4D World Models: Bridging Generation and Reconstruction

Abstract

The realistic reconstruction and generation of dynamic virtual worlds is an essential task for scene understanding and robot learning. Recent advances in NeRF and 3DGS have driven the progress of reconstruction and generative models in 3D/4D. However, 3D/4D generation and reconstruction have largely evolved as separate research tracks, despite their conceptual overlap and potential for synergy. Integrating these paradigms can offer a wider perspective to identify optimal representation and scalable data strategies for building 4D world models. This workshop aims to bring together researchers to explore the unified frameworks of understanding, reconstructing, and generating the 3D/4D world.

1. TOPIC

A longstanding goal in computer vision and computer graphics is the realistic reconstruction and generation of dynamic virtual worlds; realistic or stylized environments that evolve coherently over time and space. Achieving high-fidelity 4D world generation at scale is a critical step toward deeper scene understanding, immersive XR experiences, and robust simulation for autonomous systems. In recent years, there has been a surge of interest in 3D and 4D generation, driven by advances in generative modeling and implicit representations such as Neural Radiance Fields (NeRFs) and 3D Gaussian Splatting. These developments have enabled unprecedented progress in both the generation of novel 3D/4D content and the reconstruction of real-world dynamic scenes.

Despite their conceptual overlap, generation and reconstruction have largely evolved as parallel research tracks. Generative models benefit from realistic reconstructed data for supervision and grounding, while reconstruction pipelines can leverage generative priors for completing missing geometry, improving realism, and enhancing temporal coherence. The integration of these two paradigms offers the potential to build 4D world models - a unified frameworks capable of understanding, reconstructing, and generating consistent, temporally dynamic representations of reality.

This workshop aims to bridge the gap between 4D generation and 4D reconstruction, creating a forum for both academic and industrial researchers to:

1. Explore how generation and reconstruction approaches can be synergistically combined;
2. Identify the representations and architectures best suited for integrated 4D modeling;
3. Discuss scalable data capture, curation, and evaluation for learning 4D world models.

By fostering cross-disciplinary dialogue, this initiative seeks to move beyond static scene understanding and toward learning complete, generative 4D world models that can perceive and create the dynamic world around us.

1.1. Call for Speakers

To advance the vision of unified **World Models** that can understand, reconstruct, and generate coherent 4D environments, we invite speakers working at the intersection of generative modeling, 3D/4D reconstruction, and dynamic scene understanding. We particularly seek contributions around the following key research directions:

- **Generative and Editable World Representations**: Building world models requires controllable generation of realistic and consistent 3D/4D content. Recent works enable text- or image-driven 3D scene synthesis (**Prometheus**, **WonderWorld**), structured 3DGS representations (**UVGS**), and scalable 3D generation frameworks (**TRELLIS**). Equally important are text-driven editing approaches (**Morpheus**), which enable fine-grained control and semantic consistency across generated scenes—core capabilities for interactive world modeling and digital twins.
- **Efficient and Scalable World Rendering**: True world models must operate in real time and at scale, supporting low-latency prediction and visualization on compute-constrained devices. Advances in feed-forward reconstruction (**4DGT**), efficient 3D Gaussian Splatting (**Taming 3DGS**), and high-fidelity volumetric rendering (**Triangle Splatting**) provide critical foundations for scaling dynamic world models to immersive AR/VR and robotics applications.
- **Dynamic Scene Understanding and Data Capture**: World models must capture not just static geometry but evolving reality. Representing full spatio-temporal dynamics with Radiance Fields and Gaussian Splatting remains an open challenge, especially in achieving temporal coherence and scalability. Recent progress in large-scale 4D datasets and dynamic rendering (**DIVA-360**, **LongVolCap**, **FreeTimeGS**) offers a path toward continuously learnable world models that can perceive, predict, and simulate dynamic environments.
- **Virtual Humans as Embodied Agents in World Models**: Photo-realistic, animatable avatars are essential components of human-centric world models. Recent work on immersive volumetric humans (**RePerformer**, **Splatting Avatar**), real-time rendering (**GPS-Gaussian**), and deformable interaction (**InteractAvatar**) demonstrates progress toward embodied agents that interact naturally within learned worlds.

By uniting these threads, this workshop seeks to highlight the emerging convergence between 3D/4D vision, generative modeling, and world understanding while paving the way for next-generation **4D World Models** that can learn, imagine, and simulate the dynamic environments we live in.

