



Twelfth Annual International Conference on Learning Representations (ICLR)

***Inaugural Test of Time, 5 Outstanding Papers and 11 Honorable Mentions Awarded,
Largest Global Participation***

2024 Fact Sheet

Global Participation

- 6,533 participants spanning 79 countries
- 5,938 in-person
- 595 virtual
- 12 of the 79 countries with over 100 participants:
 1. USA 1,647
 2. China 814
 3. Germany 494
 4. UK 452
 5. Korea 315
 6. Canada 244
 7. Australia 215
 8. Switzerland 204
 9. France 160
 10. Singapore 118
 11. Japan 104
 12. Netherlands 102

Previous ICLR Locations and No. of Participants

- 2023: Kagali (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

- 7262 submissions, 2260 accepted
 - 31% acceptance rate (last year: 31%)
 - 2324 more submissions, 686 more accepted compared to last year
- 8950 reviewers (3216 more than 2023)
- 624 area chairs (211 more than 2023)
- 60 senior area chairs (20 more than 2023)
- 99.88% of papers received at least 3 reviews
- 3.1: avg. number of papers a reviewer reviewed
- 9.7: avg. number of papers in an AC's stack
- Average paper score before/after the discussion period: 5.07: ->5.41
- New in 2024 Test of Time and Test of Time Runner-Up
- 5 Outstanding Paper Awards
- 11 Honorable Mention Awards
- 7 [Invited Talks](#)
- 86 [Oral Presentations](#)
- 366 Spotlight Posters
- 10 [Socials](#)
- 20 [Workshops](#) on Saturday only
- 21 [Affinity events](#)
 - 139 Tiny Papers Posters
 - 39 Tiny Papers Oral Presentations
 - 22 [Blogpost Track Posters](#) out of 60 submissions
- 28 [Mentor Chats](#) (previously known as "Office Hours") with senior researchers:
 - Andrea Bajcsy, Samy Bengio, Yoshua Bengio, Danqi Chen, Kyunghyun Choi, Eunsol Choi, Priya Donti, Erin Grant, Moritz Hardt, Tatsunori Hashimoto, Furong Huang, Piotr Koniusz, Rosanne Liu, Lili Mou, Yao Qin, Aditi Raganathan, Andrej Risteski, Anna Rumshisky, Sasha Rush, Masashi Sugiyama, Mihaela van der Schaar, Claire Vernade, René Vidal, Xuezhi Wang, Andrew Gordon Wilson, Amy Zhang, Hanwang Zhang and Luke Zettlemoyer

Inaugural Test of Time Awards

Test of Time Winner

Diederik Kingma, Max Welling, [Auto-Encoding Variational Bayes](#)

Probabilistic modeling is one of the most fundamental ways in which we reason about the world. This paper spearheaded the integration of deep learning with scalable probabilistic inference (amortized mean-field variational inference via a so-called reparameterization trick), giving rise to the Variational Autoencoder (VAE). The lasting value of this work is rooted in its elegance. The principles used to develop VAEs deepened our understanding of the interplay between deep learning and probabilistic modeling, and sparked the development of many subsequent interesting probabilistic models and encoding approaches. A concurrent work by Rezende et al. also proposed a similar idea in the paper titled "Stochastic Backpropagation and Approximate Inference in Deep Generative Models" published at ICML 2014.

Test of Time Runner-Up

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, Rob Fergus, [Intriguing properties of neural networks](#)

With the rising popularity of deep neural networks in real applications, it is important to understand when and how neural networks might behave in undesirable ways. This paper highlighted the issue that neural networks can be vulnerable to small almost imperceptible variations to the input. This idea helped spawn the area of adversarial attacks (trying to fool a neural network) as well as adversarial defense (training a neural network to not be fooled).

