# ICML 2019 Workshop book

## Schedule Highlights

Generated Sun Dec 20, 2020

Workshop organizers make last-minute changes to their schedule. Download this document again to get the lastest changes, or use the ICML mobile application.

### June 14, 2019

| | |
|---|---|
| 101 | **ICML 2019 Workshop on Computational Biology**<br>*Pe'er, Prabhakaran, Azizi, Diallo, Kundaje, Engelhardt, Dhifli, MEPHU NGUIFO, Tansey, Vogt, Listgarten, Burdziak, CompBio* |
| 102 | **ICML 2019 Time Series Workshop**<br>*Kuznetsov, Yang, Yu, Tang, Wang* |
| 103 | **Human In the Loop Learning (HILL)**<br>*Wang, Wang, Yu, Zhang, Gonzalez, Jia, Bird, Varshney, Kim, Weller* |
| 104 A | **Climate Change: How Can AI Help?**<br>*Rolnick, Lacoste, Maharaj, Chayes, Bengio* |
| 104 B | **Workshop on the Security and Privacy of Machine Learning**<br>*Papernot, Tramer, Li, Boneh, Evans, Jha, Liang, McDaniel, Steinhardt, Song* |
| 104 C | **Theoretical Physics for Deep Learning**<br>*Lee, Pennington, Bahri, Welling, Ganguli, Bruna* |
| 201 | **AI in Finance: Applications and Infrastructure for Multi-Agent Learning**<br>*Reddy, Balch, Wellman, Kumar, Stoica, Elkind* |
| 202 | **The Third Workshop On Tractable Probabilistic Modeling (TPM)**<br>*Lowd, Vergari, Molina, Rahman, Domingos, Vergari* |
| 203 | **Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**<br>*Ravi, Kozareva, Fan, Welling, Chen, Bailer, Kulis, Hu, Dekhtiar, Lin, Marculescu* |
| 204 | **Negative Dependence: Theory and Applications in Machine Learning**<br>*Gartrell, Gillenwater, Kulesza, Mariet* |
| Grand Ballroom A | **Understanding and Improving Generalization in Deep Learning**<br>*Krishnan, Mobahi, Neyshabur, Bartlett, Song, Srebro* |
| Grand Ballroom B | **6th ICML Workshop on Automated Machine Learning (AutoML 2019)**<br>*Hutter, Vanschoren, Eggensperger, Feurer* |
| Hall A | **Generative Modeling and Model-Based Reasoning for Robotics and AI**<br>*Rajeswaran, Todorov, Mordatch, Agnew, Zhang, Pineau, Chang, Erhan, Levine, Stachenfeld, Zhang* |
| Hall B | **Uncertainty and Robustness in Deep Learning**<br>*Li, Lakshminarayanan, Hendrycks, Dietterich, Gilmer* |
| Seaside Ballroom | **Reinforcement Learning for Real Life**<br>*Li, Geramifard, Li, Szepesvari, Wang* |
| Seaside Ballroom | **Real-world Sequential Decision Making: Reinforcement Learning and Beyond**<br>*Le, Yue, Swaminathan, Boots, Cheng* |

### June 15, 2019

| | |
|---|---|
| 101 | **Workshop on AI for autonomous driving**<br>*Choromanska, Jackel, Li, Niebles, Gaidon, Posner, Chao* |
| 102 | **Workshop on Multi-Task and Lifelong Reinforcement Learning** |

| | |
|---|---|
| | *Chandar, Sodhani, Khetarpal, Zahavy, Mankowitz, Mannor, Ravindran, Precup, Finn, Gupta, Zhang, Cho, Rusu, Rob Fergus* |
| 103 | **Invertible Neural Networks and Normalizing Flows**<br>*Huang, Krueger, Van den Berg, Papamakarios, Gomez, Cremer, Chen, Courville, J. Rezende* |
| 104 A | **Stein's Method for Machine Learning and Statistics**<br>*Briol, Mackey, Oates, Liu, Goldstein* |
| 104 B | **AI For Social Good (AISG)**<br>*Luck, Sankaran, Sylvain, McGregor, Penn, Tadesse, Sylvain, Côté, Mackey, Ghani, Bengio* |
| 104 C | **Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes**<br>*Biggio, Korshunov, Mensink, Patrini, Sadhu, Rao* |
| 201 | **ICML Workshop on Imitation, Intent, and Interaction (I3)**<br>*Rhinehart, Levine, Finn, He, Kostrikov, Fu, Reddy* |
| 202 | **Coding Theory For Large-scale Machine Learning**<br>*Cadambe, Grover, Papailiopoulos, Joshi* |
| 203 | **The How2 Challenge: New Tasks for Vision & Language**<br>*Metze, Specia, Elliot, Barrault, Sanabria, Palaskar* |
| 204 | **Machine Learning for Music Discovery**<br>*Schmidt, Nieto, Gouyon, Kinnaird, Lanckriet* |
| Grand Ballroom A | **Workshop on Self-Supervised Learning**<br>*van den Oord, Aytar, Doersch, Vondrick, Radford, Sermanet, Zamir, Abbeel* |
| Grand Ballroom B | **Learning and Reasoning with Graph-Structured Representations**<br>*Fetaya, Hu, Kipf, Li, Liang, Liao, Urtasun, Wang, Welling, Xing, Zemel* |
| Hall A | **Exploration in Reinforcement Learning Workshop**<br>*Bhupatiraju, Eysenbach, Gu, Edwards, White, Oudeyer, Stanley, Brunskill* |
| Hall B | **Identifying and Understanding Deep Learning Phenomena**<br>*Sedghi, Bengio, Hata, Madry, Morcos, Neyshabur, Raghu, Rahimi, Schmidt, Xiao* |
| Seaside Ballroom | **Adaptive and Multitask Learning: Algorithms & Systems**<br>*Al-Shedivat, Platanios, Stretcu, Andreas, Talwalkar, Caruana, Mitchell, Xing* |

