywords: video generation; generative AI; diffusion models; autoregressive models; GANs; Interactive Generative Video; multimodal learning; ethics

1. Introduction

Demand for interactive, high-fidelity video spans simulation, content creation, decision support, and education. Advances in adversarial learning [1], autoregressive transformers [2–4], and diffusion models [5–7] have enabled controllable and diverse video synthesis. We use *Interactive Generative Video (IGV)* to denote systems that couple video generation with user control signals (text, image, motion, audio, or programmatic constraints), enabling closed-loop interaction.

Scope and Contributions. We (i) unify model families and their training/decoding trade-offs; (ii) categorize information representations and control interfaces; (iii) review commercial/research systems; (iv) outline evaluation practices; and (v) surface open problems and governance directions. Figure 1 provides a top-level taxonomy.

 doi:10.20944/preprints202511.1332.v1

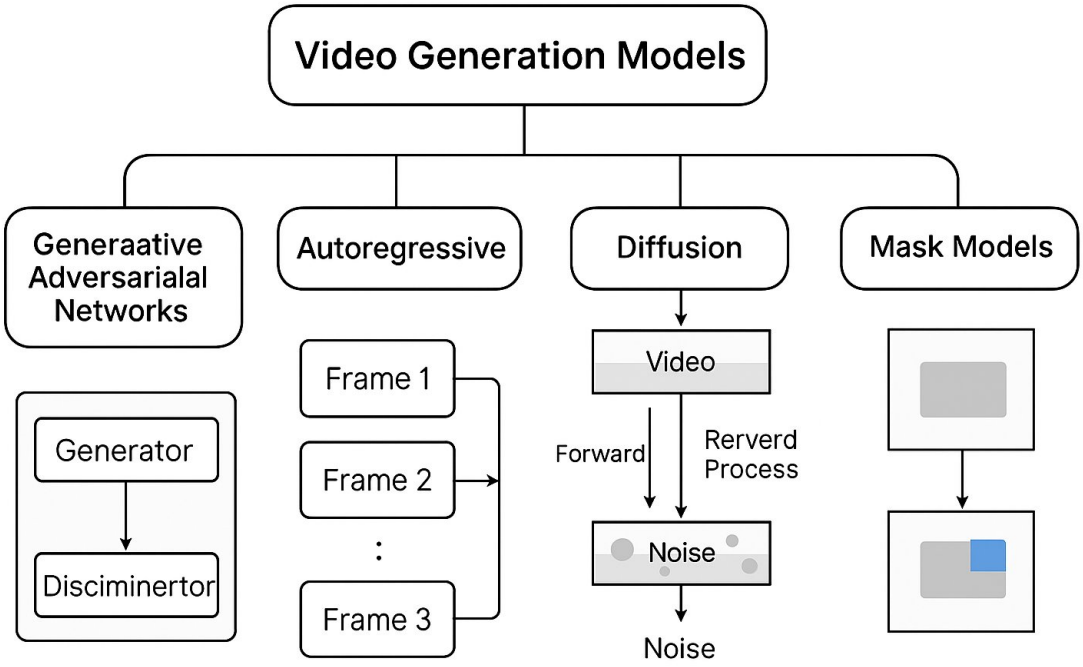

Figure 1. Taxonomy of video generation model families.

2. Foundational Models

We group methods into GAN, autoregressive (AR), diffusion, masked/MAE-style, and hybrids. Below we emphasize objectives, architectures, and practical tips.

2.1. GAN-Based Approaches

Objectives. Minimax adversarial training with hinge loss is common; feature matching, perceptual losses, and temporal coherence losses (e.g., optical-flow smoothness) stabilize training. **Architectures.** Temporal discriminators, multi-scale critics (frame/few-shot sequence), 3D spatiotemporal convolutions, and motion/content disentanglement (MoCoGAN [8]). Style-based generators extend StyleGAN to video (StyleGAN-V). **Pros/Cons.** Pros: sharp frames, low-latency inference. Cons: training instability, mode collapse, and difficulty with long-horizon dynamics. **Mitigations.** R1/R2 regularization, spectral normalization, curriculum schedules, and two-timescale updates. **Editing.** Latent space arithmetic and GAN inversion enable post-hoc control.

2.2. Autoregressive (AR) Models

Tokenization. VQ-VAE/yu2023magvit [3,4] discretize videos into spatiotemporal tokens. **Backbones.** Causal transformers predict tokens with long-context attention (sliding windows, chunk-wise decoding, rotary/ALiBi temporal encodings). **Training.** Cross-entropy with teacher forcing; scheduled sampling reduces exposure bias. **Decoding.** Nucleus sampling, classifier-free guidance, speculative decoding, and lookahead caches speed generation. **Strengths.** Exact likelihood training, modular conditioning (text, audio). **Limits.** Serial decoding; error accumulation in long videos. **Notable systems.** VideoGPT [2], VideoPoet.

2.3. Diffusion Models

Backbones. U-Nets with 2D+time or full 3D attention; DiT-style transformer backbones are increasingly common [9]. **Latent Diffusion.** Compress frames with VAEs for efficiency [6]. **Conditioning.** Text (T2V), image (I2V), pose/depth/optical flow (ControlNet-style [10]), audio beat alignment,

 doi:10.20944/preprints202511.1332.v1

camera trajectories. **Sampling.** DPM-Solver, consistency models, and few-step distillation (progressive or adversarial) reduce steps. **Cascades.** Imagen Video [7] uses cascaded diffusion for resolution/fps upscaling. **Limits.** Expensive sampling, identity drift, camera jitter. **Mitigations.** Identity anchors, keyframe-guided attention, temporal token locking, and motion priors [11].

2.4. Masked/MAE-Style Models

Masked autoencoding over spatiotemporal patches improves representations and reduces compute. Generative variants decode masked tokens directly (BERT-like) or combine with diffusion for fidelity. Temporal masking schedules and cross-frame reconstruction improve consistency.

2.5. Hybrid and Emerging Paradigms

AR+Diffusion Hybrids. AR drafts long-range content; diffusion refines quality. **World Models.** Latent dynamics (RSSM variants) produce controllable rollouts. **3D-aware Video.** NeRF/GS priors for multi-view consistency [12–14]. **Streaming/Online.**[15] Non-autoregressive decoders with keyframe memory enable low-latency streaming. **Agents.** LLM planners choose controls (pose/camera/story beats) for goal-driven video [16].

