NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2021 Fact Sheet

35th Annual Conference of Neural Information Processing Systems

Location: Second Virtual Conference

Virtual Attendance and Registration:

- Number of participants at virtual conference: 17,091

Previous Locations:

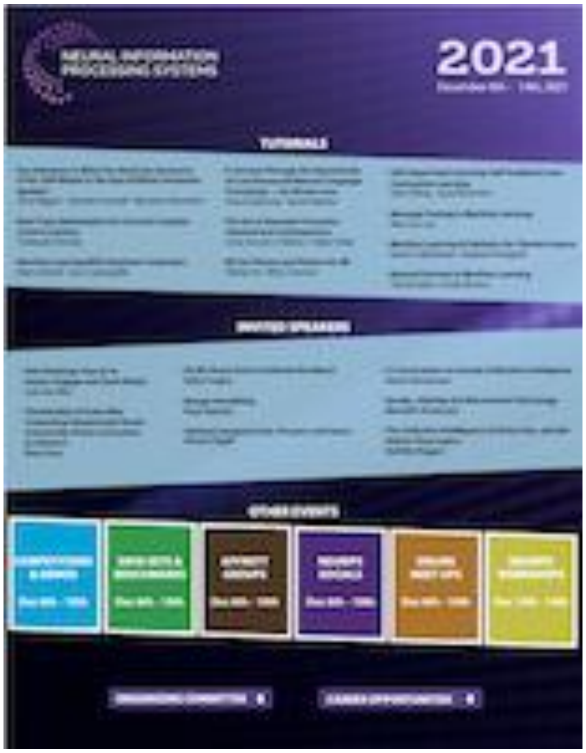
2020 First Virtual Conference	22,823
2019 Vancouver, British Columbia, Canada	13,000
2018 Montreal, Quebec, Canada	8,648
2017 Long Beach, California, United States	8,008
2016 Barcelona, Spain	5,231
2015 Montreal, Quebec, Canada	3,852
2014 Montreal, Quebec, Canada	2,581
2013 Lake Tahoe, California, United States	1,994
2012 Lake Tahoe, California, United States	1,676
2011 Granada, Spain	1,452
2010 Vancouver, British Columbia, Canada	1,354

2021 Program:

- Number of tracks in conference: 8
- Number of papers accepted: 2,334
- Number of paper submissions: 9,122
- Number of abstract submissions: 11,292
- Acceptance percentage: 25.6%
- Number of workshops: 60
- Number of tutorials: 10
- Number of competitions: 22
- Number of demonstrations: 18
- Oral presentations: 55
- Number of socials: 15
- Virtual and local Meetups: 40

New Program Features:

- Dataset and Benchmarks Track
  - Submitted papers for dataset and benchmark track: 484
  - Accepted papers for dataset and benchmark track: 174
- Career Website: A resource for finding a position from a wide range of companies and universities
  - 208 companies//institutions registered to post opportunities
  - 259 opportunities posted
- Roundtable: *How Copyright Shapes Your Datasets and What To Do About It*



[pdf version]

2020 Program:

- Number of tracks in conference: 7
- Number of papers accepted: 1,898
- Number of paper submissions: 9,467
- Acceptance percentage: 20%
- Number of workshops 60
- Number of tutorials: 16
- Number of competition tracks: 16