# June 14, 2019

## ICML 2019 Workshop on Computational Biology

***Donna Pe'er, Sandhya Prabhakaran, Elham Azizi, Abdoulaye Baniré Diallo, Anshul Kundaje, Barbara Engelhardt, Wajdi Dhifli, Engelbert MEPHU NGUIFO, Wesley Tansey, Julia Vogt, Jennifer Listgarten, Cassandra Burdziak, Workshop CompBio***

**101, Fri Jun 14, 08:30 AM**

The workshop will showcase recent research in the field of Computational Biology. There has been significant development in genomic sequencing techniques and imaging technologies. These approaches not only generate huge amounts of data but provide unprecedented resolution of single cells and even subcellular structures. The availability of high dimensional data, at multiple spatial and temporal resolutions has made machine learning and deep learning methods increasingly critical for computational analysis and interpretation of the data. Conversely, biological data has also exposed unique challenges and problems that call for the development of new machine learning methods. This workshop aims to bring together researchers working at the intersection of Machine Learning and Biology to present recent advances and open questions in Computational Biology to the ML community.

The workshop is a sequel to the WCB workshops we organized in the last three years Joint ICML and IJCAI 2018, Stockholm, ICML 2017, Sydney and ICML 2016, New York as well as Workshop on Bioinformatics and AI at IJCAI 2015 Buenos Aires, IJCAI 2016 New York, IJCAI 2017 Melbourne which had excellent line-ups of talks and were well-received by the community. Every year, we received 60+ submissions. After multiple rounds of rigorous reviewing around 50 submissions were selected from which the best set of papers were chosen for Contributed talks and Spotlights and the rest were invited as Posters. We have a steadfast and growing base of reviewers making up the Program Committee. For the past edition, a special issue of Journal of Computational Biology will be released in the following weeks with extended versions of 14 accepted papers.

We have two confirmed invited speakers and we will invite at least one more leading researcher in the field. Similar to previous years, we plan to request partial funding from Microsoft Research, Google, Python, Swiss Institute of Bioinformatics that we intend to use for student travel awards. In past years, we have also been able to provide awards for best poster/paper and partially contribute to travel expenses for at least 8 students per year.

The Workshop proceedings will be available through CEUR proceedings. We would also have an extended version to be included in a special issue of the Journal of Computational Biology (JCB) for which we already have an agreement with JCB.

## Schedule

| | |
|---|---|
| 08:30 AM | **WCB Organisers** |
| 08:40 AM | **Caroline Uhler** |
| 09:20 AM | **Jacopo Cirrone, Marttinen Pekka , Elior Rahmani, Elior Rahmani, Harri Lähdesmäki** |
| 09:50 AM | **Ali Oskooei** |
| 11:00 AM | **Francisco LePort** |
| 11:40 AM | **Ali Oskooei, Zhenqin Wu, Karren Dai Yang, Mijung Kim** |
| 12:00 PM | **Poster Session & Lunch break** *Wiese, Carter, DeBlasio, Hashir, Chan, Manica, Oskooei, Wu, Yang, FAGES, Liu, Beebe-Wang, He, Cirrone, Marttinen, Rahmani, Lähdesmäki, Yadala, Deac, Soleimany, Mane, Ernst, Cohen, Mathew, Agarwal, Zheng* |
| 02:00 PM | **Ngo Trong Trung** |
| 02:20 PM | **Achille Nazaret, François Fages, Brandon Carter, Ruishan Liu, Nicasia Beebe-Wang** |
| 02:50 PM | **Poster Session & Coffee Break** |
| 04:00 PM | **Daphne Koller** |

04:40 PM **Bryan He**

05:00 PM **Award Ceremony & Closing Remarks**

## ICML 2019 Time Series Workshop

***Vitaly Kuznetsov, Scott Yang, Rose Yu, Cheng Tang, Yuyang Wang***

**102, Fri Jun 14, 08:30 AM**

Time series data is both quickly growing and already ubiquitous. In domains spanning as broad a range as climate, robotics, entertainment, finance, healthcare, and transportation, there has been a significant shift away from parsimonious, infrequent measurements to nearly continuous monitoring and recording. Rapid advances in sensing technologies, ranging from remote sensors to wearables and social sensing, are generating rapid growth in the size and complexity of time series data streams. Thus, the importance and impact of time series analysis and modelling techniques only continues to grow.

At the same time, while time series analysis has been extensively studied by econometricians and statisticians, modern time series data often pose significant challenges for the existing techniques both in terms of their structure (e.g., irregular sampling in hospital records and spatiotemporal structure in climate data) and size. Moreover, the focus on time series in the machine learning community has been comparatively much smaller. In fact, the predominant methods in machine learning often assume i.i.d. data streams, which is generally not appropriate for time series data. Thus, there is both a great need and an exciting opportunity for the machine learning community to develop theory, models and algorithms specifically for the purpose of processing and analyzing time series data.