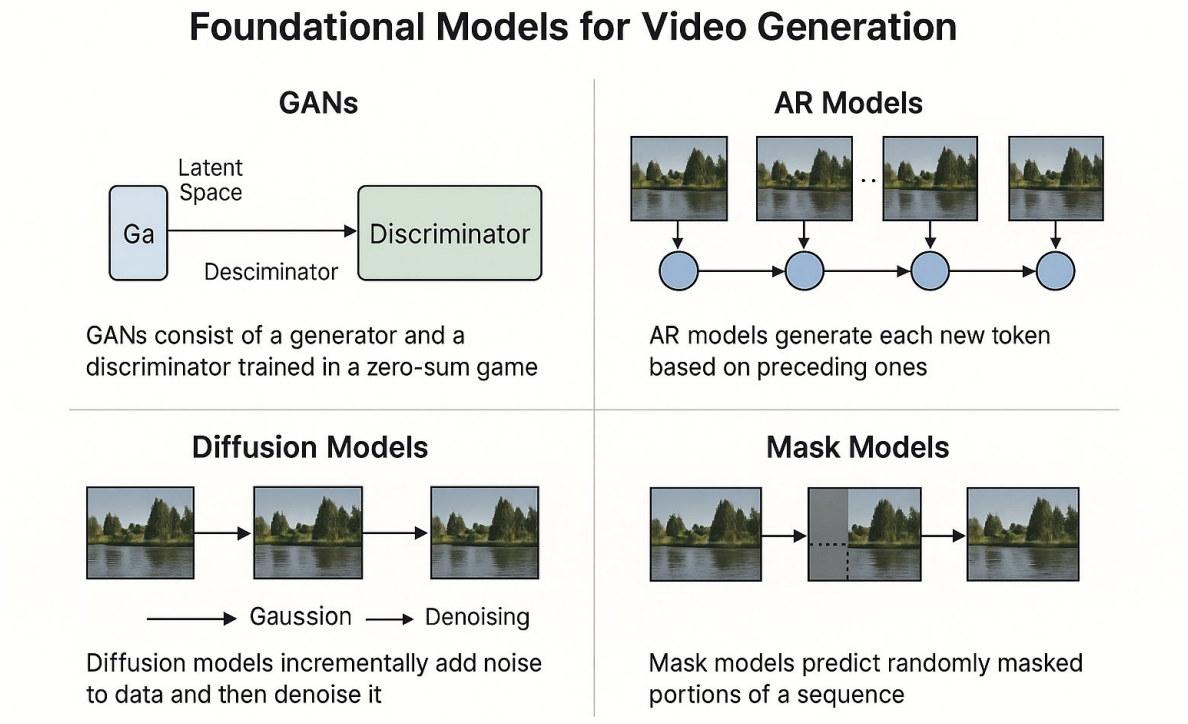

Figure 2. Foundational model families with schematic examples.

3. Information Representations

3.1. Spatiotemporal Convolution and Patches

3D convolutions couple space and time; factorized (2+1)D reduces compute. Patch tokenization feeds video transformers with temporal encodings [17]. TokenLearner/Token Merging adaptively reduces tokens while preserving semantics.

3.2. Latent Spaces (VAE/VQ-VAE)

Encoders compress frames; VQ discretizes latents for AR modeling (yu2023magvit [3,4]). Hierarchical latents (coarse → fine) align with multi-staged decoders.

 doi:10.20944/preprints202511.1332.v1

3.3. Multimodal Encoders

Vision + text (CLIP [18]), vision + audio (AST), and motion encoders (pose/flow) align heterogeneous inputs. Learned camera embeddings encode trajectories for cinematography control.

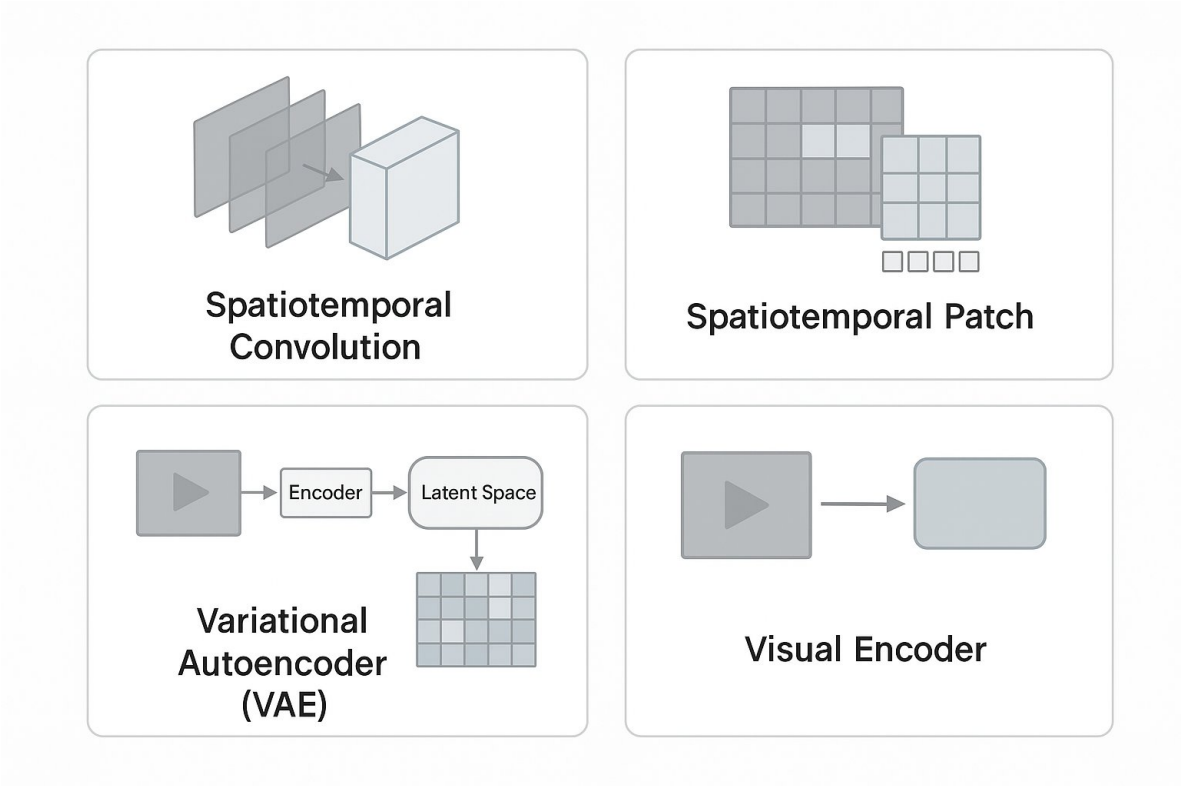

Figure 3. Information representations for video generation.

4. Generation Schemes and Control

4.1. Decoupled and Hierarchical

Decouple motion vs. content; or use keyframes/storyboards followed by inpainting, motion interpolation, and temporal super-resolution [19,20].

4.2. Multi-Staged and Cascaded

Draft low-res/low-fps videos then upsample/densify; cascades split difficulty across scales with noise schedules matched to each stage [21].

4.3. Latent Model Scheme with Control

Encode → latent backbone (GAN/AR/diffusion) → decode; control via pose/depth/camera/audio. Cross-modal attention and FiLM-style conditioning inject guidance signals.

 doi:10.20944/preprints202511.1332.v1

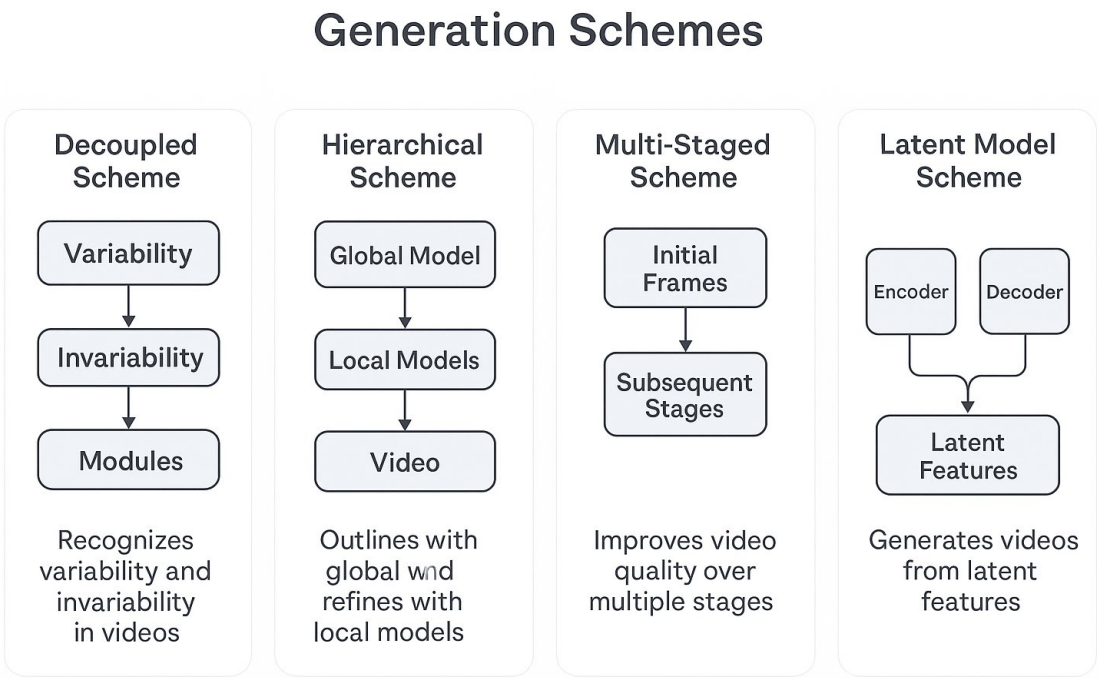

Figure 4. Four common generation schemes used in modern systems.