1.2. Broader impact statement

This workshop is highly relevant to the computer vision, graphics, and AI communities, as it unites several of the most active and rapidly evolving research directions: Radiance Fields, 3D Gaussian Splatting, dynamic scene reconstruction, and world models for perception and generation. These threads converge on a shared ambition to endow AI systems with the capacity to understand, reconstruct, and generate coherent representations of dynamic world. By bridging the traditionally distinct domains of reconstruction and generation, the workshop promotes a unified view of scene modeling that directly supports progress in embodied intelligence, simulation, and creative content creation.

At the community level, this workshop provides a timely forum for researchers to address emerging challenges: how to capture and reconstruct dynamic 4D data at scale; how to design models that are both physically grounded and generative; and how to evaluate realism, temporal coherence, and controllability in a standardized way. These questions sit at the heart of ongoing efforts to transition from static 3D perception toward interactive, generative 4D understanding — a direction strongly reflected in the increasing number of NeRF, 3DGS, and world model papers across recent CVPR editions.

Beyond the CVPR community, the relevance of this workshop extends deeply into industry and interdisciplinary research. Neural and Gaussian representations are becoming foundational tools for AR/VR, digital twins, robotics, autonomous driving, simulation for reinforcement learning, and AI-driven creative industries. The rise of world models that learn structured, predictive representations of the environment has also sparked interest across machine learning, embodied AI, and generative simulation communities. The outcomes of this workshop will thus resonate with researchers and practitioners in fields ranging from robotics and computer graphics to AI-assisted content generation and virtual environment design.

In the longer term, progress in unified 4D world modeling has the potential to transform how we capture, understand, and synthesize the physical world enabling AI systems that can perceive, imagine, and interact with dynamic environments with unprecedented realism and fidelity. By fostering collaboration between academic and industrial researchers, this workshop aims to catalyze new benchmarks, datasets, and paradigms that push the boundaries of what it means to model, simulate, and create in four dimensions.

1.3. Ethical Considerations Around the Topic

The development of **4D World Models** (systems capable of perceiving, reconstructing, and generating dynamic environments) brings both significant opportunities and important ethical responsibilities. On the positive side, advances in neural radiance fields, Gaussian splatting, and generative modeling can enable transformative applications in immersive training, education, scientific visualization, entertainment, and assistive technologies. Moreover, progress in efficient rendering and inference, a key theme of this workshop, can reduce computational costs and energy consumption, contributing to more sustainable AI systems.

At the same time, these capabilities introduce potential risks related to **misuse and authenticity**. High-fidelity dynamic scene generation can facilitate the creation of deceptive or non-consensual synthetic media, including realistic reenactments or altered identities. As these technologies mature, responsible research practices become essential. We encourage the community to develop complementary tools for **content authentication, watermarking, and fake-media detection**, and to promote transparent data collection and consent practices in human-centric capture.

Finally, we urge researchers and practitioners to consider the societal and environmental impacts of their work, ensuring that progress toward comprehensive 4D World Models supports human creativity, trust, and understanding, rather than undermining them.

1.4. Relationship to previous workshops

Our proposed workshop builds on a growing body of community activity around dynamic scene understanding and 4D modelling. At CVPR 2025, the "Workshop on 4D Vision: Modeling the Dynamic World" explicitly addressed temporal, multi-view and dynamic scene modelling. Similarly, there have been workshops combining geometry, neural fields and large-scale scene generation (e.g., the "2nd Workshop on Scalable 3D Scene Generation and Geometric Learning"). For ICCV 2025, dynamic scene reconstruction and calibration from unsynchronized multi-view videos have been highlighted in the main conference sessions (e.g., "Humans as a Calibration Pattern: Dynamic 3D Scene Reconstruction from Unsynchronized and Uncalibrated Videos").

2. ORGANIZERS AND SPEAKERS

We invited a balanced composition of the organizing and speaker teams to ensure background across gender, industry and academia, geography, and career stages.

2.1. Organizers:

The workshop organizers have prior experience running other workshops in person (2024, 2023, 2022) and virtually (2020, 2021).

2.1.1. Senior Organizers

Fernando De la Torre (Carnegie Mellon University)
Fernando De la Torre is a Research Professor at Carnegie Mellon University. He has authored over 200 peer-reviewed publications in top-tier conferences and journals in computer vision and machine learning. He has served as an Associate Editor for the *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* and regularly acts as an Area Chair for ECCV, CVPR, and ICCV. Fernando founded FacioMetrics, which was acquired by Facebook in 2016. At Facebook, he led efforts in developing real-time, on-device technologies for facial feature tracking, person segmentation, and people augmentation across Messenger, Instagram, Facebook, and Portal.