We see ICML as a great opportunity to bring together theoretical and applied researchers from around the world and with different backgrounds who are interested in the development and usage of time series analysis and algorithms. This includes methods for time series prediction, classification, clustering, anomaly and change point detection, causal discovery, and dimensionality reduction as well as general theory for learning and analyzing stochastic processes. Since time series have been studied in a variety of different fields and have many broad applications, we plan to host leading academic researchers and industry experts with a range of perspectives and interests as invited speakers. Moreover, we also invite researchers from the related areas of batch and online learning, deep learning, reinforcement learning, data analysis and statistics, and many others to both contribute and participate in this workshop.

## Human In the Loop Learning (HILL)

***Xin Wang, Xin Wang, Fisher Yu, Shanghang Zhang, Joseph Gonzalez, Yangqing Jia, Sarah Bird, Kush Varshney, Been Kim, Adrian Weller***

**103, Fri Jun 14, 08:30 AM**

https://sites.google.com/view/hill2019/home

This workshop is a joint effort between the 4th ICML Workshop on Human Interpretability in Machine Learning (WHI) and the ICML 2019 Workshop on Interactive Data Analysis System (IDAS). We have combined our forces this year to run Human in the Loop Learning (HILL) in conjunction with ICML 2019!

The workshop will bring together researchers and practitioners who study interpretable and interactive learning systems with applications in large scale data processing, data annotations, data visualization, human-assisted data integration, systems and tools to interpret machine learning models as well as algorithm designs for active learning, online learning, and interpretable machine learning algorithms. The target audience for the workshop includes people who are interested in using machines to solve problems by having a human be an integral part of the process. This workshop serves as a platform where researchers can discuss approaches that bridge the gap between humans and machines and get the best of both worlds.

We welcome high-quality submissions in the broad area of human in the loop learning. A few (non-exhaustive) topics of interest include:

Systems for online and interactive learning algorithms,
Active/Interactive machine learning algorithm design,
Systems for collecting, preparing, and managing machine learning data,
Model understanding tools (verifying, diagnosing, debugging, visualization, introspection, etc),
Design, testing and assessment of interactive systems for data analytics,
Psychology of human concept learning,
Generalized additive models, sparsity and rule learning,
Interpretable unsupervised models (clustering, topic models, etc.),
Interpretation of black-box models (including deep neural networks),
Interpretability in reinforcement learning.

## Schedule

| | |
|---|---|
| 08:25 AM | **Opening Remarks** |
| 08:30 AM | **Invited Talk: James Philbin** |
| 09:00 AM | **Invited Talk: Sanja Fidler** |
| 09:30 AM | **Invited Talk: Bryan Catanzaro** |
| 10:00 AM | **IDAS Poster Session & Coffee break** |
| 11:30 AM | **Invited Talk: Yisong Yue** |
| 12:00 PM | **Invited Talk: Vittorio Ferrari** |
| 12:30 PM | **Lunch Break** |
| 02:00 PM | **Interpretability Contributed Talks** |
| 03:00 PM | **Coffee Break** |
| 03:30 PM | **Interpretability Invited Discussion: California's Senate Bill 10 (SB 10) on Pretrial Release and Detention with Solon Barocas and Peter Eckersley** |
| 04:45 PM | **Human in the Loop Learning Panel Discussion** |

Abstracts (1):

Abstract 12: **Human in the Loop Learning Panel Discussion in Human In the Loop Learning (HILL)**, 04:45 PM

Panelists: Vittorio Ferrari, Marcin Detyniecki, and James Philbin

## Climate Change: How Can AI Help?

***David Rolnick, Alexandre Lacoste, Tegan Maharaj, Jennifer Chayes, Yoshua Bengio***

**104 A, Fri Jun 14, 08:30 AM**

Many in the machine learning community wish to take action on climate change, yet feel their skills are inapplicable. This workshop aims to show that in fact the opposite is true: while no silver bullet, ML can be an invaluable tool both in reducing greenhouse gas emissions and in helping society adapt to the effects of climate change. Climate change is a complex problem, for which action takes many forms - from designing smart electrical grids to tracking deforestation in satellite imagery. Many of these actions represent high-impact opportunities for real-world change, as well as being interesting problems for ML research.

## Schedule

| | |
|---|---|
| 08:30 AM | **Opening Remarks** |
| 08:45 AM | **AI for Climate Change: the context** *Platt* |
| 09:20 AM | **Why it's hard to mitigate climate change, and how to do better** *Kelly* |
| 09:45 AM | **Tackling climate change challenges with AI through collaboration** *Ng* |
| 10:10 AM | **Towards a Sustainable Food Supply Chain Powered by Artificial Intelligence** *Kuleshov* |
| 10:20 AM | **Deep Learning for Wildlife Conservation and Restoration Efforts** *DUHART* |
| 10:30 AM | **Morning Coffee Break + Poster Session** |

11:00 AM **Achieving Drawdown - Chad Frischmann**

12:00 PM **Networking Lunch (provided) + Poster Session** *Stanway, Robson, Rangnekar, Chattopadhyay, Pilipiszyn, LeRoy, Cheng, Zhang, Shen, Schroeder, Clough, DUHART, Fung, Ududec, Wang, Dao, wu, Giannakis, Sejdinovic, Precup, Watson-Parris, Wen, Chen, Erinjippurath, Li, Zou, van Hoof, Scannell, Mamitsuka, Zhang, Choo, Wang, Requeima, Hwang, Xu, Mathe, Binas, Lee, Ramea, Duffy, McCloskey, Sankaran, Mackey, Mones, Benabbou, Kaack, Hoffman, Mudigonda, Mahdavi, McCourt, Jiang, Kamani, Guha, Dalmasso, Pawlowski, Milojevic-Dupont, Orenstein, Hassanzadeh, Marttinen, Nair, Farhang, Kaski, Manjanna, Luccioni, Deshpande, Kim, Mouatadid, Park, Lin, Felgueira, Hornigold, Yuan, Beucler, Cui, Kuleshov, Yu, song, Wexler, Bengio, Wang, Yi, Malki*