4.4. Practical Control Interfaces

- **Geometry:** depth, normal, flow, keypoints, 2D/3D pose [22].
- **Cinematography:** camera path, focal length, shutter, DoF.
- **Audio-driven:** lip-sync, beat tracking, motion from music.
- **Structure:** layout maps, segmentation, scene graphs.
- **Programmatic:** constraints in code or node graphs (e.g., ControlNet-like blocks [10]).

5. Industry Landscape and Systems

The commercial and research ecosystem for video generation is evolving rapidly, with diverse approaches in backbone architecture, tokenization strategy, conditioning modalities, and deployment environments [23,24]. In this section, we group systems into **commercial**, **research**, and **hybrid/industry-research collaborations**.

5.1. Commercial Systems

- **OpenAI Sora** [25]: Latent diffusion with DiT-style transformer backbones. Excels in long-duration text-to-video (up to minutes) with cinematic camera controls and physics-consistent motion. Uses cascaded super-resolution and classifier-free guidance.
- **Runway Gen-3** [26]: Latent diffusion with multi-modal control (text, image, pose). Designed for professional creatives, supports inpainting and frame interpolation for seamless editing.
- **Pika** [27]: Web-based T2V and I2V platform optimized for fast iteration. Prioritizes accessibility and speed, with resolution up to 1080p.
- **Stability Video**: Part of the Stability AI ecosystem, leveraging Stable Video Diffusion backbones with open-source checkpoints for community use.
- **Google Veo** [28]: High-fidelity text-to-video with camera trajectory conditioning and temporal super-resolution. Integrates with Google Workspace for productivity applications.
- **Tencent Hunyuan Video** [**?**]: Chinese-market oriented video generation system with strong support for pose guidance and cultural style templates.

 doi:10.20944/preprints202511.1332.v1

- **Adobe Firefly Video**: Video generation integrated into Creative Cloud, emphasizing IP-safe training data and native integration with Premiere Pro/After Effects.
- **ByteDance MagicVideo**: Targets short-form social content with emphasis on style transfer and lip-sync from audio.
- **NVIDIA ACE (Avatar Cloud Engine)**: Real-time digital human platform combining speech-to-video animation, facial reenactment, and generative backgrounds.

5.2. Research Systems

- **Make-A-Video** (Meta): Early diffusion-based T2V model combining image pretraining with temporal layers.
- **Phenaki** (Google): Transformer-based long video generation via token streams and mask prediction.
- **Imagen Video** [7] (Google): High-definition cascaded diffusion pipeline for text-to-video, noted for detailed textures.
- **VideoPoet**: Autoregressive transformer conditioned on text/audio for narrative or lip-synced content.
- **yu2023magvit / yu2023magvit-v2** [3,4]: Masked generative video transformers with high compression efficiency for AR or hybrid decoding.
- **DynamiCrafter**: Diffusion model optimized for dynamic motion fidelity in complex scenes.
- **Gen-1 / Gen-2** (Runway): Predecessors to Gen-3, with strong frame-to-frame consistency for video stylization and editing.
- **MagicFight** [29]: A system for personalized martial arts combat video generation.

5.3. Hybrid and Collaborative Efforts

Some systems emerge from joint academic–industry efforts, combining cutting-edge research with product polish. For example, collaborations between NVIDIA and universities have yielded NeRF-based view-consistent video synthesis, while ByteDance AI Lab publishes open technical reports alongside product releases.

5.4. Feature Comparison

Table 1 compares representative systems on backbone type, conditioning modalities, support for cascaded refinement, and target resolution.

Table 1. Representative video generation systems and key features. "Cascade" denotes multi-stage refinement; "Max Res." is the highest publicly reported resolution.

System	Backbone	Conditioning	Cascade	Real-time Capable	Max Res.	Primary Use Case
Sora	DiT + Latent Diffusion	Text, Image, Camera Path	✓	✗	1920×1080+	Cinematic, long-form video
Runway Gen-3	Latent Diffusion	Text, Image, Pose	✓	✗	1920×1080	Creative production
Pika	Latent Diffusion	Text, Image	✗	✓	1920×1080	Rapid prototyping
Stability Video	Latent Diffusion	Text, Image	✓	✗	2048×1152	Open-source creation
Veo	Latent Diffusion	Text, Camera Path	✓	✗	1280×768+	Productivity and creative
Hunyuan Video	Latent Diffusion	Text, Pose	✓	✗	1920×1080	Regional/cultural content
Firefly Video	Latent Diffusion	Text, Image, Audio	✓	✗	1920×1080	Professional editing
MagicVideo	Latent Diffusion	Text, Audio	✗	✓	1080×1920	Social media
NVIDIA ACE	Multi-Modal Transformers	Audio, Facial Pose	✗	✓	1920×1080	Digital humans

 doi:10.20944/preprints202511.1332.v1

5.5. Trends and Observations

- **Modalities:** Most modern systems accept at least text and image; adding audio, depth, or pose is becoming common.
- **Cascades:** Cascaded architectures remain dominant for high resolution, trading off latency.
- **Deployment:** Few systems offer real-time performance; those that do target lower resolutions or highly specialized domains (avatars, lip-sync).
- **Integration:** Increasing trend toward embedding generation tools directly into creative suites or productivity platforms.

6. Applications of IGV

Interactive Generative Video (IGV) systems enable on-demand creation of dynamic content tailored to user intent, often with direct control over motion, appearance, and structure [30,31]. Applications span entertainment, industry, science, and public services, with varying requirements for fidelity, interactivity, and safety [32,33].

6.1. Gaming and Generative Engines

In gaming, IGV facilitates:

- **Procedural World Generation**: Creating expansive, explorable environments at runtime, with variation seeded from player history or random inputs [34].
- **Dynamic Asset Creation**: Generating characters, textures, and animations on-the-fly, reducing the need for large asset libraries.
- **Player-Driven Narrative**: Adjusting cutscenes or environmental changes based on in-game events or player choices.
- **Simulation of NPC Behavior**: Producing responsive, contextually appropriate animations for non-playable characters [35].

IGV-powered game engines such as *GameGen-X* and *Genie2* integrate world modeling with reinforcement learning, enabling AI agents to train in environments that evolve in response to their actions [36,37].

6.2. Embodied AI and Robotics

Embodied AI agents benefit from IGV for:

- **Physics-Aware Simulation**: Generating rich, physically consistent scenes for robotic manipulation, locomotion, or human–robot interaction [38–40].
- **Cross-Domain Adaptation**: Training in synthetic environments with controllable complexity and noise, then transferring policies to the real world [41].
- **Multi-Sensor Fusion Testing**: Simulating not only RGB frames but also depth, segmentation, and infrared streams [42,43].

Sim-to-real transfer remains a challenge, requiring careful domain randomization and calibration [44].