Angela Dai (Technical University of Munich)
Angela Dai is an Associate Professor at the Technical University of Munich where she leads the 3D AI Lab. Angela's research focuses on understanding how real-world 3D scenes around us can be modeled and semantically understood. Previously, she received her PhD in computer science from Stanford in 2018, advised by Pat Hanrahan, and her BSE in computer science from Princeton in 2013. Her research has been recognized through an ECVA Young Researcher Award, ERC Starting Grant, Eurographics Young Researcher Award, German Pattern Recognition Award, Google Research Scholar Award, and an ACM SIGGRAPH Outstanding Doctoral Dissertation Honorable Mention.

Srinath Sridhar (Brown University)
Srinath Sridhar is the John E. Savage Assistant Professor in the Department of Computer Science at Brown University, where he leads the Interactive 3D Vision and Learning (IVL) Lab. His research focuses on 3D computer vision and artificial intelligence, with an emphasis on 4D spatiotemporal understanding of the world including objects, humans in motion, and human–object interactions. He is the recipient of an NSF CAREER Award and a Google Research Scholar Award. His work has received a Best Student Paper Award at WACV 2025 and a Best Paper Honorable Mention at Eurographics 2019. He is also an Amazon Scholar and a Visiting Faculty at the Indian Institute of Science (IISc). Srinath previously held a postdoctoral position at Stanford University with Leonidas Guibas and earned his Ph.D. from the Max Planck Institute for Informatics under Christian Theobalt and Antti Oulasvirta.

Xiaowei Zhou (Zhejiang University)
Xiaowei Zhou is a Professor in the College of Computer Science and the State Key Laboratory of CAD&CG at Zhejiang University. Before joining ZJU in 2017, he was a postdoctoral researcher at the GRASP Laboratory, University of Pennsylvania. His research interests lie in 3D computer vision, including 3D reconstruction, pose estimation, motion capture, and scene understanding, as well as their applications in mixed reality and robotics. His recent publications include *EasyVolcap*, *4K4D*, *ENeRF*, *Manhattan SDF*, *OnePose*, *LoFTR*, *SpatialTracker*, *FreeTimeGS*, *LongVolcap*, and *MapAnything*.

Lingjie Liu (University of Pennsylvania)
Lingjie Liu is an Assistant Professor in the Department of Computer and Information Science at the University of Pennsylvania, where she leads the Penn Computer Graphics Lab and is a member of the GRASP Laboratory. She was previously a Postdoctoral Research Fellow at the Max Planck Institute for Informatics and earned her Ph.D. from the University of Hong Kong in 2019. Her research lies at the intersection of

computer graphics, computer vision, and AI, with a focus on neural scene representations, neural rendering, human performance modeling, and 3D reconstruction. She is particularly interested in combining classical graphics pipelines with deep learning techniques to enable high-quality 3D reconstruction and rendering of humans and general scenes.

Aayush Prakash (Meta Reality Labs)
Aayush Prakash is an Engineering Manager leading the machine learning team within the Synthetic Data organization at Meta Reality Labs. His team focuses on problems at the intersection of machine learning, computer vision, and computer graphics, addressing challenges in domain adaptation, neural rendering, and sim-to-real learning for mixed reality. Before joining Meta, Aayush was the Head of Machine Learning at the synthetic data startup AI Reverie. Prior to that, he spent six years at NVIDIA, where his group delivered several prominent works in synthetic data generation for computer vision. He holds a B.Tech in E&ECE from the Indian Institute of Technology (IIT) Kharagpur and an MASc in Computer Engineering from the University of Waterloo.

Nikolaos Sarafianos (Meta Reality Labs)
Nikolaos Sarafianos is a Research Scientist at Meta Reality Labs working on 3D generative models that encompass textures, assets, and full virtual worlds aimed at creating immersive environments for user interaction and exploration. Previously, he worked on human-centric 3D reconstruction, dense correspondences, and neural rendering. He received the Best Paper Honorable Mention at ECCV 2022 and holds a Ph.D. from the University of Houston.

