December 2–5, 2025 | San Diego Convention Center

NVIDIA Experiences at NeurIPS 2025: **Expo Hall, Conference Talks, & Workshop Talks**

<table>
<tr><th>MONDAY
12/1</th><th>TUESDAY
12/2</th><th>WEDNESDAY
12/3</th><th>THURSDAY
12/4</th><th>FRIDAY
12/5</th><th>SATURDAY
12/6</th><th>SUNDAY
12/7</th></tr>
<tr><td>4 -8 PM | Omni Hotel Gallery 2 | 3
NVIDIA Nemotron Summit</td><td>12 -1 PM | Expo Hall Booth #1139
NVIDIA Cosmos: A World Model Platform for Physical AI</td><td>9 - 10 AM | Expo Hall Booth #1139
NVIDIA Cosmos: A World Model Platform for Physical AI</td><td>8:30 - 9:30 AM | Exhibit Hall F, G, H
NeurIPS Invited Talk: The Art of Artificial Reasoning
Yejin Choi</td><td>9 - 10 AM | Expo Hall Booth #1139
Ship Your Stack: How to Share and Scale AI Projects with Brev Launchables</td><td>9 - 9:30 AM | Upper Level Room 30A-E
Embodied World Models for Decision-Making
Sanja Fidler</td><td>2:20 - 2:55 PM | Upper Level Room 7
Frontiers in Probabilistic Inference: Sampling Meets Learning
Karsten Kreis</td></tr>
<tr><td></td><td>1 -1:30 PM | Expo Hall Booth #1139
CuFOLIO: GPU-Acceleration Portfolio Optimization</td><td>10 - 10:30 AM | Expo Hall Booth #1139
Object Manipulation Data Capture</td><td>9 - 9:30 AM | Expo Hall Booth #1139
Object Manipulation Data Capture</td><td>10 - 11 AM | Expo Hall Booth #1139
NVIDIA Cosmos: A World Model Platform For Physical AI</td><td>9 - 9:30 AM | Upper Level Room 3
NeurIPS 2025 Workshop on Generative AI in Finance
Ioana Boier</td><td></td></tr>
<tr><td></td><td>1:30 - 2 PM | Expo Hall Booth #1139
AI Model Distillation for Financial Datasets</td><td>10:30 -11:30 AM | Expo Hall Booth #1139
Real-Time 480p Video Generation with FastVideo and SGLang On Brev</td><td>9:30 - 10:30 AM | Expo Hall Booth #1139
Ship Your Stack: How to Share and Scale AI Projects with Brev Launchables</td><td>11 -12:30 PM | Omni Hotel Gallery 2 | 3
Launching Legends Workshop</td><td>1:30 - 2 PM | Room 20C
Structured Probabilistic Inference and Generative Modeling
Arash Vahdat</td><td></td></tr>
<tr><td></td><td>2 - 3 PM | Expo Hall Booth #1139
Identifying and Transcribing Speakers in Real-Time</td><td>11:30 - 12:30 PM | Expo Hall Booth #1139
GPU-Accelerated Financial Data Pipeline in Action</td><td>10:30 -11 AM | Expo Hall Booth #1139
Nemo-Evaluator: Open Models Need Open Evaluations</td><td>11 - 11:30 AM | Expo Hall Booth #1139
AI Model Distillation for Financial Datasets</td><td>6 -10 PM | Upper Level Room 23ABC