01:30 PM **Personalized Visualization of the Impact of Climate Change** *Bengio*

01:55 PM **Advances in Climate Informatics: Machine Learning for the Study of Climate Change** *Monteleoni*

02:30 PM **Detecting anthropogenic cloud perturbations with deep learning** *Watson-Parris*

02:40 PM **Evaluating aleatoric and epistemic uncertainties of time series deep learning models for soil moisture predictions** *Shen*

02:50 PM **Targeted Meta-Learning for Critical Incident Detection in Weather Data** *Kamani, Farhang, Mahdavi, Wang*

03:00 PM **Afternoon Coffee Break + Poster Session**

03:30 PM **Geoscience data and models for the Climate Change AI community** *Mukkavilli*

03:45 PM **ML vs. Climate Change, Applications in Energy at DeepMind** *Witherspoon*

04:20 PM **Truck Traffic Monitoring with Satellite Images** *Kaack, Chen*

04:30 PM **Machine Learning for AC Optimal Power Flow** *Guha, Wang*

04:40 PM **Planetary Scale Monitoring of Urban Growth in High Flood Risk Areas** *Clough, Nair, Erinjippurath*

04:50 PM **"Ideas" mini-spotlights** *McCloskey, Milojevic-Dupont, Binas, Schroeder, Luccioni*

05:15 PM **Panel Discussion** *Bengio, Ng, Hadsell, Platt, Monteleoni, Chayes*

Abstracts (14):

Abstract 2: **AI for Climate Change: the context in Climate Change: How Can AI Help?**, *Platt* 08:45 AM

This talk will set the context around "AI for climate change". This context will include the difference between long-term energy research and shorter-term research required to mitigate or adapt to climate change. I'll illustrate the urgency of the latter research by discussing the carbon budget of the atmosphere. The talk will also highlight some examples of how AI can be applied to climate change mitigation and energy research, including ML for fusion and for flood prediction.

Abstract 3: **Why it's hard to mitigate climate change, and how to do better in Climate Change: How Can AI Help?**, *Kelly* 09:20 AM

It's hard to have climate impact! Lots of projects look great from a distance but fail in practice. The energy system is enormously complex, and there are many non-technical bottlenecks to having impact. In this talk, I'll describe some of these challenges, so you can try to avoid them and hence reduce emissions more rapidly! Let's say you've built a great ML algorithm and written a paper. Now what? Your paper is completely invisible to the climate. How do you get your research used by the energy system? I don’t claim to have all the answers; but I’d like to discuss some of the challenges, and some ideas for how to get round them.

Abstract 4: **Tackling climate change challenges with AI through collaboration in Climate Change: How Can AI Help?**, *Ng* 09:45 AM

The time is now for the AI community to collaborate with the climate community to help

understand, mitigate, and adapt to climate change. In this talk, I will present two projects as part of interdisciplinary collaborations, one in the earth system sciences and one in the energy space, to illustrate specific use cases where AI is making an impact on climate change. I hope this talk will motivate you to contribute to tackling one of the greatest challenges of our time.

Abstract 5: **Towards a Sustainable Food Supply Chain Powered by Artificial Intelligence in Climate Change: How Can AI Help?**, *Kuleshov* 10:10 AM

About 30-40% of food produced worldwide is wasted. This puts a severe strain on the environment and represents a $165B loss to the US economy. This paper explores how artificial intelligence can be used to automate decisions across the food supply chain in order to reduce waste and increase the quality and affordability of food. We focus our attention on supermarkets — combined with downstream consumer waste, these contribute to 40% of total US food losses — and we describe an intelligent decision support system for supermarket operators that optimizes purchasing decisions and minimizes losses. The core of our system is a model-based reinforcement learn- ing engine for perishable inventory management; in a real-world pilot with a US supermarket chain, our system reduced waste by up to 50%. We hope that this paper will bring the food waste problem to the attention of the broader machine learning research community.

Abstract 6: **Deep Learning for Wildlife Conservation and Restoration Efforts in Climate Change: How Can AI Help?**, *DUHART* 10:20 AM

Climate change and environmental degradation are causing species extinction worldwide. Automatic wildlife sensing is an urgent requirement to track biodiversity losses on Earth. Recent improvements in machine learning can accelerate the development of large-scale monitoring systems that would help track conservation outcomes and target efforts. In this paper, we present one such system we developed. 'Tidzam' is a Deep Learning framework for wildlife detection, identification, and geolocalization, designed for the Tidmarsh Wildlife Sanctuary, the site of the largest freshwater wetland restoration in Massachusetts.