6.3. Autonomous Driving

IGV supports the automotive sector through:

- **Closed-Loop Scenario Generation**: Creating realistic driving scenes to test perception and planning modules in varied weather, lighting, and traffic conditions [45–47].
- **Rare Event Synthesis**: Simulating edge cases (e.g., jaywalking pedestrians, sudden braking) that are rare in collected data but critical for safety [48,49].
- **Multi-Modal Sensor Simulation**: Generating synchronized LiDAR, RADAR, and camera feeds for sensor fusion pipelines [50–56].

Recent systems integrate IGV with differentiable simulators for end-to-end training and evaluation [57–60].

 doi:10.20944/preprints202511.1332.v1

6.4. Education and Knowledge Transfer

In education, IGV enables:

- **Automated Educational Video Creation**: Turning lecture notes or textbooks into animated explainers with synchronized narration.
- **Immersive Virtual Field Trips**: Rendering historical sites or scientific phenomena from textual prompts, enabling experiences otherwise inaccessible.
- **Personalized Learning Content**: Tailoring examples and pacing based on learner performance and preferences [61,62].

Generative models like *Genie3* can convert high-level descriptions into fully interactive 3D scenes for classroom VR.

6.5. Film and Media Production

Filmmakers use IGV for:

- **Pre-Visualization**: Quickly drafting scenes for pitch, budgeting, and planning.
- **Special Effects**: Generating or replacing complex shots without expensive location shoots or post-production [63].
- **Storyboarding and Style Transfer**: Maintaining consistent characters and environments across shots while applying artistic styles.

This accelerates the iterative cycle between concept and final cut.

6.6. Biomedicine and Healthcare

Biomedical applications demand domain-specific adaptations:

- **Surgical Training**: Simulating procedures in high resolution, with realistic tissue deformation and tool interaction.[64,65]
- **Medical Imaging Synthesis**: Generating ultrasound, endoscopy, or microscopy videos for rare conditions to augment datasets [66–68].
- **Physiological Process Visualization**: Rendering internal biological processes (e.g., blood flow, cell division) for education and diagnosis [69–74].
- **Augmented Rehabilitation**: Creating privacy-preserving frameworks for physical therapy, such as augmented knee rehabilitation programs [75].

Systems like *Bora* demonstrate multi-modal biomedical generation from text prompts.[76]

6.7. Security and Surveillance Simulation

Security agencies and researchers use IGV to:

- **Crowd Behavior Modeling**: Simulating public spaces under varying densities and events for safety planning [77,78].
- **Incident Replay and Forensics**: Reconstructing events from partial data[79,80] to test hypotheses or train response systems.

6.8. Industrial and Civil Engineering Applications

- **Infrastructure Inspection**: Generating synthetic data[81] for training models to detect defects in critical infrastructure, such as sewer pipes [82–84].
- **Geotechnical Evaluation**: Simulating geological conditions to evaluate the integrity of structures like tunnels [85–87].
- **Fault Diagnosis**: Creating simulations for hierarchical fault diagnosis in complex industrial systems [88–90].
- **Prognostics**: Simulating component wear-and-tear for AI-driven health prognostics, such as in Li-ion batteries [91].
- **Warehouse Automation**: Using reinforcement learning for efficient robot task scheduling, picking, and packing in automated warehouses [92,93].

 doi:10.20944/preprints202511.1332.v1

6.9. Business and Finance Applications

- **Decision Support**: Using generative models to create scenarios for business decision support systems [94].
- **Risk Analysis**: Generating synthetic supply chain data to train GNNs for credit risk analysis [95].

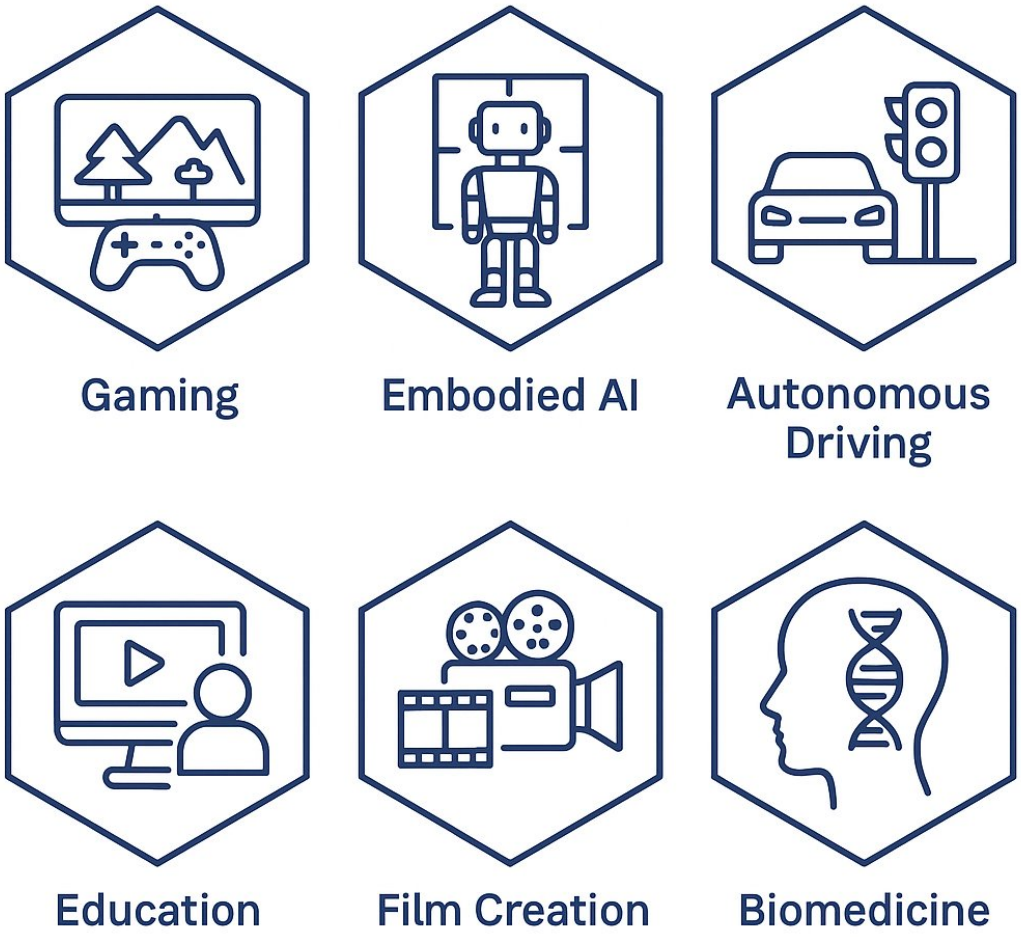

Figure 5. Representative application verticals for IGV, from entertainment to scientific and industrial domains.

6.10. Emerging Niches

- **Virtual Fashion Try-On**: Combining IGV with human pose transfer to simulate clothing on different body types.
- **Sports Analytics**: Visualizing alternative plays or player movements for coaching and broadcasting, as well as detailed performance analysis of specific actions like a golf swing or personalizing pedometer algorithms [96,97].
- **Telepresence and Digital Humans**: Creating avatars that respond in real-time for meetings, events, and customer service.

7. Systems Design: Toward Real-Time IGV

Latency. Few-step diffusion and consistency models reduce sampling; early-exit decoders and keyframe interpolation further cut latency. **Streaming.** Sliding-window attention with recurrent latent caches; online conditioning buffers for pose/depth/audio. **Scheduling.** Quality-of-service schedulers allocate compute across stages; adaptive guidance strength stabilizes identity.

 doi:10.20944/preprints202511.1332.v1

8. Data, Training, and Optimization

Data Quality. Long-video corpora with hierarchical metadata (shots, scenes, characters); deduplication and safety filtering. **Objective Design.** Multi-task training (recon + diffusion + adversarial + temporal).[98] **Optimization.** Mixed precision, gradient checkpointing, ZeRO/FS, [99]and distillation for deployment on edge devices.

