2ND WORKSHOP ON WORLD MODELS: UNDERSTANDING, MODELLING AND SCALING

Mengyue Yang[1], **Xidong Feng**[2], **Nicklas Hansen**[3], **Francesco Faccio**[2], **Dima Damen**[1,2]

[1] University of Bristol
[2] Google DeepMind
[3] University of California San Diego

1 WORKSHOP SUMMARY

Over the past two years, the notion of a world model (Ha & Schmidhuber, 2018; Sutton, 1991) has undergone both conceptual clarification and significant scaling. Moving beyond a loose idea of "models that can imagine the future in the world", increasingly framed as scalable engines for modeling and simulating dynamic interactions with the world, positioned at the intersection of generative modeling, sequential decision-making, multimodal representation learning, simulation and interaction, causal robustness, and spatial intelligence. In practice, contemporary world models learn composable dynamics, often primarily from large-scale video and related multimodal data, and expose interfaces for prediction, planning, and behavior learning.

From 2024–2025, several milestones have pushed the frontier of scaled world models (Genie (Bruce et al., 2024), V-JEPA Assran et al. (2025), and Marble from World Labs etc.). Large video-pretrained models show markedly improved long-horizon rollouts and cross-frame consistency; interactive generative environments support minute-level persistence, object permanence, and promptable events; and imagination-based training for agents yields greater stability in complex object interaction and control. Together, these advances move world models from a conceptual aspiration toward a plausible systems infrastructure for general-purpose intelligence (Hafner et al., 2025; Raad et al., 2024).

To reflect this shift, the second edition organizes discussion along a pipeline: *(i) understanding and knowledge extraction* $\rightarrow$ *(ii) training and evaluation at scale* $\rightarrow$ *(iii) cross-modal and control-centric scaling*. Systems-level sessions, robotics/Open-World agent case studies, and *failure-mode post-mortems* will knit these threads together.

1.1 UNDERSTANDING THE WORLD AND EXTRACTING KNOWLEDGE

Goal: establish representations and mechanisms that support causal reasoning, temporal abstraction, and knowledge retention, feeding directly into training and evaluation.

- What constitutes **meaningful information** in partially observed environments, and how do we improve **causal/physical rules understanding** and **counterfactual reasoning** in learned models?
- What **theoretical foundations** guide the construction of World Models (e.g., sufficiency, identifiability, abstraction granularity, uncertainty quantification)?
- How should models acquire, consolidate, and update **long-horizon knowledge** (memory, schemas, and world facts) without catastrophic interference?

Transition to Training and Evaluation: The desiderata above specify what our training objectives must capture (e.g., causality, uncertainty, and temporal abstraction), and what our evaluations must measure beyond short-term fidelity.

1.2 WORLD MODEL TRAINING AND EVALUATION

Goal: examine scalable architectures, objectives, data pipelines, and rigorous evaluation protocols that stress test long-horizon competence and robustness.

- What are the strengths, limitations, and open challenges of current **modeling architectures** (Transformers, RNNs, SSMs, and hybrid latent-variable models (Micheli et al., 2022))?
- What are the trade-offs among **training algorithms**, autoregressive training (Sutskever et al., 2014), diffusion modeling (Sohl-Dickstein et al., 2015), and normalizing flows (Papamakarios et al., 2021), and when do auxiliary losses help (e.g., consistency, policy-aware rollouts)?
- How should we construct and curate **datasets** and **simulators** that expose long-horizon credit assignment, rare events, and distribution shift; and what are best practices for data governance?
- What **evaluation protocols** capture stability under rollout, sample efficiency, generalization, calibration/uncertainty, and safety (including failure taxonomy and negative-result reporting)?

Transition to Cross-Modal Scaling: The algorithms and benchmarks above establish the substrate on which we integrate language, vision, and control and evaluate end-to-end agents.

1.3 Scaling World Models Across Language, Vision, and Control

Goal: unify modeling across modalities and actions, and surface systems issues that arise when deploying agents in open-world settings.