Generative and Protective AI for Content Creation
Frank Wang</td><td></td></tr>
<tr><td></td><td>3 - 4 PM | Expo Hall Booth #1139
Medical Reasoning Demo</td><td>12:30 - 1:30 PM | Expo Hall Booth #1139
Training and Fine-Tuning on DGX Spark</td><td>11 -11:30 AM | Expo Hall Booth #1139
CuFOLIO: GPU-Acceleration Portfolio Optimization</td><td>11:30 - 12 PM | Expo Hall Booth #1139
NVIDIA Dynamo - Consolidated Exemplar</td><td></td><td></td></tr>
<tr><td></td><td>4 - 5 PM | Expo Hall Booth #1139
Structure Prediction Demo</td><td>1:30 -2:00 PM | Expo Hall Booth #1139
Accelerated Image And Video Generation With Sglang Diffusion</td><td>11:30 - 12:30 PM | Expo Hall Booth #1139
NVIDIA Cosmos: A World Model Platform For Physical AI</td><td>3:30 -6 PM | The New Children's Museum
AI for Science: Algorithms to Atoms
Bill Dally</td><td></td><td></td></tr>
<tr><td></td><td>5 - 6 PM | Expo Hall Booth #1139
Ship Your Stack: How to Share and Scale AI Projects with Brev Launchables</td><td>2 -2:30 PM | Expo Hall Booth #1139
NeMo Evaluator: Open Models Need Open Evaluations</td><td>6 -10 PM | Expo Hall Booth #1139
Inference With vLLM and PyTorch on DGX Spark</td><td></td><td></td><td></td></tr>
<tr><td></td><td>5:45 - 6 PM | Expo Hall
The State of Open Source AI and a Look Into the Future
Nader Kahill</td><td>2:30 - 4 PM | Expo Hall Booth #1139
Structure Prediction</td><td>1:30 - 2:30 PM | Expo Hall Booth #1139
CuFOLIO: GPU-Acceleration Portfolio Optimization</td><td></td><td></td><td></td></tr>
<tr><td></td><td>6 - 7 PM | Expo Hall Booth #1139
ProtoMotion</td><td>6 -10 PM | Expo Hall Booth #1139
Medical Reasoning</td><td>2:30 -3:30 PM | Expo Hall Booth #1139
AI Model Distillation for Financial Datasets</td><td></td><td></td><td></td></tr>
<tr><td></td><td>7 - 8 PM | Expo Hall Booth #1139
NVIDIA Cosmos: A World Model Platform for Physical AI</td><td>4:30 - 5:30 PM | Upper Level Ballroom 6AB
Cosmos World Foundation Model Platform for Physical AI
Ming-Yu Liu</td><td>3 - 4:30 PM | Expo Hall Booth #1139
Structure Prediction</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td>6 -10 PM | Puesto, W Harbor Dr Unit 155
NVIDIA Researcher Reception</td><td>4:30 - 5:00 PM | Expo Hall Booth #1139
Medical Reasoning</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td>5 -6:30 PM | Omni Hotel Gallery 2|3
Launching Legends Reception</td><td></td><td></td><td></td></tr>
</table>