Abstract 9: **Networking Lunch (provided) + Poster Session in Climate Change: How Can AI Help?**, *Stanway, Robson, Rangnekar, Chattopadhyay, Pilipiszyn, LeRoy, Cheng, Zhang, Shen, Schroeder, Clough, DUHART, Fung, Ududec, Wang, Dao, wu, Giannakis, Sejdinovic, Precup, Watson-Parris, Wen, Chen, Erinjippurath, Li, Zou, van Hoof, Scannell, Mamitsuka, Zhang, Choo, Wang, Requeima, Hwang, Xu, Mathe, Binas, Lee, Ramea, Duffy, McCloskey, Sankaran, Mackey, Mones, Benabbou, Kaack, Hoffman, Mudigonda, Mahdavi, McCourt, Jiang, Kamani, Guha, Dalmasso, Pawlowski, Milojevic-Dupont, Orenstein, Hassanzadeh, Marttinen, Nair, Farhang, Kaski, Manjanna, Luccioni, Deshpande, Kim, Mouatadid, Park, Lin, Felgueira, Hornigold, Yuan, Beucler, Cui, Kuleshov, Yu, song, Wexler, Bengio, Wang, Yi, Malki* 12:00 PM

Catered sandwiches and snacks will be provided (including vegetarian/vegan and gluten-free options). Sponsored by Element AI.

Abstract 12: **Detecting anthropogenic cloud perturbations with deep learning in Climate Change: How Can AI Help?**, *Watson-Parris* 02:30 PM

One of the most pressing questions in climate science is that of the effect of anthropogenic aerosol on the Earth's energy balance. Aerosols provide the `seeds' on which cloud droplets form, and changes in the amount of aerosol available to a cloud can change its brightness and other physical properties such as optical thickness and spatial extent. Clouds play a critical role in moderating global temperatures and small perturbations can lead to significant amounts of cooling or warming. Uncertainty in this effect is so large it is not currently known if it is negligible, or provides a large enough cooling to largely negate present-day warming by CO2. This work uses deep convolutional neural networks to look for two particular perturbations in clouds due to anthropogenic aerosol and assess their properties

and prevalence, providing valuable insights into their climatic effects.

Abstract 13: **Evaluating aleatoric and epistemic uncertainties of time series deep learning models for soil moisture predictions in Climate Change: How Can AI Help?**, *Shen* 02:40 PM

Soil moisture is an important variable that determines floods, vegetation health, agriculture productivity, and land surface feedbacks to the atmosphere, etc. Accurately modeling soil moisture has important implications in both weather and climate models. The recently available satellite-based observations give us a unique opportunity to build data-driven models to predict soil moisture instead of using land surface models, but previously there was no uncertainty estimate. We tested Monte Carlo dropout (MCD) with an aleatoric term for our long short-term memory models for this problem, and asked if the uncertainty terms behave as they were argued to. We show that the method successfully captures the predictive error after tuning a hyperparameter on a representative training dataset. We show the MCD uncertainty estimate, as previously argued, does detect dissimilarity. In this talk, several important challenges with climate modeling where machine learning may help are also introduced to open up a discussion.

Abstract 14: **Targeted Meta-Learning for Critical Incident Detection in Weather Data in Climate Change: How Can AI Help?**, *Kamani, Farhang, Mahdavi, Wang* 02:50 PM

Due to imbalanced or heavy-tailed nature of weather- and climate-related datasets, the performance of standard deep learning models significantly deviates from their expected behavior on test data. Classical methods to address these issues are mostly data or application dependent, hence burdensome to tune.
Meta-learning approaches, on the other hand, aim to learn hyperparameters in the learning process using different objective functions on training and validation data. However, these methods suffer from high computational complexity and are not scalable to large datasets. In this paper, we aim to apply a novel framework named as targeted meta-learning to rectify this issue, and show its efficacy in dealing with the aforementioned biases in datasets. This framework employs a small, well-crafted target dataset that resembles the desired nature of test data in order to guide the learning process in a coupled manner. We empirically show that this framework can overcome the bias issue, common to weather-related datasets, in a bow echo detection case study.

Abstract 16: **Geoscience data and models for the Climate Change AI community in Climate Change: How Can AI Help?**, *Mukkavilli* 03:30 PM

This talk will outline how to make climate science datasets and models accessible for machine learning. The focus will be on climate science challenges and opportunities associated with two distinct projects, 1) EnviroNet: a project focused on bridging gaps between geoscience and machine learning research through a global data repository of ImageNet analogs and AI challenges, and 2) a Mila project on changing people's minds and behavior through visualization of future extreme climate events. The discussion related to EnviroNet will be on how datasets and climate science problems can be framed for the machine learning research community at large. The discussion related to the Mila project will include climate science forecast model prototype developments in progress for accurately visualizing future extreme climate impacts of events such as floods, that particularly impact individual's neighborhoods and households.

Abstract 17: **ML vs. Climate Change, Applications in Energy at DeepMind in Climate Change: How Can AI Help?**, *Witherspoon* 03:45 PM

DeepMind has proved that machine learning can help us solve challenges in the Energy sector that contribute to climate change. DeepMind Program Manager Sims Witherspoon will share how they have applied ML to reduce energy consumption in data centers as well as to increase the value of wind power by 20% (compared to a baseline of

no realtime commitments to the grid). Sims will also highlight insights the team has learned in their application of ML to the real-world as well as the potential for these kinds of techniques to be applied in other areas, to help tackle climate change on an even grander scale.

Abstract 18: **Truck Traffic Monitoring with Satellite Images in Climate Change: How Can AI Help?**, *Kaack, Chen* 04:20 PM

The road freight sector is responsible for a large and growing share of greenhouse gas emissions, but reliable data on the amount of freight that is moved on roads in many parts of the world are scarce. Many low- and middle-income countries have limited ground-based traffic monitoring and freight surveying activities. In this proof of concept, we show that we can use an object detection network to count trucks in satellite images and predict average annual daily truck traffic from those counts. We describe a complete model, test the uncertainty of the estimation, and discuss the transfer to developing countries.