9. Challenges and Future Directions

Real-time and Streaming. Reduce steps via distillation/consistency; non-autoregressive decoders.
Control and Editing. Open-domain conditioning (text, pose, depth, audio) and camera controls; interactive node-graph UIs.
Memory and Long-horizon Consistency. Segment memory, identity locking, and camera trajectory anchors.[100–102]
Physics and Dynamics. Learned physics priors, differentiable simulators, and safety constraints.
Generalization and Multimodality. Unified frameworks that jointly reason over text–audio–vision–kinematics, [103,104]especially for enhancing intent understanding from ambiguous prompts [105].
Governance and Safety. Detection, authentication/watermarking, usage policies, transparency, and energy efficiency.[106,107]

10. Evaluation

Spatial. FID, IS, SSIM, PSNR, LPIPS, CLIPSIM. **Temporal.** FVD, KVD, temporal LPIPS, warping error. **Human Studies.** MOS, task success, preference tests. **Robustness.** Sensitivity to prompt/seed, stability across edits, safety filters [108]. **Domain Metrics.** Driving safety proxies, clinical/educational utility, gameplay KPIs.

11. Ethical and Legal Considerations

Misinformation and Deepfakes. Risk of realistic fabricated content; invest in provenance and detection [109,110]. **IP and Licensing.** Data consent and derivative rights; licensing and opt-out mechanisms. **Bias, Fairness, and Privacy.** Inclusive datasets, audits, privacy-preserving learning (DP, federated)[111,112]. **Accountability and Explainability.** Responsibility for deployment harms; interpretable controls in sensitive domains [113]. **Environmental Impact.** Efficient architectures, mixed-precision, and renewable-powered training. **Fraud Detection.** Evaluating models for fraud on imbalanced transaction data is a key concern [114].

12. Conclusion

Video generation is converging toward interactive, controllable, and reliable media engines. Progress depends on efficient backbones, robust control, principled evaluation, and strong governance.

References

1. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Advances in Neural Information Processing Systems* **2014**, [1406.2661].
2. Yan, R.; Jiang, Y.; Song, W.; et al. VideoGPT: Video Generation using VQ-VAE and Transformers. *arXiv preprint arXiv:2104.10157* **2021**.
3. Yu, L.; Cheng, Y.; Sohn, K.; Lezama, J.; Zhang, H.; Chang, H.; Hauptmann, A.G.; Yang, M.H.; Hao, Y.; Essa, I.; et al. Magvit: Masked generative video transformer. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 10459–10469.
4. Yu, L.; Lezama, J.; Gundavarapu, N.B.; Versari, L.; Sohn, K.; Minnen, D.; Cheng, Y.; Birodkar, V.; Gupta, A.; Gu, X.; et al. Language Model Beats Diffusion–Tokenizer is Key to Visual Generation. *arXiv preprint arXiv:2310.05737* **2023**.
5. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. *NeurIPS* **2020**, [2006.11239].

 doi:10.20944/preprints202511.1332.v1

6. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10684–10695.
7. Ho, J.; Chan, W.; Saharia, C.; Whang, J.; Gao, R.; Gritsenko, A.; Kingma, D.P.; Poole, B.; Norouzi, M.; Fleet, D.J.; et al. Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303* **2022**.
8. Tulyakov, S.; Liu, M.Y.; Yang, X.; Kautz, J. MoCoGAN: Decomposing Motion and Content for Video Generation. In Proceedings of the CVPR, 2018.
9. Huang, J.; Zhang, G.; Jie, Z.; Jiao, S.; Qian, Y.; Chen, L.; Wei, Y.; Ma, L. M4V: Multi-Modal Mamba for Text-to-Video Generation. *arXiv preprint arXiv:2506.10915* **2025**.
10. Zhang, L.; Agrawala, M. Adding conditional control to text-to-image diffusion models. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 3836–3847.
11. Huang, Y.; Huang, J.; Liu, Y.; Yan, M.; Lv, J.; Liu, J.; Xiong, W.; Zhang, H.; Cao, L.; Chen, S. Diffusion model-based image editing: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2025**.
12. Qu, D.; Chen, Q.; Zhang, P.; Gao, X.; Li, J.; Zhao, B.; Wang, D.; Li, X. Livescene: Language embedding interactive radiance fields for physical scene rendering and control. *arXiv preprint arXiv:2406.16038* **2024**.
13. Chen, Q.; Qu, D.; Tang, Y.; Song, H.; Zhang, Y.; Zhao, B.; Wang, D.; Li, X. FreeGaussian: Guidance-free Controllable 3D Gaussian Splats with Flow Derivatives **2024**.
14. Ding, T.; Xiang, D.; Rivas, P.; Dong, L. Neural Pruning for 3D Scene Reconstruction: Efficient NeRF Acceleration. *arXiv preprint arXiv:2504.00950* **2025**.
15. Li, J.; Zhou, Y. BIDeepLab: An Improved Lightweight Multi-scale Feature Fusion Deeplab Algorithm for Facial Recognition on Mobile Devices. *Computer Simulation in Application* **2025**, *3*, 57–65.
16. Liang, X.; He, Y.; Tao, M.; Xia, Y.; Wang, J.; Shi, T.; Wang, J.; Yang, J. Cmat: A multi-agent collaboration tuning framework for enhancing small language models. *arXiv preprint arXiv:2404.01663* **2024**.
17. Tang, X.; Li, X.; Tasdizen, T. Dynamic Scale for Transformer. In Proceedings of the Medical Imaging with Deep Learning-Short Papers, 2025.
18. Radford, A.; et al. Learning Transferable Visual Models From Natural Language Supervision. *ICML* **2021**, [2103.00020].
19. Li, Z.; Fu, Z.; Hu, Y.; Chen, Z.; Wen, H.; Nie, L. FineCIR: Explicit Parsing of Fine-Grained Modification Semantics for Composed Image Retrieval. *https://arxiv.org/abs/2503.21309* **2025**.
20. Li, Z.; Chen, Z.; Wen, H.; Fu, Z.; Hu, Y.; Guan, W. Encoder: Entity mining and modification relation binding for composed image retrieval. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2025, Vol. 39, pp. 5101–5109.
21. Chen, Z.; Hu, Y.; Li, Z.; Fu, Z.; Song, X.; Nie, L. OFFSET: Segmentation-based Focus Shift Revision for Composed Image Retrieval, 2025, [arXiv:cs.CV/2507.05631].
22. Liu, J.; Wang, G.; Ye, W.; Jiang, C.; Han, J.; Liu, Z.; Zhang, G.; Du, D.; Wang, H. DifFlow3D: Toward Robust Uncertainty-Aware Scene Flow Estimation with Iterative Diffusion-Based Refinement. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 15109–15119.
23. Wang, J.; Zhang, Z.; He, Y.; Zhang, Z.; Song, Y.; Shi, T.; Li, Y.; Xu, H.; Wu, K.; Yi, X.; et al. Enhancing Code LLMs with Reinforcement Learning in Code Generation: A Survey. *arXiv preprint arXiv:2412.20367* **2024**.
24. Yi, Q.; He, Y.; Wang, J.; Song, X.; Qian, S.; Yuan, X.; Sun, L.; Xin, Y.; Tang, J.; Li, K.; et al. Score: Story coherence and retrieval enhancement for ai narratives. *arXiv preprint arXiv:2503.23512* **2025**.
25. OpenAI. Video generation models as world simulators. https://openai.com/research/video-generation-models-as-world-simulators, 2024.
26. Runway. Introducing Gen-3 Alpha. https://runwayml.com/blog/introducing-gen-3-alpha/, 2024.
27. Pika. Pika. https://pika.art, 2024.
28. Google. Veo: a generative video model. https://deepmind.google/technologies/veo/, 2024.
29. Huang, J.; Yan, M.; Chen, S.; Huang, Y.; Chen, S. Magicfight: Personalized martial arts combat video generation. In Proceedings of the Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 10833–10842.
30. Zheng, Z.; Liu, K.; Zhu, X. Machine Learning-Based Prediction of Metal-Organic Framework Materials: A Comparative Analysis of Multiple Models. *arXiv preprint arXiv:2507.04493* **2025**.
31. Wang, J.; Ding, W.; Zhu, X. Financial analysis: Intelligent financial data analysis system based on llm-rag. *arXiv preprint arXiv:2504.06279* **2025**.