NVIDIA | NEURAL INFORMATION PROCESSING SYSTEMS

December 2–5, 2025 | San Diego Convention Center

NVIDIA Experiences at NeurIPS 2025: **Posters**

WEDNESDAY 12/3		THURSDAY 12/4		FRIDAY 12/5	
11 AM - 2 PM	4:30 - 7:30 PM	11 AM - 2 PM	4:30 - 7:30 PM	11 AM - 2 PM	4:30 - 7:30 PM
Jet-Nemotron: Efficient Language Model With Post-Neural Architecture Search	Efficient Hybrid Language Model Compression Through Group-Aware SSM Pruning	Nemotron-Flash: Towards Latency-Optimal Hybrid Small Language Models	Audio Flamingo 3: Advancing Audio Intelligence With Fully Open Large Audio Language Models	ProRL: Prolonged Reinforcement Learning Expands Reasoning Boundaries in Large Language Models	Beyond Token Probes: Hallucination Detection via Activation Tensors With ACT-ViT
RobotSmith: Generative Robotic Tool Design for Acquisition of Complex Manipulation Skills	Unified Reinforcement and Imitation Learning for Vision-Language Models	FFN Fusion: Rethinking Sequential Computation in Large Language Models	CLIMB: CLustering-Based Iterative Data Mixture Bootstrapping for Language Model Pretraining	Scaling Up Reasoning to Long Videos in VLMs	Attention on the Sphere
Seeing What Matters: Generalizable AI-Generated Video Detection With Forensic-Oriented Augmentation	Win Fast or Lose Slow: Balancing Speed and Accuracy in Latency-Sensitive Decisions of LLMs	ReliabilityRAG: Effective and Provably Robust Defense for RAG-Based Web Search	GSPN-2: Efficient Parallel Sequence Modeling	Diversifying Parallel Ergodic Search: A Signature Kernel Evolution Strategy	ThinkAct: Vision-Language-Action Reasoning via Reinforced Visual Latent Planning
Efficient Part-Level 3D Object Generation via Dual Volume Packing	GradMetaNet: An Equivariant Architecture for Learning on Gradients	Image-Based Gaussian Splatting	Align Your Flow: Scaling Continuous-Time Flow Map Distillation	Towards Predicting Any Human Trajectory In Context	Trust Region Constrained Measure Transport in Path Space for Stochastic Optimal Control and Inference
WorldModelBench: Judging Video Generation Models as World Models	LuxDiT: Lighting Estimation with Video Diffusion Transformer	Policy Optimized Text-to-Image Pipeline Design	BlurDM: A Blur Diffusion Model for Image Deblurring	Elucidated Rolling Diffusion Models for Probabilistic Forecasting of Complex Dynamics	Sparse VideoGen2: Accelerate Video Generation With Sparse Attention via Semantic-Aware Permutation
VAGEN: Reinforcing World Model Reasoning for Multi-Turn VLM Agents	Reliable World Simulation for Autonomous Driving	UniRelight: Learning Joint Decomposition and Synthesis for Video Relighting	Rectified Point Flow: Generic Point Cloud Pose Estimation	PanTS: The Pancreatic Tumor Segmentation Dataset	AneuG-Flow: A Large-Scale Synthetic Dataset of Diverse Intracranial Aneurysm Geometries and Hemodynamics
Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond)	Model-Based Policy Adaptation for Closed-Loop End-to-End Autonomous Driving	Eagle 2.5: Boosting Long-Context Post-Training for Frontier Vision-Language Models	Feed-Forward Bullet-Time Reconstruction of Dynamic Scenes From Monocular Videos	Scaling RL to Long Videos	Twilight: Adaptive Attention Sparsity With Hierarchical Top-p Pruning
Prismatic Synthesis: Gradient-Based Data Diversification Boosts Generalization in LLM Reasoning	Unmasking Puppeteers: Leveraging Biometric Leakage to Expose Impersonation in AI-Based Videoconferencing	IBGS: Image-Based Gaussian Splatting	Martian World Model: Controllable Video Synthesis With Physically Accurate 3D Reconstructions	Test-Time Scaling of Diffusion Models via Noise Trajectory Search	SutureBot: A Precision Framework and Benchmark for Autonomous End-to-End Suturing
	Why and How LLMs Hallucinate: Connecting the Dots With Subsequence Associations		Radial Attention: $\mathcal O(n \log n)$ Sparse Attention for Long Video Generation		
	HelpSteer3-Preference: Open Human-Annotated Preference Data Across Diverse Tasks and Languages		Machine Unlearning Doesn't Do What You Think: Lessons for Generative AI Policy and Research		
	Better Tokens for Better 3D: Advancing Vision-Language Modeling in 3D Medical Imaging		Broken Tokens? Your Language Model Can Secretly Handle Non-Canonical Tokenizations		
			AI Debate Aids Assessment of Controversial Claims		
			Generalizable Domain Adaptation for Sim-and-Real Policy Co-Training		
			AceReason: Advancing Math and Code Reasoning Through Reinforcement Learning		
			Frame Context Packing and Drift Prevention in Next-Frame-Prediction Video Diffusion Models		