5 Outstanding Paper Awards

Generalization in diffusion models arises from geometry-adaptive harmonic representations

Zahra Kadkhodaie, Florentin Guth, Eero P Simoncelli, Stéphane Mallat

<https://openreview.net/forum?id=ANvmVS2Yr0>

This paper provides an important in-depth analysis on generalization and memorization aspects of image diffusion models. The authors empirically study when an image generative model switches from memorizing the input to a generalization regime and they further provide an explanation of this phenomenon in terms of architectural inductive biases by making a connection to ideas from harmonic analysis via "geometry-adaptive harmonic representations". The paper covers a critical missing piece of our understanding of visual generative models and will likely inspire future important theory work in this area.

Learning Interactive Real-World Simulators

Sherry Yang, Yilun Du, Seyed Kamyar Seyed Ghasemipour, Jonathan Tompson, Leslie Pack Kaelbling, Dale Schuurmans, Pieter Abbeel

<https://openreview.net/forum?id=sFyTZEqmUY>

Aggregating data across multiple sources to train foundation models for robotics is a long-term ambitious goal. It poses significant challenges due to different robots having different sensory-motor interfaces which hinder training across large-scale datasets. This work, UniSim, is a significant step in this direction and an engineering feat, aggregating data using a unified interface based on visual perceptions and text descriptions of controls, and training a robotics simulator from the data by leveraging latest developments in vision and language domains.

Never Train from Scratch: Fair Comparison of Long-Sequence Models Requires Data-Driven Priors

Ido Amos, Jonathan Berant, Ankit Gupta

<https://openreview.net/forum?id=PdaPky8MUn>

This paper dives deep into understanding the ability of recently proposed state-space models and transformer architectures to model long-term sequential dependencies. Surprisingly, the authors find that training transformer models from scratch leads to an under-estimation of their performance and demonstrates dramatic gains can be achieved with a pre-training and fine-tuning setup. The paper is exceptionally well executed and exemplary in its focus on simplicity and systematic insights.

Protein Discovery with Discrete Walk-Jump Sampling

Nathan C. Frey, Dan Berenberg, Karina Zadorozhny, Joseph Kleinhenz, Julien Lafrance-Vanasse, Isidro Hotzel, Yan Wu, Stephen Ra, Richard Bonneau, Kyunghyun Cho, Andreas Loukas, Vladimir Gligorijevic, Saeed Saremi

<https://openreview.net/forum?id=zMPHKOmQNb>

This paper addresses the problem of sequence-based antibody design, a timely and important application for generative models of protein sequences. To this end, the authors introduce an innovative and effective new modeling approach specifically tailored to the problem of handling discrete protein sequence data. In addition to validating the method *in silico*, the authors perform extensive wet lab experiments to measure antibody binding affinity *in vitro*, demonstrating the effectiveness of their generative method.

Vision Transformers Need Registers

Timothée Darcet, Maxime Oquab, Julien Mairal, Piotr Bojanowski

<https://openreview.net/forum?id=2dnO3LLiJ1>

This paper identifies artifacts in feature maps of vision transformer networks, characterized by high-norm tokens in low-informative background areas. The authors provide key hypotheses for why this is happening and provide a simple yet elegant solution to address these artifacts using additional register tokens, enhancing model performance on various tasks. The insights gained from this work can also impact other application areas. The paper is very well-written and provides a great example of conducting research – identifying an issue, understanding why it is happening, and then providing a solution.

11 Honorable Mentions

Amortizing intractable inference in large language models

Edward J Hu, Moksh Jain, Eric Elmoznino, Younesse Kaddar, Guillaume Lajoie, Yoshua Bengio, Nikolay Malkin

<https://openreview.net/forum?id=Ouj6p4ca60>

The paper proposes a promising alternative to autoregressive decoding in LLMs from a Bayesian inference perspective that can inspire follow-up studies.

Approximating Nash Equilibria in Normal-Form Games via Stochastic Optimization

Ian Gemp, Luke Marris, Georgios Piliouras

<https://openreview.net/forum?id=cc8h3l3V4E>

An exceptionally clearly written paper, making progress on the important problem of developing efficient and scalable Nash solvers.