 doi:10.20944/preprints202511.1332.v1

32. Yuan, T.; Zhang, X.; Chen, X. Machine Learning based Enterprise Financial Audit Framework and High Risk Identification. *arXiv preprint arXiv:2507.06266* **2025**.
33. Li, Z.; Qiu, S.; Ke, Z. Revolutionizing Drug Discovery: Integrating Spatial Transcriptomics with Advanced Computer Vision Techniques. In Proceedings of the 1st CVPR Workshop on Computer Vision For Drug Discovery (CVDD): Where are we and What is Beyond?, 2025.
34. Li, Z.; Ke, Z. Cross-Modal Augmentation for Low-Resource Language Understanding and Generation. In Proceedings of the Proceedings of the 1st Workshop on Multimodal Augmented Generation via Multimodal Retrieval (MAGMaR 2025), 2025, pp. 90–99.
35. Zhang, Z.; Wang, J.; Li, Z.; Wang, Y.; Zheng, J. AnnCoder: A Mti-Agent-Based Code Generation and Optimization Model. *Symmetry* **2025**, *17*. https://doi.org/10.3390/sym17071087.
36. Li, Z.; Ji, Q.; Ling, X.; Liu, Q. A Comprehensive Review of Multi-Agent Reinforcement Learning in Video Games. *IEEE Transactions on Games* **2025**, pp. 1–21. https://doi.org/10.1109/TG.2025.3588809.
37. Li, Z. Language-Guided Multi-Agent Learning in Simulations: A Unified Framework and Evaluation, 2025, [arXiv:cs.AI/2506.04251].
38. Jin, S.; Wang, X.; Meng, Q. Spatial memory-augmented visual navigation based on hierarchical deep reinforcement learning in unknown environments. *Knowledge-Based Systems* **2024**, *285*, 111358.
39. Qian, H.; Jin, S.; Chen, L. I2KEN: Intra-Domain and Inter-Domain Knowledge Enhancement Network for Lifelong Loop Closure Detection. *IEEE Robotics and Automation Letters* **2025**.
40. Gu, N.; Kosuge, K.; Hayashibe, M. TactileAloha: Learning Bimanual Manipulation With Tactile Sensing. *IEEE Robotics and Automation Letters* **2025**, *10*, 8348–8355. https://doi.org/10.1109/LRA.2025.3585396.
41. Jin, S.; Dai, X.; Meng, Q. "Focusing on the right regions"—Guided saliency prediction for visual SLAM. *Expert Systems with Applications* **2023**, *213*, 119068.
42. Zhao, Z.; Chen, B.M. Benchmark for Evaluating Initialization of Visual-Inertial Odometry. In Proceedings of the 2023 42nd Chinese Control Conference (CCC), 2023, pp. 3935–3940.
43. Zhao, Z. Balf: Simple and efficient blur aware local feature detector. In Proceedings of the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 3362–3372.
44. Liu, J.; Wang, G.; Liu, Z.; Jiang, C.; Pollefeys, M.; Wang, H. RegFormer: an efficient projection-aware transformer network for large-scale point cloud registration. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 8451–8460.
45. Zeng, S.; Chang, X.; Xie, M.; Liu, X.; Bai, Y.; Pan, Z.; Xu, M.; Wei, X. FutureSightDrive: Thinking Visually with Spatio-Temporal CoT for Autonomous Driving. *arXiv preprint arXiv:2505.17685* **2025**.
46. Ma, Z.; Luo, Y.; Zhang, Z.; Sun, A.; Yang, Y.; Liu, H. Reinforcement Learning Approach for Highway Lane-Changing: PPO-Based Strategy Design. *Preprints* **2025**. https://doi.org/10.20944/preprints202506.2087.v1.
47. Chen, Y.; Greer, R. Technical Report for Argoverse2 Scenario Mining Challenges on Iterative Error Correction and Spatially-Aware Prompting, 2025, [arXiv:cs.CV/2506.11124].
48. Huang, Z.; Qian, H.; Cai, Z.; Wang, X.; Xie, L.; Niu, X. An intelligent multilane roadway recognition method based on pseudo-tagging. *Cartography and Geographic Information Science* **2025**, pp. 1–16.
49. Qian, Y.; Li, X.; Zhang, J.; Meng, X.; Li, Y.; Ding, H.; Wang, M. A Diffusion-TGAN Framework for Spatio-Temporal Speed Imputation and Trajectory Reconstruction. *IEEE Transactions on Intelligent Transportation Systems* **2025**.
50. Zhou, S.; Tian, Z.; Chu, X.; Zhang, X.; Zhang, B.; Lu, X.; Feng, C.; Jie, Z.; Chiang, P.Y.; Ma, L. FastPillars: A Deployment-friendly Pillar-based 3D Detector. *arXiv preprint arXiv:2302.02367* **2023**.
51. Zhou, S.; Li, L.; Zhang, X.; Zhang, B.; Bai, S.; Sun, M.; Zhao, Z.; Lu, X.; Chu, X. LiDAR-PTQ: Post-Training Quantization for Point Cloud 3D Object Detection. In Proceedings of the The Twelfth International Conference on Learning Representations, 2024.
52. Zhou, S.; Nie, J.; Zhao, Z.; Cao, Y.; Lu., X. FocusTrack: One-Stage Focus-and-Suppress Framework for 3D Point Cloud Object Tracking. In Proceedings of the Proceedings of the 33rd ACM International Conference on Multimedia, 2025.
53. Liu, J.; Zhuo, D.; Feng, Z.; Zhu, S.; Peng, C.; Liu, Z.; Wang, H. Dvlo: Deep visual-lidar odometry with local-to-global feature fusion and bi-directional structure alignment. In Proceedings of the European Conference on Computer Vision. Springer, 2025, pp. 475–493.
54. Yao, S.; Guan, R.; Peng, Z.; Xu, C.; Shi, Y.; Yue, Y.; Lim, E.G.; Seo, H.; Man, K.L.; Zhu, X.; et al. Exploring radar data representations in autonomous driving: A comprehensive review. *IEEE Transactions on Intelligent Transportation Systems* **2025**, *26*, 7401–7425. https://doi.org/10.1109/TITS.2025.3554781.

