

FINAL

**13th Annual International Conference on Learning Representations (ICLR)  
2026 Fact Sheet****Global Participation**

- 9,954 participants spanning 80 countries
- 7,054 in-person
- 2,900 virtual
- List of countries with over 200 participants:
  1. United States: 2,375
  2. China: 1,804
  3. United Kingdom: 523
  4. South Korea: 472
  5. Germany: 406
  6. Brazil: 377
  7. Canada: 345
  8. Switzerland: 226

**Previous ICLR Locations and No. of Participants**

- 2025: Singapore (Republic of Singapore) 11,137 participants from 85 countries
- 2024: Vienna (Austria), 6,533 participants from 79 countries
- 2023: Kigali, Rwanda (Africa) 3,758 participants from 73 countries
- 2022: Virtual (Global) 5,200 participants from 81 countries
- 2021: Virtual (Global) 6,300 participants from 64 countries
- 2020: Virtual (Global) 5,600 participants from 76 countries
- 2019: New Orleans (USA) 2,600 participants from 50 countries
- 2018: Vancouver (Canada) 1,950 participants from 38 countries
- 2017: Toulon (France)
- 2016: San Juan (Puerto Rico)
- 2015: San Diego (USA)
- 2014: Banff (Canada)
- 2013: Scottsdale (USA)

**Program Committee Statistics**

- 19,525 submissions, 5357 accepted (11,603 submissions/3704 accepted in 2025)
  - 27% acceptance rate (last year: 32%)
  - 7,922 more submissions, 1,653 more accepted compared to last year
- 21,674 reviewers (18,325 in 2025)
- 1,634 area chairs (823 in 2025)
- 79 senior area chairs (71 in 2025)
- [Retrospective on ICLR 2026 Review Process](#)
  
- 6 [Invited Talks](#)
- 224 [Oral Presentations](#)
- 23 [Socials](#)



- 40 [Workshops](#) (122 Workshop submissions)
- 3 [Affinity events](#) –
  - [Women in Machine Learning](#) Social
  - [Queer in AI](#) Social
  - [Humanists in AI](#) Social
- 58 JMLR Posters
- 58 [Blogpost Track Posters](#)
- 6 Mentorship Round-table Q&A's with 16 mentors (3 mentors/session)

## Test of Time Winners: [Announcement](#) / ICLR [Presentation](#)

### [Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks](#)

*Alec Radford, Luke Metz, Soumith Chintala*

Abstract: In recent years, supervised learning with convolutional networks (CNNs) has seen huge adoption in computer vision applications. Comparatively, unsupervised learning with CNNs has received less attention. In this work we hope to help bridge the gap between the success of CNNs for supervised learning and unsupervised learning. We introduce a class of CNNs called deep convolutional generative adversarial networks (DCGANs), that have certain architectural constraints, and demonstrate that they are a strong candidate for unsupervised learning. Training on various image datasets, we show convincing evidence that our deep convolutional adversarial pair learns a hierarchy of representations from object parts to scenes in both the generator and discriminator. Additionally, we use the learned features for novel tasks - demonstrating their applicability as general image representations.

### **Continuous control with deep reinforcement learning** (paper download in [Announcement](#))

*Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, Daan Wierstra*

Abstract: We adapt the ideas underlying the success of Deep Q-Learning to the continuous action domain. We present an actor-critic, model-free algorithm based on the deterministic policy gradient that can operate over continuous action spaces. Using the same learning algorithm, network architecture and hyper-parameters, our algorithm robustly solves more than 20 simulated physics tasks, including classic problems such as cartpole swing-up, dexterous manipulation, legged locomotion and car driving. Our algorithm is able to find policies whose performance is competitive with those found by a planning algorithm with full access to the dynamics of the domain and its derivatives. We further demonstrate that for many of the tasks the algorithm can learn policies “end-to-end”: directly from raw pixel inputs.

## Outstanding Paper Winners: [Announcement](#)

### [Transformers are Inherently Succinct](#)

*Pascal Bergstrauser, Ryan Cotterell, Anthony Widjaja Lin*

Abstract: We propose succinctness as a measure of expressive power of a transformer in describing a concept. To this end, we prove that transformers are highly expressive in that they can represent formal languages substantially more succinctly than standard representations of formal languages like finite automata and Linear Temporal Logic (LTL) formulas. As a by-product of this expressivity, verifying even simple properties of transformers is shown to be provably intractable (i.e. EXPSPACE-complete).

[LLMs Get Lost In Multi-Turn Conversation](#)

*Philippe Laban, Hiroaki Hayashi, Yingbo Zhou, Jennifer Neville*

**Abstract.** Large Language Models (LLMs) are conversational interfaces. As such, LLMs have the potential to assist their users not only when they can fully specify the task at hand, but also to help them define, explore, and refine what they need through multi-turn conversational exchange. Although analysis of LLM conversation logs has confirmed that underspecification occurs frequently in user instructions, LLM evaluation has predominantly focused on the single-turn, fully-specified instruction setting. In this work, we perform large-scale simulation experiments to compare LLM performance in single- and multi-turn settings. Our experiments confirm that all the top open- and closed-weight LLMs we test exhibit significantly lower performance in multi-turn conversations than single-turn, with an average drop of 39% across six generation tasks. Analysis of 200,000+ simulated conversations decomposes the performance degradation into two components: a minor loss in aptitude and a significant increase in unreliability. We find that LLMs often make assumptions in early turns and prematurely attempt to generate final solutions, on which they overly rely. In simpler terms, we discover that when LLMs take a wrong turn in a conversation, they get lost and do not recover.

**Honorable Mention:**[The Polar Express: Optimal Matrix Sign Methods and their Application to the Muon Algorithm](#)

*Noah Amsel, David Persson, Christopher Musco, Robert M. Gower*

**Abstract:** Computing the polar decomposition and the related matrix sign function has been a well-studied problem in numerical analysis for decades. Recently, it has emerged as an important subroutine within the Muon algorithm for training deep neural networks. However, the requirements of this application differ sharply from classical settings: deep learning demands GPU-friendly algorithms that prioritize high throughput over high precision. We introduce Polar Express, a new method for computing the polar decomposition. Like Newton–Schulz and other classical polynomial methods, our approach uses only matrix-matrix multiplications, making it very efficient on GPUs. Inspired by earlier work of Chen & Chow and Nakatsukasa & Freund, Polar Express adapts the update rule at each iteration by solving a minimax optimization problem. We prove that this strategy minimizes error in a worst-case sense, allowing Polar Express to converge as rapidly as possible both in the early iterations and asymptotically. We also address finite-precision issues, making it practical to use in bfloat16. When integrated into Muon, our method yields consistent improvements in validation loss for a GPT-2 model on one to ten billion tokens from the FineWeb dataset, outperforming recent alternatives across a range of learning rates.

Session recordings will be public one month after ICLR 2026 concludes.

###

Media Contact:

Vivian Kelly

Interpose, PR for ICLR

press@iclr.cc