Abstract 19: **Machine Learning for AC Optimal Power Flow in Climate Change: How Can AI Help?**, *Guha, Wang* 04:30 PM

We explore machine learning methods for AC Optimal Powerflow (ACOPF) - the task of optimizing power generation in a transmission network according while respecting physical and engineering constraints. We present two formulations of ACOPF as a machine learning problem: 1) an end-to-end prediction task where we directly predict the optimal generator settings, and 2) a constraint prediction task where we predict the set of active constraints in the optimal solution. We validate these approaches on two benchmark grids.

Abstract 20: **Planetary Scale Monitoring of Urban Growth in High Flood Risk Areas in Climate Change: How Can AI Help?**, *Clough, Nair, Erinjippurath* 04:40 PM

Climate change is increasing the incidence of flooding. Many areas in the developing world are experiencing strong population growth but lack adequate urban planning. This represents a significant humanitarian risk. We explore the use of high-cadence satellite imagery provided by Planet, whose flock of over one hundred 'Dove' satellites image the entire earth's landmass everyday at 3-5m resolution. We use a deep learning-based computer vision approach to measure flood-related humanitarian risk in 5 cities in Africa.

## Workshop on the Security and Privacy of Machine Learning

***Nicolas Papernot, Florian Tramer, Bo Li, Dan Boneh, David Evans, Somesh Jha, Percy Liang, Patrick McDaniel, Jacob Steinhardt, Dawn Song***

**104 B, Fri Jun 14, 08:30 AM**

As machine learning has increasingly been deployed in critical real-world applications, the dangers of manipulation and misuse of these models has become of paramount importance to public safety and user privacy. In applications such as online content recognition to financial analytics to autonomous vehicles all have shown the be vulnerable to adversaries wishing to manipulate the models or mislead models to their malicious ends.
This workshop will focus on recent research and future directions about the security and privacy problems in real-world machine learning systems. We aim to bring together experts from machine learning, security, and privacy communities in an attempt to highlight recent work in these area as well as to clarify the foundations of secure and private machine learning strategies. We seek to come to a consensus on a rigorous framework to formulate adversarial attacks targeting machine learning models, and to characterize the properties that ensure the security and privacy of machine learning systems. Finally, we hope to chart out important directions for future work and cross-community collaborations.

## Schedule

09:00 AM **Patrick McDaniel**

09:30 AM **Una-May O'Reilly**

10:00 AM **Enhancing Gradient-based Attacks with Symbolic Intervals**

10:20 AM **Adversarial Policies: Attacking Deep Reinforcement Learning**

10:45 AM **Le Song**

11:15 AM **Allen Qi**

11:45 AM **Private vqSGD: Vector-Quantized Stochastic Gradient Descent**

01:15 PM **Ziko Kolter**

01:45 PM **Provable Certificates for Adversarial Examples:Fitting a Ball in the Union of Polytopes**

02:05 PM **Poster Session #1**

02:45 PM **Alexander Madry**

03:15 PM **Been Kim**

03:45 PM **Theoretically Principled Trade-off between Robustness and Accuracy**

04:05 PM **Model weight theft with just noise inputs: The curious case of the petulant attacker**

04:15 PM **Panel**

05:15 PM **Poster Sesson #2**

## Theoretical Physics for Deep Learning

***Jaehoon Lee, Jeffrey Pennington, Yasaman Bahri, Max Welling, Surya Ganguli, Joan Bruna***

**104 C, Fri Jun 14, 08:30 AM**

Though the purview of physics is broad and includes many loosely connected subdisciplines, a unifying theme is the endeavor to provide concise, quantitative, and predictive descriptions of the often large and complex systems governing phenomena that occur in the natural world. While one could debate how closely deep learning is connected to the natural world, it is undeniably the case that deep learning systems are large and complex; as such, it is reasonable to consider whether the rich body of ideas and powerful tools from theoretical physicists could be harnessed to improve our understanding of deep learning. The goal of this workshop is to investigate this question by bringing together experts in theoretical physics and deep learning in order to stimulate interaction and to begin exploring how theoretical physics can shed light on the theory of deep learning.

We believe ICML is an appropriate venue for this gathering as members from both communities are frequently in attendance and because deep learning theory has emerged as a focus at the conference, both as an independent track in the main conference and in numerous workshops over the last few years. Moreover, the conference has enjoyed an increasing number of papers using physics tools and ideas to draw insights into deep learning.

### Schedule

08:30 AM **Opening Remarks** *Lee, Pennington, Bahri, Welling, Ganguli, Bruna*

08:40 AM **Linearized two-layers neural networks in high dimension** *Montanari*

09:10 AM **Loss landscape and behaviour of algorithms in the spiked matrix-tensor model** *Zdeborova*

09:40 AM **Poster spotlights** *Novak, Dreyer, Golkar, Higgins, Antognini, Karakida, Ghosh*

10:20 AM **Break and poster discussion**

11:00 AM **On the Interplay between Physics and Deep Learning** *Cranmer*

11:30 AM **Why Deep Learning Works: Traditional and Heavy-Tailed Implicit Self-Regularization in Deep Neural Networks** *Mahoney*

12:00 PM **Analyzing the dynamics of online learning in over-parameterized two-layer neural networks** *Goldt*

12:15 PM **Convergence Properties of Neural Networks on Separable Data** *Tachet des Combes*

12:30 PM **Lunch**

02:00 PM **Is Optimization a sufficient language to understand Deep Learning?** *Arora*