 doi:10.20944/preprints202511.1332.v1

55. Yao, S.; Guan, R.; Wu, Z.; Ni, Y.; Huang, Z.; Liu, R.W.; Yue, Y.; Ding, W.; Lim, E.G.; Seo, H.; et al. Waterscenes: A multi-task 4d radar-camera fusion dataset and benchmarks for autonomous driving on water surfaces. *IEEE Transactions on Intelligent Transportation Systems* **2024**, *25*, 16584–16598.
56. Yao, S.; Guan, R.; Huang, X.; Li, Z.; Sha, X.; Yue, Y.; Lim, E.G.; Seo, H.; Man, K.L.; Zhu, X.; et al. Radar-camera fusion for object detection and semantic segmentation in autonomous driving: A comprehensive review. *IEEE Transactions on Intelligent Vehicles* **2023**, *9*, 2094–2128.
57. Li, X.; Mangin, T.; Saha, S.; Mohammed, R.; Blanchard, E.; Tang, D.; Poppe, H.; Choi, O.; Kelly, K.; Whitaker, R. Real-time idling vehicles detection using combined audio-visual deep learning. In *Emerging Cutting-Edge Developments in Intelligent Traffic and Transportation Systems*; IOS Press, 2024; pp. 142–158.
58. Li, X.; Mohammed, R.; Mangin, T.; Saha, S.; Kelly, K.; Whitaker, R.; Tasdizen, T. Joint audio-visual idling vehicle detection with streamlined input dependencies. In Proceedings of the Proceedings of the Winter Conference on Applications of Computer Vision, 2025, pp. 885–894.
59. Lu, B.; Lu, Z.; Qi, Y.; Guo, H.; Sun, T.; Zhao, Z. Predicting Asphalt Pavement Friction by Using a Texture-Based Image Indicator. *Lubricants* **2025**, *13*, 341.
60. Lu, B.; Dan, H.C.; Zhang, Y.; Huang, Z. Journey into automation: Image-derived pavement texture extraction and evaluation. *arXiv preprint arXiv:2501.02414* **2025**.
61. Cui, X.; Lu, W.; Tong, Y.; Li, Y.; Zhao, Z. Multi-Modal Multi-Behavior Sequential Recommendation with Conditional Diffusion-Based Feature Denoising. In Proceedings of the Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2025, pp. 1593–1602.
62. Liu, S.; Zhang, Y.; Li, X.; Liu, Y.; Feng, C.; Yang, H. Gated Multimodal Graph Learning for Personalized Recommendation. *INNO-PRESS: Journal of Emerging Applied AI* **2025**, *1*.
63. Zhang, H.; Xu, H.; Liu, H.; Yu, X.; Zhang, X.; Wu, C. Conditional variational underwater image enhancement with kernel decomposition and adaptive hybrid normalization. *Neurocomputing* **2025**, p. 130845.
64. Yue, L.; Xing, S.; Lu, Y.; Fu, T. Biomamba: A pre-trained biomedical language representation model leveraging mamba. *arXiv preprint arXiv:2408.02600* **2024**.
65. Wang, Y.; Fu, T.; Xu, Y.; Ma, Z.; Xu, H.; Du, B.; Lu, Y.; Gao, H.; Wu, J.; Chen, J. TWIN-GPT: digital twins for clinical trials via large language model. *ACM Transactions on Multimedia Computing, Communications and Applications* **2024**.
66. Zhong, J.; Wang, Y.; Zhu, D.; Wang, Z. A Narrative Review on Large AI Models in Lung Cancer Screening, Diagnosis, and Treatment Planning. *arXiv preprint arXiv:2506.07236* **2025**.
67. Wang, Y.; Wang, Z.; Zhong, J.; Zhu, D.; Li, W. Applications of Small Language Models in Medical Imaging Classification with a Focus on Prompt Strategies. *arXiv preprint arXiv:2508.13378* **2025**.
68. Zhao, Y.; Zhu, Z.; Yang, S.; Li, W. YeastSAM: A Deep Learning Model for Accurate Segmentation of Budding Yeast Cells. *bioRxiv* **2025**, pp. 2025–09.
69. Ding, T.; Xiang, D.; Schubert, K.E.; Dong, L. GKAN: Explainable Diagnosis of Alzheimer's Disease Using Graph Neural Network with Kolmogorov-Arnold Networks. *arXiv preprint arXiv:2504.00946* **2025**.
70. Wang, Y.; Zhong, J.; Kumar, R. A systematic review of machine learning applications in infectious disease prediction, diagnosis, and outbreak forecasting **2025**.
71. Zhong, J.; Wang, Y. A Comparative Study of Ensemble Models for Thyroid Disease Prediction under Class Imbalance **2025**.
72. Zhao, Y.; Li, C.; Shu, C.; Wu, Q.; Li, H.; Xu, C.; Li, T.; Wang, Z.; Luo, Z.; He, Y. Tackling Small Sample Survival Analysis via Transfer Learning: A Study of Colorectal Cancer Prognosis. *arXiv preprint arXiv:2501.12421* **2025**.
73. Lu, Y.; Wu, C.T.; Parker, S.J.; Cheng, Z.; Saylor, G.; Van Eyk, J.E.; Yu, G.; Clarke, R.; Herrington, D.M.; Wang, Y. COT: an efficient and accurate method for detecting marker genes among many subtypes. *Bioinformatics Advances* **2022**, *2*, vbac037.
74. Fu, Y.; Lu, Y.; Wang, Y.; Zhang, B.; Zhang, Z.; Yu, G.; Liu, C.; Clarke, R.; Herrington, D.M.; Wang, Y. Ddn3. 0: Determining significant rewiring of biological network structure with differential dependency networks. *Bioinformatics* **2024**, *40*, btae376.
75. Bačić, B.; Vasile, C.; Feng, C.; Ciucă, M.G. Towards nation-wide analytical healthcare infrastructures: A privacy-preserving augmented knee rehabilitation case study. *arXiv preprint arXiv:2412.20733* **2024**.
76. Lu, Y.; Sato, K.; Wang, J. Deep learning based multi-label image classification of protest activities. *arXiv preprint arXiv:2301.04212* **2023**.
77. Xu, Z.; Liu, Y. Robust Anomaly Detection in Network Traffic: Evaluating Machine Learning Models on CICIDS2017, 2025, [arXiv:cs.CR/2506.19877].