- How can **language and vision models** be grounded for temporal reasoning, semantics, and physical consistency?
- How can World Models integrate with **control and reinforcement learning** for imagination and planning? What enables truly **multimodal world models** with coherent text–image–action semantics?
- What are the key **scaling bottlenecks** in compute, data, memory, and deployment?

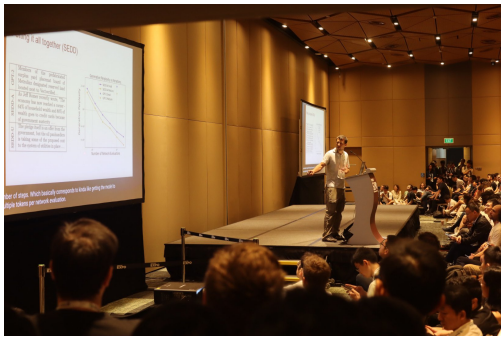

Figure 1: First edition of World Model Workshop at ICLR 2025.

1.4 Lessons and Impact from the First Edition (Context for Expansion)

The past inaugural workshop on World Models: Understanding, Modelling and Scaling at ICLR 2025 attracted **over 1,500 participants (stats from Whova)** from academia and industry, with lively discussion across keynotes, panels, and posters. This momentum evidences strong community demand and directly motivates the second edition's deeper emphasis on systems, robotics/Open-World agents, standardized evaluation, and *failure-mode post-mortems*.

Program Structure and Deliverables. The 2026 edition is more sharply focused on:

- **Systems-level sessions:** end-to-end pipelines (data→model→agent→feedback), scaling training/inference, and engineering lessons.
- **Robotics/Open-World case studies:** sim-to-real transfers, long-horizon tasks, and real-time perception–planning–control loops.
- **Failure-mode post-mortems:** negative results, brittleness under shift, safety-relevant errors, and remediation playbooks.

- **Evaluation track:** shared tasks/benchmarks for rollout stability, sample efficiency, and OOD generalization; standardized metrics and artifact checklists.

2 WORKSHOP SCHEDULE

Our one-day workshop mainly includes 3 poster sessions, 7 invited talks, 1 industry demo, an oral sessions with 6 contributed talks in total, and a panel discussion. We are glad to announce that **9 confirmed top-tier researchers as keynote speakers or panelists**. For the contributed paper sessions, regarding the recent surge in publications in related areas and the success of the previous workshop, we project over 150 paper submissions and over 1,500 participants.

2.1 **CONFIRMED** INVITED SPEAKERS AND PANELISTS

Jürgen Schmidhuber (**Confirmed**) is the Scientific Director of Swiss AI Lab, IDSIA and the co-chair of the Center of Excellence in Generative AI at KAUST. The main goal of Professor Jürgen Schmidhuber has been to build a self-improving Artificial Intelligence (AI) smarter than himself. His lab's Deep Learning Neural Networks (NNs) based on ideas published in the "Annus Mirabilis" 1990-1991 have revolutionised machine learning and AI. He was one of the first to work on LSTM, feedforward NNs on GPUs, DanNet, deep NN for medical imaging, GANs, Transformers. His research group also established the fields of mathematically rigorous universal AI and recursive self-improvement in meta-learning machines that learn to learn (since 1987). He also generalized algorithmic information theory and the many-worlds theory of physics. He is recipient of numerous awards, author of about 400 peer-reviewed papers, and Chief Scientist of the company NNAISENSE, which aims at building the first practical general purpose AI.