Beyond Weisfeiler-Lehman: A Quantitative Framework for GNN Expressiveness

Bohang Zhang, Jingchu Gai, Yiheng Du, Qiwei Ye, Di He, Liwei Wang

<https://openreview.net/forum?id=HSKaGOi7Ar>

Expressivity of GNNs is an important topic for which current solutions, such as the Weisfeiler-Lehman test, still come with significant limitations. The authors propose a new "expressivity theory" based on homomorphism counts.

Flow Matching on General Geometries

Ricky T. Q. Chen, Yaron Lipman

<https://openreview.net/forum?id=g7ohDITL>

This paper tackles the challenging yet important problem of generative modeling on general geometric manifolds, for which it proposes a practical and efficient algorithm. The paper is presented exceptionally well and features a comprehensive experimental validation on a wide range of tasks.

Is ImageNet worth 1 video? Learning strong image encoders from 1 long unlabelled video

Shashanka Venkataramanan, Mamshad Nayeem Rizve, Joao Carreira, Yuki M Asano, Yannis Avrithis

<https://openreview.net/forum?id=Yen1IGNs2o>

The paper proposes a novel path to self-supervised image pre-training, by learning from continuous videos. The paper contributes both new types of data and a method to learn from novel data.

Meta Continual Learning Revisited: Implicitly Enhancing Online Hessian Approximation via Variance Reduction

Yichen Wu, Long-Kai Huang, Renzhen Wang, Deyu Meng, Ying Wei

<https://openreview.net/forum?id=TpD2aG1h0D>

The authors propose a new variance reduction approach to meta continuous learning. The approach is presented well and has not only practical impact but is also backed up with regret analysis.

Model Tells You What to Discard: Adaptive KV Cache Compression for LLMs

Suyu Ge, Yunan Zhang, Liyuan Liu, Minjia Zhang, Jiawei Han, Jianfeng Gao

<https://openreview.net/forum?id=uNrFpDPMyo>

The paper targets the critical KV cache compression problem with great impact on transformer based LLMs, reducing the memory with a simple idea that can be deployed without resource intensive fine-tuning or re-training. The approach is quite simple and yet is shown to be quite effective.

Proving Test Set Contamination in Black-Box Language Models

Yonatan Oren, Nicole Meister, Niladri S. Chatterji, Faisal Ladhak, Tatsunori Hashimoto

<https://openreview.net/forum?id=KS8mlvetg2>

A simple yet elegant method to test whether a supervised-learning dataset has been included in LLM training.

Robust agents learn causal world models

Jonathan Richens, Tom Everitt

<https://openreview.net/forum?id=pOoKI3ouv1>

This paper makes progress on laying theoretical foundations towards understanding the role of causal reasoning in agents' ability to generalize to new domains and has potential implications for a range of related fields.

The mechanistic basis of data dependence and abrupt learning in an in-context classification task

Gautam Reddy

<https://openreview.net/forum?id=aN4Jf6Cx69>

A timely and exceptionally systematic study of the mechanics that underlie in-context vs. in-weight learning at a point where we are only starting to understand these phenomena.

Towards a statistical theory of data selection under weak supervision

Germain Kolossov, Andrea Montanari, Pulkit Tandon



<https://openreview.net/forum?id=HhfcNgQn6p>

The paper establishes statistical foundations for data subset selection and identifies the shortcomings of popular data selection methods.

7 Invited Talks

- Kyunghyun Cho [*Machine Learning in Prescient Design's Lab-in-the-Loop Antibody Design*](#)
- Priya Donti [*Why your work matters for climate in more ways than you think*](#)
- Kate Downing [*Copyright Fundamentals for AI Researchers*](#)
- Raia Hadsell, Google DeepMind, [*Learning through AI's winters and springs: unexpected truths on the road to AGI*](#)
- Moritz Hartd [*The emerging science of benchmarks*](#)
- Devi Parikh [*Stories from my life*](#)
- Jie Tang [*The ChatGLM's Road to AGI*](#)

Session recordings will be public one month after ICLR 2024 concludes.

###

Media Contact:
Becky Obbema
Interpose, PR for ICLR
press@iclr.cc