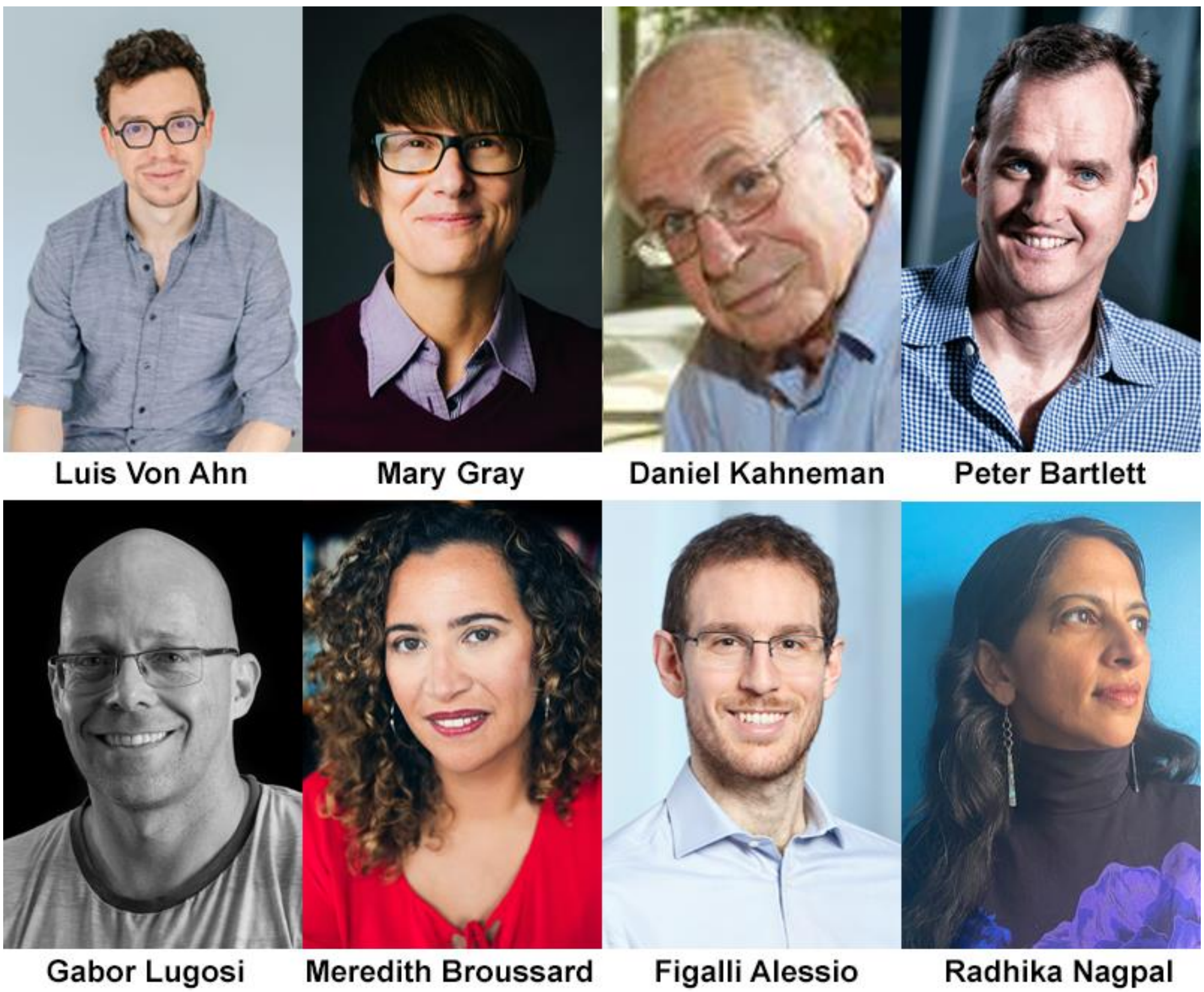
All accepted papers can be found in the proceedings:
https://proceedings.neurips.cc/paper/2021

Affinity Groups Represented:

- Black in AI
- Indigenous in AI
- LatinX in AI
- Queer in AI
- Women in ML (WiML)

Invited Keynote Speakers:

- *Featured plenary interview with Daniel Kahneman - A Conversation on Human and Machine Intelligence*
- *Roundtable - Amanda Levendowski - How Copyright Shapes Your Datasets and What To Do About It*
- *Luis von Ahn - How Duolingo Uses AI to Assess, Engage and Teach Better*
- *Peter Bartlett - Benign Overfitting*
- *Meredith Broussard - Gender, Allyship & Public Interest Technology*
- *Alessio Figalli - Optimal Transport: Past, Present, and Future*
- *Mary L. Gray - The Banality of Scale: A Theory on the Limits of Modeling Bias and Fairness Frameworks for Social Justice (and other lessons from the Pandemic)*
- *Gábor Lugosi - Do We Know How to Estimate the Mean?*
- *Radhika Nagpal - The Collective Intelligence of Army Ants, and the Robots They Inspire*



Paper Award Recipients:

- **Outstanding Paper Award**:
    - **Title**: A Universal Law of Robustness via Isoperimetry
    - **Authors:** Sébastien Bubeck and Mark Sellke
    - **Presented**: Tuesday, December 7 at 08:20 GMT (12:20 am PST) in the session on Deep Learning Theory and Causality
    - **About**: This paper proposes a theoretical model to explain why many state-of-the-art deep networks require many more parameters than are necessary to smoothly fit the training data. In particular, under certain regularity conditions about the training distribution, the number of parameters needed for an O(1)-Lipschitz function to interpolate training data below the label noise scales as *nd*, where *n* is the number of training examples, and *d* is the dimensionality of the data. This result stands in stark contrast to conventional results stating that one needs *n* parameters for a function to interpolate the training data, and this extra factor of *d* appears necessary in order to smoothly interpolate. The theory is simple and elegant, and consistent with some empirical observations about the size of models that have robust generalization on MNIST classification. This work also offers a testable prediction about the model sizes needed to develop robust models for ImageNet classification.