 doi:10.20944/preprints202511.1332.v1

78. Lyu, J.; Zhao, M.; Hu, J.; Huang, X.; Chen, Y.; Du, S. VADMamba: Exploring State Space Models for Fast Video Anomaly Detection, 2025, [arXiv:cs.CV/2503.21169].
79. Chen, L.; Lu, Y.; Wu, C.T.; Clarke, R.; Yu, G.; Van Eyk, J.E.; Herrington, D.M.; Wang, Y. Data-driven detection of subtype-specific differentially expressed genes. *Scientific reports* **2021**, *11*, 332.
80. Chen, H.; Zhao, W.; Zhang, R.; Li, N.; Li, D. Multiple Object Tracking in Video SAR: A Benchmark and Tracking Baseline. *IEEE Geoscience and Remote Sensing Letters* **2025**, *22*, 1–5. https://doi.org/10.1109/LGRS.2025.3592711.
81. Lu, Y.; Shen, M.; Wang, H.; Wang, X.; van Rechem, C.; Fu, T.; Wei, W. Machine learning for synthetic data generation: a review. *arXiv preprint arXiv:2302.04062* **2023**.
82. Zhao, C.; Hu, C.; Shao, H.; Liu, J. PipeMamba: State Space Model for Efficient Video-based Sewer Defect Classification. *IEEE Transactions on Artificial Intelligence* **2025**.
83. Zhao, C.; Hu, C.; Shao, H.; Dunkin, F.; Wang, Y. Trusted video-based sewer inspection via support clip-based pareto-optimal evidential network. *IEEE Signal Processing Letters* **2024**.
84. Hu, C.; Zhao, C.; Shao, H.; Deng, J.; Wang, Y. TMFF: Trustworthy multi-focus fusion framework for multi-label sewer defect classification in sewer inspection videos. *IEEE Transactions on Circuits and Systems for Video Technology* **2024**.
85. Wu, C.; Huang, H.; Zhang, L.; Chen, J.; Tong, Y.; Zhou, M. Towards automated 3D evaluation of water leakage on a tunnel face via improved GAN and self-attention DL model. *Tunnelling and Underground Space Technology* **2023**, *142*, 105432.
86. Wu, C.; Huang, H.; Ni, Y.Q. Evaluation of Tunnel Rock Mass Integrity Using Multi-Modal Data and Generative Large Model: Tunnel Rip-Gpt. *Available at SSRN 5348429* **2025**.
87. Wu, C.; Huang, H.; Chen, J.; Zhou, M.; Han, S. A novel Tree-augmented Bayesian network for predicting rock weathering degree using incomplete dataset. *International Journal of Rock Mechanics and Mining Sciences* **2024**, *183*, 105933.
88. Sha, Y.; Gou, S.; Liu, B.; Faber, J.; Liu, N.; Schramm, S.; Stoecker, H.; Steckenreiter, T.; Vnucec, D.; Wetzstein, N.; et al. Hierarchical knowledge guided fault intensity diagnosis of complex industrial systems. In Proceedings of the Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2024, pp. 5657–5668.
89. Sha, Y.; Gou, S.; Faber, J.; Liu, B.; Li, W.; Schramm, S.; Stoecker, H.; Steckenreiter, T.; Vnucec, D.; Wetzstein, N.; et al. Regional-local adversarially learned one-class classifier anomalous sound detection in global long-term space. In Proceedings of the Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 3858–3868.
90. Sha, Y.; Faber, J.; Gou, S.; Liu, B.; Li, W.; Schramm, S.; Stoecker, H.; Steckenreiter, T.; Vnucec, D.; Wetzstein, N.; et al. A multi-task learning for cavitation detection and cavitation intensity recognition of valve acoustic signals. *Engineering Applications of Artificial Intelligence* **2022**, *113*, 104904.
91. Ding, T.; Xiang, D.; Sun, T.; Qi, Y.; Zhao, Z. AI-driven prognostics for state of health prediction in Li-ion batteries: A comprehensive analysis with validation. *arXiv preprint arXiv:2504.05728* **2025**.
92. Wu, S.; Fu, L.; Chang, R.; Wei, Y.; Zhang, Y.; Wang, Z.; Liu, L.; Zhao, H.; Li, K. Warehouse robot task scheduling based on reinforcement learning to maximize operational efficiency. *Authorea Preprints* **2025**.
93. Yu, D.; Liu, L.; Wu, S.; Li, K.; Wang, C.; Xie, J.; Chang, R.; Wang, Y.; Wang, Z.; Ji, R. Machine learning optimizes the efficiency of picking and packing in automated warehouse robot systems. In Proceedings of the 2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE). IEEE, 2025, pp. 1325–1332.
94. Zhang, Y.; Chen, N.; Zhang, Y.; Wu, W. Research on business decision support system based on big data and artificial intelligence. *Available at SSRN 5332298* **2025**.
95. Zhang, Z.; Shen, Q.; Hu, Z.; Liu, Q.; Shen, H. Credit Risk Analysis for SMEs Using Graph Neural Networks in Supply Chain. *arXiv preprint arXiv:2507.07854* **2025**.
96. Bačić, B.; Feng, C.; Li, W. Jy61 imu sensor external validity: A framework for advanced pedometer algorithm personalisation. *ISBS Proceedings Archive* **2024**, *42*, 60.
97. Feng, C.; Bačić, B.; Li, W. Sca-lstm: A deep learning approach to golf swing analysis and performance enhancement. In Proceedings of the International Conference on Neural Information Processing. Springer, 2024, pp. 72–86.
98. Li, X.; Ma, Y.; Huang, Y.; Wang, X.; Lin, Y.; Zhang, C. Synergized data efficiency and compression (sec) optimization for large language models. In Proceedings of the 2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS). IEEE, 2024, pp. 586–591.

 doi:10.20944/preprints202511.1332.v1

99. He, L.; Wang, X.; Lin, Y.; Li, X.; Ma, Y.; Li, Z. BOANN: Bayesian-Optimized Attentive Neural Network for Classification. In Proceedings of the 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), 2024, pp. 1860–1864. https://doi.org/10.1109/ICICML63543.2024.10957983.
100. Fei, Y.; He, Y.; Chen, F.; You, P.; Zhai, H. Optimal Planning and Design for Sightseeing Offshore Island Microgrids. In Proceedings of the E3S Web of Conferences. EDP Sciences, 2019, Vol. 118, p. 02044.
101. Zhai, M.; Abu-Znad, O.; Wang, S.; Du, L. A Bayesian Potential-based Architecture for Mutually Beneficial EV Charging Infrastructure-DSO Coordination. *IEEE Transactions on Transportation Electrification* **2025**, pp. 1–1. https://doi.org/10.1109/TTE.2025.3576721.
102. Zhai, M.; Tian, X.; Liu, Z.; Zhao, Y.; Deng, Y.; Yang, W. Advancing just transition: The role of biomass co-firing in emission reductions and employment for coal regions. *Sustainable Energy Technologies and Assessments* **2025**, *75*, 104246.
103. Li, X.; Ma, Y.; Ye, K.; Cao, J.; Zhou, M.; Zhou, Y. Hy-Facial: Hybrid Feature Extraction by Dimensionality Reduction Methods for Enhanced Facial Expression Classification, 2025, [arXiv:cs.CV/2509.26614].
104. Zhao, M.; Chen, Y.; Lyu, J.; Du, S.; Lv, Z.; Wang, L. SDAFE: A Dual-filter Stable Diffusion Data Augmentation Method for Facial Expression Recognition. In Proceedings of the ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2025, pp. 1–5. https://doi.org/10.1109/ICASSP49660.2025.10888031.
105. He, Y.; Wang, J.; Wang, Y.; Li, K.; Zhong, Y.; Song, X.; Sun, L.; Lu, J.; Tang, J.; Zhang, M.; et al. Enhancing intent understanding for ambiguous prompt: A human-machine co-adaption strategy. *arXiv preprint arXiv:2501.15167* **2025**.
106. Zhang, J.; Cao, J.; Chang, J.; Li, X.; Liu, H.; Li, Z. Research on the Application of Computer Vision Based on Deep Learning in Autonomous Driving Technology. *arXiv preprint arXiv:2406.00490* **2024**.
107. Li, Z.; Guan, B.; Wei, Y.; Zhou, Y.; Zhang, J.; Xu, J. Mapping new realities: Ground truth image creation with pix2pix image-to-image translation. *arXiv preprint arXiv:2404.19265* **2024**.
108. Li, Z. Investigating Spurious Correlations in Vision Models Using Counterfactual Images. In Proceedings of the First Workshop on Experimental Model Auditing via Controllable Synthesis at CVPR 2025, 2025.
109. Tong, Y.; Lu, W.; Zhao, Z.; Lai, S.; Shi, T. MMDFND: Multi-modal multi-domain fake news detection. In Proceedings of the Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 1178–1186.
110. Lu, W.; Tong, Y.; Ye, Z. DAMMFND: Domain-Aware Multimodal Multi-view Fake News Detection. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2025, Vol. 39, pp. 559–567.
111. Tao, Y.; Jia, Y.; Wang, N.; Wang, H. The FacT: Taming Latent Factor Models for Explainability with Factorization Trees. In Proceedings of the Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA, 2019; SIGIR'19, p. 295–304. https://doi.org/10.1145/3331184.3331244.
112. Tao, Y.; Wang, Z.; Zhang, H.; Wang, L. NEVLP: Noise-Robust Framework for Efficient Vision-Language Pre-training. *arXiv preprint arXiv:2409.09582* **2024**, [arXiv:cs.CV/2409.09582].
113. Lu, Y.; Yang, W.; Zhang, Y.; Chen, Z.; Chen, J.; Xuan, Q.; Wang, Z.; Yang, X. Understanding the Dynamics of DNNs Using Graph Modularity. In Proceedings of the Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XII, Berlin, Heidelberg, 2022; p. 225–242. https://doi.org/10.1007/978-3-031-19775-8_14.
114. Wang, C.; Nie, C.; Liu, Y. Evaluating Supervised Learning Models for Fraud Detection: A Comparative Study of Classical and Deep Architectures on Imbalanced Transaction Data. *arXiv preprint arXiv:2505.22521* **2025**.