Sergey Levine (**Confirmed**) is Associate Professor in UC Berkeley, EECS. His work focuses on machine learning for decision making and control, with an emphasis on deep learning and reinforcement learning algorithms. Applications of his work include autonomous robots and vehicles, as well as applications in other decision-making domains. His research includes developing algorithms for end-to-end training of deep neural network policies that combine perception and control, scalable algorithms for inverse reinforcement learning, deep reinforcement learning algorithms, and more. `https://people.eecs.berkeley.edu/~svlevine/`

Jiajun Wu (**Confirmed**) is an Assistant Professor of Computer Science at Stanford University, working on computer vision, machine learning, and computational cognitive science. Dr. Wu is affiliated with the Stanford Vision and Learning Lab (SVL) and the Stanford AI Lab (SAIL) and studies machine perception, reasoning, and its interaction with the physical world, drawing inspiration from human cognition. Wu's research has been recognized through the ACM Doctoral Dissertation Award Honorable Mention, the AAAI/ACM SIGAI Doctoral Dissertation Award, the MIT George M. Sprowls PhD Thesis Award in Artificial Intelligence and Decision-Making, the 2020 Samsung AI Researcher of the Year, the IROS Best Paper Award on Cognitive Robotics, and fellowships from Facebook, Nvidia, Samsung, and Adobe. `https://jiajunwu.com/`

Danijar Hafner (**Confirmed**) is a staff Research Scientist at Google DeepMind. He received his PhD at the University of Toronto with Jimmy Ba, was a visiting student at UC Berkeley with Pieter Abbeel, was a Vanier Scholar, and interned at Google Brain for many years. He completed his MRes at UCL and the Gatsby Unit with Tim Lillicrap and Karl Friston. Danijar's research aims at building generally intelligent machines that understand and interact with the world. `https://danijar.com/#short-biography`

Mido Assran (**Confirmed**) is a Research Scientist at Meta in Fundamental AI Research (FAIR). Previously, Mido obtained his PhD in Electrical and Computer Engineering from McGill University and Mila, the Quebec AI Institute, where he was primarily advised by Michael Rabbat. His current research focuses on advancing self-supervised representation learning and low-shot prediction. In the past, Mido worked on parallelizing deep reinforcement learning, and large-scale optimization. His research on representation learning has been featured in several media outlets, including VentureBeat, TechCrunch, and SiliconANGLE. He was also fortunate to have a featured profile piece in the ICCV daily magazine. `https://www.midoassran.ca/`

Table 1: Tentative schedule (subject to change).

Morning		**Afternoon**	
Time	**Planned Event**	**Time**	**Planned Event**
08:15 am - 08:45 am	Poster Session I	13:00–13:30	Invited Talk 5 + Q&A
08:45 am - 09:00 am	Opening Remarks	13:30–14:00	Invited Talk 6 + Q&A
09:00 am - 09:30 am	Invited talk 1 + Q&A	14:00–15:00	Panel Discussion
09:30 am - 10:00 am	Invited talk 2 + Q&A	15:00–16:00	Poster Session III
10:00 am - 10:30 am	Poster Session II	16:00–16:30	Invited Talk 7 + Q&A
10:30 am - 11:00 am	Industry demo + Q&A	16:30–17:30	Contributed Talks (×6)
11:00 am - 11:30 am	Invited talk 3 + Q&A	17:30–17:45	Awarding
11:30 am - 12:00 am	Invited talk 4 + Q&A	17:45–18:00	Closing Remarks

Katja Hofmann (**Confirmed**) is a Principal Researcher and lead of Game Intelligence at Microsoft Research Cambridge. She and her team drive research into new AI capabilities, enabled by the rich data and environments provided by modern video games. Her long-term goal is to develop AI systems that learn to collaborate with people, to empower their users and help solve complex real-world problems. https://www.microsoft.com/en-us/research/people/kahofman/

Hao Su (**Confirmed**) is an Associate Professor of Computer Science at the University of California, San Diego (UCSD), and also serves as the Founder and Chief Technology Officer of Hillbot, an intelligent robotics startup. His research focuses on developing algorithms to simulate, understand, and interact with the physical world. His interests span computer vision, machine learning, computer graphics, and robotics, with extensive publications and teaching experience in these fields. He developed widely-used datasets and software, such as ImageNet, ShapeNet, PointNet, PartNet, SAPIEN, and the recent ManiSkill. These works have greatly advanced research progress in areas such as 3D vision and robot manipulation. Su serves as the Program Chair of CVPR 2025. He has been awarded the SIGGRAPH Best Doctoral Dissertation Honorable Mention and the NSF CAREER Award. https://cseweb.ucsd.edu/~haosu/index.html