- **Outstanding Paper Award**:
    - **Title**: On the Expressivity of Markov Reward
    - **Authors**: David Abel, Will Dabney, Anna Harutyunyan, Mark K. Ho, Michael Littman, Doina Precup, and Satinder Singh
    - **Presented**: Tuesday, December 7 at 09:20 GMT (1:20 am PST) in the session on Reinforcement Learning
    - **About:** Markov reward functions are the dominant framework for sequential decision making under uncertainty and reinforcement learning. This paper provides a careful, clear exposition of when Markov rewards are, or are not, sufficient to enable a system designer to specify a task, in terms of their preference for a particular behavior, preferences over behaviors, or preferences over state and action sequences. The authors demonstrate with simple, illustrative examples that there exist some tasks for which no Markov reward function can be specified that induces the desired task and result. Fortunately, they also show that it is possible in polynomial time to decide if a compatible Markov reward exists for a desired setting, and if it does, there also exists a polynomial time algorithm to construct such a Markov reward in the finite decision process setting. This work sheds light on the challenge of reward design and may open up future avenues of research into when and how the Markov framework is sufficient to achieve performance desired by human stakeholders.

- **Outstanding Paper Award:**
    - **Title**: MAUVE: Measuring the Gap Between Neural Text and Human Text using Divergence Frontiers
    - **Authors**: Krishna Pillutla, Swabha Swayamdipta, Rowan Zellers, John Thickstun, Sean Welleck, Yejin Choi and Zaid Harchaoui
    - **Presented**: Tuesday, December 7 at 8:00 GMT (midnight PST) in the session on Deep Learning

- **About**: This paper presents MAUVE, a divergence measure to compare the distribution of model-generated text with the distribution of human-generated text. The idea is simple and elegant, and it basically uses a continuous family of (soft) KL divergence measures of quantized embeddings of the two texts being compared. The proposed MAUVE measure is essentially an integration over the continuous family of measures, and aims to capture both Type I error (generating unrealistic text) and Type II error (not capturing all possible human text). The empirical experiments demonstrate that MAUVE identifies the known patterns of model-generated text and correlates better with human judgments compared to previous divergence metrics. The paper is well-written, the research question is important in the context of rapid progress of open-ended text generation, and the results are clear.

- **Outstanding Paper Award:**
    - **Title**: Continuized Accelerations of Deterministic and Stochastic Gradient Descents, and of Gossip Algorithms
    - **Authors**: Mathieu Even, Raphaël Berthier, Francis Bach, Nicolas Flammarion, Pierre Gaillard, Hadrien Hendrikx, Laurent Massoulié and Adrien Taylor
    - **Presented**: Wednesday, December 8 at 16:00 GMT (8:00 am PST) in the session on Optimization
    - **About**: This paper describes a “continuized” version of Nesterov’s accelerated gradient method in which the two separate vector variables evolve jointly in continuous-time—much like previous approaches that use differential equations to understand acceleration—but uses gradient updates that occur at random times determined by a Poisson point process. This new approach leads to a (randomized) discrete-time method that: (1) enjoys the same accelerated convergence as Nesterov’s method; (2) comes with a clean and transparent analysis that leverages continuous-time arguments, which is arguably easier to understand than prior analyses of accelerated gradient methods; and (3) avoids additional errors from discretizing a continuous-time process, which stands in stark contrast to several previous attempts to understand accelerated methods using continuous-time processes.