02:30 PM **Towards Understanding Regularization in Batch Normalization**

02:45 PM **How Noise during Training Affects the Hessian Spectrum**

03:00 PM **Break and poster discussion**

03:30 PM **Understanding overparameterized neural networks** *Sohl-Dickstein*

04:00 PM **Asymptotics of Wide Networks from Feynman Diagrams** *Gur-Ari*

04:15 PM **A Mean Field Theory of Quantized Deep Networks: The Quantization-Depth Trade-Off** *Gilboa*

04:30 PM **Deep Learning on the 2-Dimensional Ising Model to Extract the Crossover Region** *Walker*

04:45 PM **Learning the Arrow of Time** *Rahaman*

05:00 PM **Poster discussion** *Novak, Gabella, Dreyer, Golkar, Tong, Higgins, Milletari, Antognini, Goldt, Ramírez Rivera, Bondesan, Karakida, Tachet des Combes, Mahoney, Walker, Fort, Smith, Ghosh, Baratin, Granziol, Roberts, Vetrov, Wilson, Laurent, Thomas, Lacoste-Julien, Gilboa, Soudry, Gupta, Goyal, Bengio, Elsen, De, Jastrzebski, Martin, Shabanian, Courville, Akaho, Zdeborova, Dyer, Weiler, de Haan, Cohen, Welling, Luo, peng, Rahaman, Matthey, J. Rezende, Choi, Cranmer, Xiao, Lee, Bahri, Pennington, Yang, Hron, Sohl-Dickstein, Gur-Ari*

Abstracts (9):

Abstract 2: **Linearized two-layers neural networks in high dimension in Theoretical Physics for Deep Learning**, *Montanari* 08:40 AM

Speaker: Andrea Montanari (Stanford)

Abstract: Abstract: We consider the problem of learning an unknown function f on the d-dimensional sphere with respect to the square loss, given i.i.d. samples (y_i,x_i) where x_i is a feature vector uniformly distributed on the sphere and y_i = f(x_i). We study two popular classes of models that can be regarded as linearizations of two-layers neural networks around a random initialization: (RF) The random feature model of Rahimi-Recht; (NT) The neural tangent kernel model of Jacot-Gabriel-Hongler. Both these approaches can also be regarded as randomized approximations of kernel ridge regression (with respect to different kernels), and hence enjoy universal approximation properties when the number of neurons N diverges, for a fixed dimension d.

We prove that, if both d and N are large, the behavior of these models is instead remarkably simpler.
If N is of smaller order than d^2, then RF performs no better than linear regression with respect to the raw features x_i, and NT performs no better than linear regression with respect to degree-one and two monomials in the x_i's. More generally, if N is of smaller order than d^{k+1} then RF fits at most a degree-k polynomial in the raw features, and NT fits at most a degree-(k+ 1) polynomial.
We then focus on the case of quadratic functions, and N= O(d). We show that the gap in generalization error between fully trained neural networks and the linearized models is potentially unbounded.
[based on joint work with Behrooz Ghorbani, Song Mei, Theodor Misiakiewicz]

Abstract 3: **Loss landscape and behaviour of algorithms in the spiked matrix-tensor model in Theoretical Physics for Deep Learning**, *Zdeborova* 09:10 AM

Speaker: Lenka Zdeborova (CEA/SACLAY)

Abstract: A key question of current interest is: How are properties of optimization and sampling algorithms influenced by the properties of the loss function in noisy high-dimensional non-convex settings? Answering this question for deep neural networks is a landmark goal of many ongoing works. In this talk I will answer this question in unprecedented detail for the spiked matrix-tensor model. Information theoretic limits, and Kac-Rice analysis of the loss landscapes, will be compared to the analytically studied performance of message passing algorithms, of the Langevin dynamics and of the gradient flow.

Several rather non-intuitive results will be unveiled and explained.

Abstract 4: **Poster spotlights in Theoretical Physics for Deep Learning**, *Novak, Dreyer, Golkar, Higgins, Antognini, Karakida, Ghosh* 09:40 AM

A Quantum Field Theory of Representation Learning Robert Bamler (University of California at Irvine)*; Stephan Mandt (University of California, Irivine)

Covariance in Physics and Convolutional Neural Networks Miranda Cheng (University of Amsterdam)*; Vassilis Anagiannis (University of Amsterdam); Maurice Weiler (University of Amsterdam); Pim de Haan (University of Amsterdam); Taco S. Cohen (Qualcomm AI Research); Max Welling (University of Amsterdam)

Scale Steerable Filters for Locally Scale-Invariant Convolutional Neural Networks Rohan Ghosh (National University of Singapore)*; Anupam Gupta (National University of Singapore)

Towards a Definition of Disentangled Representations Irina Higgins (DeepMind)*; David Amos (DeepMind); Sebastien Racaniere (DeepMind); David Pfau (); Loic Matthey (DeepMind); Danilo Jimenez Rezende (Google DeepMind)

Bayesian Deep Convolutional Networks with Many Channels are Gaussian Processes Roman Novak (Google Brain)*; Lechao Xiao (Google Brain); Jaehoon Lee (Google Brain); Yasaman Bahri (Google Brain); Greg Yang (Microsoft Research AI); Jiri Hron (University of Cambridge); Daniel Abolafia (Google Brain); Jeffrey Pennington (Google Brain); Jascha Sohl-Dickstein (Google Brain)

Finite size corrections for neural network Gaussian processes Joseph M Antognini (Whisper AI)*

Pathological Spectrum of the Fisher Information Matrix in Deep Neural Networks Ryo Karakida (National Institute of Advanced Industrial Science and Technology)*; Shotaro Akaho (AIST); Shun-ichi Amari (RIKEN)