Noor Sajid (**Confirmed**) is a Kempner Research Fellow at Harvard University and a Guest Scientist at the Max Planck Institute for Biological Cybernetics. She completed her PhD at University College London under Karl Friston and a postdoctoral fellowship with Peter Dayan at the Max Planck Institute. Her research develops probabilistic world models to explain how biological systems predict, generalize, and adapt to changing structures, bridging neuroscience and machine learning. Noor's recent work advances biologically inspired world models with temporal abstraction and structured context access (NeurIPS 2020; Nat Mach Intell 2023; ICLR 2024). She has delivered invited talks at major venues including NeurIPS, OHBM, MILA, and Oxford, reflecting her active engagement across neuroscience and AI communities. https://www.noorsajid.com/

Genie3 team (Confirmed): Genie3 is a general purpose world model that can generate an unprecedented diversity of interactive environments, from Google DeepMind. https://deepmind.google/discover/blog/genie-3-a-new-frontier-for-world-models/

Importantly, our team is fortunate to have **Jürgen Schmidhuber, one of the founders of the World Models concept**, as our keynote speaker. In addition, we have invited leading researchers behind several recent, groundbreaking foundation world model projects that have attracted significant attention across the AI community, including **Danijar Hafner (Dreamer4 (Hafner et al., 2025), Google DeepMind)**, **Mido Assran (V-JEPA2 (Assran et al., 2025), Meta FAIR)**, **Jiajun Wu (World Labs)** and speakers from **Genie3 Team (Google DeepMind)** as our keynote speakers. They are expected to present the latest advancements and insights from their respective teams.

3 Organizing Team

Our team includes researchers specializing in World Models, causality, Reinforcement Learning, robotics, and video understanding, all of which are closely aligned with the workshop's theme. Our team not only brings a wealth of experience across various research domains but also showcases remarkable diversity in backgrounds, and genders. With 5 members from 3 different institutions, we are well-positioned to integrate a broad range of perspectives and expertise into the workshop. None

Table 2: Key dates aligned with ICLR 2026 workshop requirements. We follow the official ICLR 2026 guidance of workshops. All times are AoE.

Date	Milestone
3 Dec 2025	OpenReview site launches; CFP posted on iclr.cc, social media, and mailing lists.
12 Jan 2026	First mentoring clinic and checklist walkthrough (recording shared for asynchronous access).
30 Jan 2026	Recommended submission date so reviewers receive full papers before the official discussion window.
7 Feb 2026	Final submission deadline (AoE) with artifacts and governance checklist uploaded.
14 Feb 2026	Reviews due; ethics and safety flagging window opens for organisers.
21 Feb 2026	Author rebuttal and discussion close; meta-reviewers consolidate decisions.
24 Feb 2026	Final decisions signed off internally to allow notification preparation.
1 Mar 2026	Authors notified and accepted papers made public on OpenReview, matching the official ICLR requirement.
3 Mar 2026	Camera-ready packages (paper, operator card, artifact links) due for proceedings layout.
11 Mar 2026	Metadata and PDFs imported into iclr.cc and the central ICLR schedule, satisfying the compliance checklist.

of the organizers will serve as speakers in this workshop. Every team member organizes less than or equal to one workshop in ICLR 2026, and is fully committed to going to the venue in person.

Our team members have experience in organizing workshops, making us well-prepared to handle workshops organization at large conferences. The organized important events of our team are listed.