- **Outstanding Paper Award:**
    - **Title**: Moser Flow: Divergence-based Generative Modeling on Manifolds
    - **Authors**: Noam Rozen, Aditya Grover, Maximilian Nickel and Yaron Lipman
    - **Presented**: Saturday, December 11 at 00:00 GMT (Friday, December 10 at 4:00 pm PST) in the session on Generative Modeling
    - **About**: This paper proposes a method for training continuous normalizing flow (CNF) generative models over Riemannian manifolds. The key idea is to leverage a result by Moser (1965) that characterizes the solution of a CNF (which Moser called an orientation preserving automorphism on manifolds) using a restricted class of ODEs that enjoys geometric regularity conditions, and is explicitly defined using the divergence of the target density function. The proposed Moser Flow method uses this solution concept to develop a CNF approach based on a parameterized target density estimator (which can be a neural network). Training amounts to simply optimizing the divergence of the density estimator, which side-steps running an ODE solver (required for standard backpropagation training). The experiments show faster training times and superior test performance compared to prior CNF work, as well as the ability to model densities on

implicit surfaces with non-constant curvature such as the Stanford Bunny model. More generally, this concept of exploiting geometric regularity conditions to side-step expensive backpropagation training may be of broader interest.

- **Test of Time Award:**
  - **Title**: [Online Learning for Latent Dirichlet Allocation]
  - **Year**: NeurIPS 2010
  - **Authors**: Matthew Hoffman, David Blei and Francis Bach
  - **Presented:** Saturday, December 11 at 01:00 GMT (Friday, December 10 at 5:00 pm PST) in the final session of the conference
  - **About:** This paper introduces a stochastic variational gradient based inference procedure for training Latent Dirichlet Allocation (LDA) models on very large text corpora. On the theoretical side it is shown that the training procedure converges to a local optimum and that, surprisingly, the simple stochastic gradient updates correspond to a stochastic natural gradient of the evidence lower bound (ELBO) objective. On the empirical side the authors show that for the first time LDA can be comfortably trained on text corpora of several hundreds of thousands of documents, making it a practical technique for "big data" problems. The idea has made a large impact in the ML community because it represented the first stepping stone for general stochastic gradient variational inference procedures for a much broader class of models. After this paper, there would be no good reason to ever use full batch training procedures for variational inference anymore.

New to NeurIPS 2021 - Best Paper Awards for the datasets and benchmarks track:

- **Datasets & Benchmarks Best Paper Award**:
  - **Title**: [Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research]
  - **Authors**: Bernard Koch, Emily Denton, Alex Hanna, Jacob Gates Foster
  - **Presented:** Wednesday, December 8 at 16:00 GMT (8:00 PST) in Dataset and Benchmark Track 2
  - **About**: This paper analyzes thousands of papers and studies the evolution of dataset use within different machine learning subcommunities, as well as the interplay between dataset adoption and creation. It finds that in most communities, there is an evolution towards using fewer different datasets over time, and that these datasets come from a handful of elite institutions. This evolution is problematic, since benchmarks become less generalizable, biases that exist within the sources of these datasets may be amplified, and it becomes harder for new datasets to be accepted by the research community. This is an important 'wake up call' for the machine learning community as a whole, to think more critically about which datasets are used for benchmarking, and to put more emphasis on the creation of new and more varied datasets.

- **Datasets & Benchmarks Best Paper Award**:
  - **Title**: [ATOM3D: Tasks on Molecules in Three Dimensions]
  - **Authors**: Raphael John Lamarre Townshend, Martin Vögele, Patricia Adriana Suriana, Alexander Derry, Alexander Powers, Yianni Laloudakis, Sidhika Balachandar, Bowen Jing, Brandon M. Anderson, Stephan Eismann, Risi Kondor, Russ Altman, Ron O. Dror

- **Presented:** Wednesday, December 8 at 16:00 GMT (8:00 PST) in Dataset and Benchmark Track 2
- **About:** This paper introduces a collection of benchmark datasets with 3D representations of small molecules and/or biopolymers for solving a wide range of problems, spanning single molecular structure prediction and interactions between biomolecules as well as molecular functional and design/engineering tasks. Simple yet robust implementations of 3D models are then benchmarked against state-of-the-art models with 1D or 2D representation, and show better performance over lower-dimensional counterparts. This work provides important insight about how to choose and design models for a given task. Not only does this work provide benchmarking datasets, it also provides baseline models and open source tools to leverage these datasets and models, dramatically lowering the entry barrier for machine learning people to get into computational biology and molecule design.

2021 Organizing Committee:
This year's organizing committee is led by general chair Marc'Aurelio Ranzato, DeepMind and program chair Alina Beygelzimer, Yahoo Research, along with the program co-chairs Yann Dauphin, Google Brain, Percy Liang, Stanford University and Jenn Wortman Vaughan, Microsoft Research.

Accessing NeurIPS content:
The tutorials and Invited talks are available on the virtual site to anyone with a free NeurIPS account, no NeurIPS registration required.

All content will be available to the public online 30 days after the conference.

# # #