Jonathon Luiten (Meta Reality Labs)
Jonathon Luiten is Head of Volumetric 3D Video at Meta Reality Labs. His team develops large-scale systems for Gaussian Splatting capture, 3D reconstruction, VR, and dynamic scene understanding. His group's work on high-quality Gaussian Splatting in VR was recently featured at Meta Connect 2025. Jonathon's research includes *MapAnything*, *Dynamic 3D Gaussian Splatting*, *SplaTAM*, and the widely used *HOTA* tracking metric. He earned his Ph.D. from RWTH Aachen University.

2.1.2. Junior Organizers

- Chaerin Min, Aashish Rai, Angela Xing, Xiaoyan Cong, Zekun Li, Rao Fu, Yiqing Liang.

2.2. Speakers

The workshop will feature six invited speakers. Confirmed speakers include:

Peter Kontschieder (Website)
Peter Kontschieder is a Director at Reality Labs Research, Meta, where he works at the intersection of novel-view synthesis and 4D. Previously, he was a postdoctoral researcher with the Machine Intelligence and Perception Group at Microsoft Research in Cambridge, UK. In 2016, he joined Mapillary and founded Mapillary Research – Mapillary's Research Lab focusing on basic research in computer vision and machine learning. With the acquisition of Mapillary in 2020, he became a research scientist manager with Facebook He received the MSc and PhD degrees from the Graz University of Technology, Graz, Austria. He received the Marr Prize in 2015 for his contribution of Deep Neural Decision Forests, joining deep learning with decision forests.

Sara Fridovich-Keil (Website)
Sara Fridovich-Keil is an assistant professor at Georgia Tech. Her research interest lies in foundations of machine learning, particularly connections to signal processing and optimization, and applications to inverse

problems that arise in computational imaging. She was a NSF fellow at Stanford, working with Mert Pilanci and Gordon Wetzstein. She received her Ph.D. at University of California, Berkeley with Benjamin Recht. Before that, she finished her bachelor's degree at Princeton University.

Youngjoong Kwon (Website)
Youngjoong Kwon is an Assistant Professor in the Department of Computer Science at Emory University. Her research interests lie in the intersection field of Computer Vision and Computer Graphics, especially on the tasks of human performance capture, avatar modeling, neural rendering, and telepresence. Previously, she was a postdoctoral researcher at Stanford Translational AI (STAI) Lab and Stanford Vision and Learning (SVL) Lab, advised by Professor Ehsan Adeli. She was also a postdoctoral fellow at the Robotics Institute at Carnegie Mellon University, advised by Professor Fernando de la Torre. She received her Ph.D. from the Department of Computer Science at the University of North Carolina at Chapel Hill, advised by Professor Henry Fuchs.

Qianqian Wang (Website)
Qianqian Wang is an assistant professor at Hardvard Universiity and the Kempner Institute. Her research focuses on dynamic 3D world from everyday images and videos. Her long-term goal is to build intelligent systems that continualy learn from the ever-changing physical world. She was a postdoctoral researcher at UC Berkeley with Angjoo Kanazawa and Alexei A. Efros. She completed her Ph.D. at Cornell Tech with Noah Snavely and Bharath Hariharan. Before that, she received her bachelor's defree from Zhejiang University, working with Xiaowei Zhou.

Jiahui Lei (Website)
Jiahui Lei is a post-doc researcher at UC Berkeley working with Angjoo Kanazawa and Trevor Darrell. His focus is on 4D Vision for Robotics, and he has been working on 4D vision and equivariant neural networks. He obtained his Ph.D at University of Pennsylvania GRASP Lab advised Kostas Daniilidis. He was a student researcher at Google DeepMind with Leo Guibas, and received his bachelor's degree in Automation with ranking 1st/141 in Zhejiang University.

Pratul Srinivasan (Website)
Pratul Srinivasan is a research scientist at Google DeepMind in San Francisco, where he work on radiance fields research and other problems at the intersection of computer vision, graphics, and machine learning. He received his PhD from the EECS Department at UC Berkeley in 2020, where he was advised by Ren Ng and Ravi Ramamoorthi. During his PhD, he interned twice at Google Research, hosted by Jon Barron in Mountain View and Noah Snavely in New York. He graduated from Duke University in 2014, where he worked with Sina Farsiu on research problems in medical computer vision. He receive the 2025 SIGGRAPH Significant New Researcher Award, the 2021 ACM Doctoral Disseration Award Honorable Mention, the 2020 David J. Sakrison Memorial Prize, and best paper awards at ECCV 2020, ICCV 2021, and CVPR 2022.