- Embodied World Models at NeurIPS 2025;
- Program co-chair of European Workshop of Reinforcement Learning;
- World Models: Understanding, Modelling and Scaling at ICLR 2025;
- Reasoning and Planning for Large Language Models Workshop at ICLR 2025;
- Artificial Intelligence with Causal Techniques Workshop at AAAI 2025;
- Causality and Large Model Workshop at NeurIPS 2024;
- Causal Representation Learning Workshop at NeurIPS 2024;
- Artificial Intelligence with Causal Techniques Workshop at AAAI 2025;
- Causal Structure Learning from Event Sequences and Prior Knowledge Competition at NeurIPS 2023;
- Causal Representation Learning Workshop at ICDM 2024;
- Universal Learning Algorithms and Optimal Search Workshop at NeurIPS 2022;
- Our senior organizer Dima Damen served as Program Chair of ICCV 2021.

3.1 Organizer Introductions

Mengyue Yang is a Lecturer (equals to US Assistant Professor) in AI at University of Bristol. She obtained her Ph.D. from University College London, under the supervision of Professor Jun Wang. Her research interests are causality, reinforcement learning and world models. She has been recognized as a Rising Star in AI by KAUST. She has extensive experience in organizing workshops at conferences. She was the organizer of the first World Model: Understanding, Modelling and Scaling workshop at ICLR 2025, co-organizer of the Causal Structure Learning Competition at NeurIPS 2023, Causal Representation Learning Workshop at ICDM 2024, Causality and Large Model Workshop at NeurIPS 2024 and Artificial Intelligence with Causal Techniques Workshop at AAAI 2025.

Xidong Feng is a research scientist at Google DeepMind Discovery team. His research spans over Large Langauge Model, Reinforcement Learning, and Multi-agent Learning. He has published over 10 papers in top AI conferences or journals like NeurIPS, ICML and JMLR. Previously, he obtained his Ph.D. at Computer Science, University College London, advised by Prof. Jun Wang. He co-organized the first workshop on World Models: Understanding, Modelling and Scaling; and Reasoning and Planning for Large Language Models Workshop at ICLR 2025.

Nicklas Hansen is a PhD candidate at UC San Diego advised by Professors Xiaolong Wang and Hao Su. Their research focuses on developing generalist AI agents that learn from physical interaction, with a particular focus on world models and model-based reinforcement learning. Nick has spent time at NVIDIA Research, Meta AI, and UC Berkeley, and received their BS and MS degrees from Technical University of Denmark. They are a recipient of the 2024 NVIDIA Graduate Fellowship, have authored more than 20 publications at top venues in machine learning and robotics such as ICLR, NeurIPS, and ICRA, and they co-organized the Embodied World Models for Decision-Making workshop at NeurIPS 2025.

Francesco Faccio is Senior Research Scientist at Google DeepMind. Before this, he completed a PhD in IDSIA mentored by Prof. Jürgen Schmidhuber. Before his PhD, he earned bachelor's and master's degrees in Mathematical Engineering at Politecnico di Milano. His research interests lie at the intersection of General AI, Reinforcement Learning, and AI for Science, focusing on developing artificial scientists to automate scientific research using artificial curiosity and meta-learning techniques. He has authored 11 publications in top-tier AI conferences and journals, with four receiving oral presentations at NeurIPS, ICML, and AAAI, and one winning the Best Paper Award at the NeurIPS Ro-FoMo Workshop. He also co-organized the KAUST Rising Stars in AI Symposium in 2023 and 2024 and served as program co-chair of the European Workshop of Reinforcement Learning.

Dima Damen is a Professor of Computer Vision at the University of Bristol and Senior Research Scientist at Google DeepMind. Dima is currently an EPSRC Fellow (2020-2026), focusing her research interests in the automatic understanding of object interactions, actions and activities using wearable visual (and depth) sensors. She is best known for her leading works in Egocentric Vision, and has also contributed to novel research questions including mono-to-3D, video object segmentation, assessing action completion, domain adaptation, skill/expertise determination from video sequences, discovering task-relevant objects, dual-domain and dual-time learning as well as multi-modal fusion using vision, audio and language. She is the project lead for EPIC-KITCHENS, the seminal dataset in egocentric vision, with accompanying open challenges and follow-up works: EPIC-Sounds, VISOR and EPIC Fields, as well as the recent HD-EPIC. She is part of the large-scale consortium effort Ego4D and Ego-Exo4D. She is an ELLIS Fellow, associate editor (AE) of IJCV, and was a program chair for ICCV 2021 and Associate Editor-in-Chief (AEIC) of IEEE TPAMI (2023-2025).