3. FORMAT AND LOGISTICS

3.1. Format

We plan for a full-day workshop to accommodate 6 speakers, a panel discussion, and oral presentations from the accepted paper submissions. Given the recent growth of research on world models, a full-day format allows for in-depth discussion of current methods and future challenges. Through this discussion, we can effectively cover multiple areas, including reconstruction, generation, and the integration of the two. Furthermore, this structure provides opportunities for discussion from various backgrounds, from earlycareer researchers to established experts, and community engagement through accepted paper presentations.

The tentative structure is as follows:

- **First half: Invited-speaker program**
  - **Invited Talks:** Four oral sessions. Each invited talk is allocated 30 minutes, followed by a 10-minute Q&A.
- **Second half: Invited talks, Panel, Accepted papers**
  - **Invited Talks:** Two oral sessions. Each invited talk is allocated 30 minutes, followed by a 10-minute Q&A.
  - **Panel Discussion:** A 45-minute panel with all invited speakers participating.
  - **Paper Presentations:** Presentations from accepted papers.
- **Expected audience size:** Medium workshop (approximately 100–300 attendees).

Tentative Schedule (full day)

09:30–11:10	**Invited Talks (first block)**: Two invited talks (30 min + 10 min Q&A each)
11:10–11:30	Coffee break
11:30–13:10	**Invited Talks (second block)**: Two invited talks (30 min + 10 min Q&A each)
13:10–14:10	Lunch
14:10–15:50	**Invited Talks (third block)**: Two invited talks (30 min + 10 min Q&A each)
15:50–16:10	Coffee break
16:10–16:55	**Panel Discussion:** 45 min with all invited speakers
16:55–17:45	**Accepted Paper Presentations:** Short oral talks (schedule to be finalized depending on the number of papers received.)

3.2. Contingency plans

Along with the in-person workshop, a virtual mode would consist of a Zoom webinar, which runs throughout the workshop schedule. Keynote speakers and panelists will be added to this webinar as audio/video participants and will give their talks through screen sharing and camera feed. The audience will have the chance to submit questions via the Q&A interface in Zoom. The workshop program will not need to be changed if moving to a fully virtual/hybrid workshop.

3.3. Call for Papers

We invite researchers and practitioners from computer vision, computer graphics, and machine learning communities to submit original contributions to the Workshop on 4D World Models: Bridging Generation and Reconstruction, co-located with CVPR 2026. This workshop seeks papers that explore the intersection of 4D reconstruction and generative modeling, advancing toward unified 4D world models capable of capturing, representing, and synthesizing dynamic and static real-world environments. We encourage both theoretical and applied works that connect reconstruction pipelines with generative frameworks for static, dynamic, time-varying, and photorealistic scene understanding.

Topics of Interest (include but are not limited to):

1. 4D reconstruction of dynamic scenes and human activities
2. Free-viewpoint and time-varying scene generation

3. Neural representations and dynamic radiance fields
4. Generative priors for reconstruction (Diffusion, GANs, VAEs, etc.)
5. Video-based geometry learning and motion modeling
6. Multi-view dynamic capture and novel-view synthesis
7. Temporal consistency and dynamic texture generation
8. Scalable datasets, benchmarks, and evaluation protocols for 4D modeling
9. Applications in XR, robotics, and simulation

Submission Details

All submissions will undergo double-blind peer review and must follow the CVPR 2025 author guidelines. Accepted papers will be published in the IEEE/CVF CVPR 2026 Workshop Proceedings. Submissions will be handled via the OpenReview system, ensuring transparency and open feedback within the research community. We also encourage authors to submit works describing participation results or system reports from our 4D-View Challenge, which will be featured as part of the workshop's presentations and discussions.

Tentative Program Committee

Each submitted paper will undergo a thorough review process by at least three independent reviewers, overseen by area chairs. The area chairs will consist of senior workshop organizers specializing in their respective domains. We have already assembled a review committee of over 15 qualified reviewers who have committed to evaluating the submissions for this workshop. Each reviewer is a PhD student or higher (or holds an equivalent research experience).

Paper review timeline The tentative schedule of the paper submission is shown below.

February 20, 2026	Submission start
March 17, 2026	Submission deadline
March 31, 2026	Reviews released & Final decisions
April 7, 2026	Camera ready submission