3.2 Program Committee

Building on the success of last year's edition, which attracted **125 qualified reviewers** from both academia and industry, we have further strengthened our reviewer network through collaboration among seven participating institutions. This collective effort ensures that each submission will receive at least two high-quality reviews, and no reviewer will be assigned more than three papers. We are pleased to report that **43 reviewers have already accepted the invitation** through an internal sign-up process across three partner institutions. While this provides a strong starting point, we continue to recruit additional experts to maintain a fair, balanced, and comprehensive review process. Several members of the organizing committee will serve as area chairs, leveraging their expertise in related subfields to oversee and coordinate the reviewing efficiently.

To ensure integrity and fairness, we will strictly manage **conflicts of interest** in line with ICLR's policies: no organizer or reviewer will assess submissions from their own institution or close collaborators. All conflicts will be identified and resolved prior to reviewer assignment, and the area chairs will verify that each paper receives independent, unbiased evaluations.

4 Inclusion and Diversity

To account for underrepresented researchers who may not have access to large amounts of compute, we accept research papers pages, tiny paper, in which the working-in-progress submissions with proof-of-concept demonstrations with live demos are welcome. Meanwhile, to incentivize participants to submit new and unique work, our workshops will not accept the work that has already been published at other machine learning conferences (including the main ICLR conference). We will have 3 poster presentation sessions during the workshop, giving authors and attendees plenty of time to discuss

ideas. These sessions will help authors improve their work and give attendees a chance to see new perspectives. We will set a paper travel grant give support to the unrepresentative group of students.

The organizing committee presents the diversity in the following aspects.

- **Gender**: *female* (Mengyue, Dima), *non-binary* (Nicklas), *male* (Xidong, Francesco).
- **Seniority**: *Professor* (Dima), *Assistant Professor* (Mengyue), *PhD Student* (Nicklas), and *Research Scientist* (Xidong and Francesco).
- **Institutional affiliation**: *Academia* (Mengyue, Dima, Nicklas) and *industry* (Xidong and Francesco at Google DeepMind).
- **New member compared with previous edition:** (Nicklas, Dima).

The collection of invited speakers presents diversity in the following aspects.

- **Gender**: *female* (Katja, Noor), *male* (Jürgen, Sergey, Jiajun, Danijar, Mido, Hao).
- **Seniority**: *Professor* (Jürgen), *Associate Professor* (Sergey, Hao), *Assistant Professor* (Jiajun), *Postdoc Fellow* (Noor), *Senior Research Scientist* (Danijar, Katja), and *Research Scientist* (Mido).
- **Institutional affiliation**: *Academia* (Jürgen, Sergey, Hao, Jiajun, Noor) and *industry* (Danijar, Mido, Katja).
- **New member compared with previous edition:** (No overlaping of speakers).

5 Other Statements

5.1 Submission guidelines

Our Call for Papers (CFP) is synchronized with the ICLR 2026 schedule, while providing contributors ample time to prepare their submission. Submissions open on 3 December 2025. The workshop welcomes submissions in three complementary categories: **Research Papers** (up to 8 pages + references) and **Tiny Papers / Practice Notes** (2–4 pages).

This **Tiny Paper track** aims to make the workshop more inclusive and accessible to contributors outside the traditional ML publication circuit, encouraging early-stage, self-contained work that stimulates discussion. In line with ICLR guidelines, AI-generated papers are not permitted in this track; AI assistance is allowed, but all submissions must be primarily human-authored and reflect original thought and analysis. To support new research teams, we will organize office-hour clinics in December and mid-January focused on artifact curation and safety documentation. Our internal review timeline follows the official ICLR 2026 milestones, as shown in Table 2.

5.2 LLM Usage Policy

We will follow the official ICLR 2026 Policies on Large Language Model (LLM) Usage and ensure transparent, responsible, and human-supervised use of AI tools throughout the workshop process. Workshop participants and organizers may use established LLM tools (e.g., GPT-4, Claude, Gemini, or open-source equivalents) for language editing, idea structuring, or code documentation, provided such use is disclosed. No automated system will independently generate or submit research papers or reviews. AI tools may assist in writing refinement, reference formatting, or discussion summarization, but may not act as a primary author, reviewer, or autonomous decision-maker. All submissions and reviews will remain fully human-authored and validated. Organizers will oversee AI-assisted content for factual accuracy, originality, and compliance with ICLR's authorship standards.

5.3 Online Resources

Given the rapid growth of ICLR and uncertainty in venue capacity, we will strengthen online components to ensure inclusive participation. All talks will be recorded and shared online, with presentation titles, abstracts, and accepted materials (papers, posters, demos) published on the workshop website. A dedicated online platform will host videos and discussion materials as a lasting archive. While **the workshop will be in person**, online participation will be supported if needed to ensure smooth communication and accessibility for all contributors.

Post-workshop Summary Article: Following the second workshop, we plan to prepare a comprehensive post-workshop summary article that synthesizes insights from both editions of the event. This article will consolidate key discussions, highlight the field's developmental trajectory over the past two years, and identify emerging research frontiers and future directions. The organizing team will coordinate with contributing authors from both years to co-author this synthesis piece, which we intend to submit to a leading venue or publish as part of a special issue dedicated to advancing research on world models to summarize the current development and potential future trends.

5.4 Sponsor and Promotion

Upon acceptance, we will actively seek sponsorship from companies such as Apple, Meta, and Google to enhance the accessibility and inclusivity of the workshop. In particular, we plan to cover the registration fees of student participants, prioritizing members from under-represented groups. We will also promote the workshop widely across major academic and industry channels to increase visibility and attract additional sponsorship and community engagement.

5.5 Relevance to Previous Event

The first Edition of our workshop World Models: Understanding, Modelling and Scaling, at ICLR 2025, is the first comprehensive discussion on world models. Following events included ICML 2025 Workshop on Assessing World Models, ICML 2025 Building Physically Plausible World Models, the NeurIPS 2025 Embodied World Models focus world models in embodied AI and robotics, NeurIPS 2025 Workshop on Scaling Environments for Agents focus on world models with LLMS. Previous workshops have focused on specific areas, such as robotics, the agent problem and language models or only assessing it. In contrast, our workshop aims to present a more comprehensive discussion on world models. We include the widest range of discussion on world models, such as classical world model methods, and initiate broader discussions on the evolution of world models, such as multimodal modelling, generative AI and their integration with more general application areas. Our goal is to connect AI researchers widely and encourage a holistic exchange.

Selected References

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, et al. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025.

Jake Bruce, Michael D Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, et al. Genie: Generative interactive environments. In *Forty-first International Conference on Machine Learning*, 2024.

David Ha and Jürgen Schmidhuber. World models. *Advances in Neural Information Processing Systems*, 2018.

Danijar Hafner, Wilson Yan, and Timothy Lillicrap. Training agents inside of scalable world models. *arXiv preprint arXiv:2509.24527*, 2025.

Vincent Micheli, Eloi Alonso, and François Fleuret. Transformers are sample-efficient world models. *arXiv preprint arXiv:2209.00588*, 2022.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

Maria Abi Raad, Arun Ahuja, Catarina Barros, Frederic Besse, Andrew Bolt, Adrian Bolton, Bethanie Brownfield, Gavin Buttimore, Max Cant, Sarah Chakera, et al. Scaling instructable agents across many simulated worlds. *arXiv preprint arXiv:2404.10179*, 2024.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. PMLR, 2015.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*, 2014.

Richard S Sutton. Dyna, an integrated architecture for learning, planning, and reacting. *ACM Sigart Bulletin*, 2(4):160–163, 1991.