Inferring the quantum density matrix with machine learning Kyle Cranmer (New York University); Siavash Golkar (NYU)*; Duccio Pappadopulo (Bloomberg)

Jet grooming through reinforcement learning Frederic Dreyer (University of Oxford)*; Stefano Carrazza (University of Milan)

Abstract 5: **Break and poster discussion in Theoretical Physics for Deep Learning**, 10:20 AM

Bayesian Deep Convolutional Networks with Many Channels are Gaussian Processes Roman Novak (Google Brain)*; Lechao Xiao (Google Brain); Jaehoon Lee (Google Brain); Yasaman Bahri (Google Brain); Greg Yang (Microsoft Research AI); Jiri Hron (University of Cambridge); Daniel Abolafia (Google Brain); Jeffrey Pennington (Google Brain); Jascha Sohl-Dickstein (Google Brain)
Topology of Learning in Artificial Neural Networks Maxime Gabella (Magma Learning)*
Jet grooming through reinforcement learning Frederic Dreyer (University of Oxford)*; Stefano Carrazza (University of Milan)
Inferring the quantum density matrix with machine learning Kyle Cranmer (New York University); Siavash Golkar (NYU)*; Duccio Pappadopulo (Bloomberg)
Backdrop: Stochastic Backpropagation Siavash Golkar (NYU)*; Kyle Cranmer (New York University)
Explain pathology in Deep Gaussian Process using Chaos Theory Anh Tong (UNIST)*; Jaesik Choi (Ulsan National Institute of Science and Technology)
Towards a Definition of Disentangled Representations Irina Higgins (DeepMind)*; David Amos (DeepMind); Sebastien Racaniere (DeepMind); David Pfau (DeepMind); Loic Matthey (DeepMind); Danilo Jimenez Rezende (DeepMind)
Towards Understanding Regularization in Batch Normalization Ping Luo (The Chinese University of Hong Kong); Xinjiang Wang (); Wenqi Shao (The Chinese University of HongKong)*; Zhanglin Peng (SenseTime)
Covariance in Physics and Convolutional Neural Networks Miranda Cheng (University of

Amsterdam)*; Vassilis Anagiannis (University of Amsterdam); Maurice Weiler (University of Amsterdam); Pim de Haan (University of Amsterdam); Taco S. Cohen (Qualcomm AI Research); Max Welling (University of Amsterdam)

Meanfield theory of activation functions in Deep Neural Networks Mirco Milletari (Microsoft)*; Thiparat Chotibut (SUTD) ; Paolo E. Trevisanutto (National University of Singapore)

Finite size corrections for neural network Gaussian processes Joseph M Antognini (Whisper AI)*

SWANN: Small-World Neural Networks and Rapid Convergence Mojan Javaheripi (UC San Diego)*; Bita Darvish Rouhani (UC San Diego); Farinaz Koushanfar (UC San Diego)

Analysing the dynamics of online learning in over-parameterised two-layer neural networks Sebastian Goldt (Institut de Physique théorique, Paris)*; Madhu Advani (Harvard University); Andrew Saxe (University of Oxford); Florent Krzakala (École Normale Supérieure); Lenka Zdeborova (CEA Saclay)

A Halo Merger Tree Generation and Evaluation Framework Sandra Robles (Universidad Autónoma de Madrid); Jonathan Gómez (Pontificia Universidad Católica de Chile); Adín Ramírez Rivera (University of Campinas)*; Jenny Gonzáles (Pontificia Universidad Católica de Chile); Nelson Padilla (Pontificia Universidad Católica de Chile); Diego Dujovne (Universidad Diego Portales)

Learning Symmetries of Classical Integrable Systems Roberto Bondesan (Qualcomm AI Research)*, Austen Lamacraft (Cavendish Laboratory, University of Cambridge, UK)

Cosmology inspired generative models Uros Seljak (UC Berkeley)*; Francois Lanusse (UC Berkeley)

Pathological Spectrum of the Fisher Information Matrix in Deep Neural Networks Ryo Karakida (National Institute of Advanced Industrial Science and Technology)*; Shotaro Akaho (AIST); Shun-ichi Amari (RIKEN)

How Noise during Training Affects the Hessian Spectrum Mingwei Wei (Northwestern University); David Schwab (Facebook AI Research)*

A Quantum Field Theory of Representation Learning Robert Bamler (University of California at Irvine)*; Stephan Mandt (University of California, Irivine)

Convergence Properties of Neural Networks on Separable Data Remi Tachet des Combes (Microsoft Research Montreal)*; Mohammad Pezeshki (Mila & University of Montreal); Samira Shabanian (Microsoft, Canada); Aaron Courville (MILA, Université de Montréal); Yoshua Bengio (Mila)

Universality and Capacity Metrics in Deep Neural Networks Michael Mahoney (University of California, Berkeley)*; Charles Martin (Calculation Consulting)

Feynman Diagrams for Large Width Networks Guy Gur-Ari (Google)*; Ethan Dyer (Google)

Deep Learning on the 2-Dimensional Ising Model to Extract the Crossover Region Nicholas Walker (Louisiana State Univ - Baton Rouge)*

Large Scale Structure of the Loss Landscape of Neural Networks Stanislav Fort (Stanford University)*; Stanislaw Jastrzebski (New York University)

Momentum Enables Large Batch Training Samuel L Smith (DeepMind)*; Erich Elsen (Google); Soham De (DeepMind)

Learning the Arrow of Time Nasim Rahaman (University of Heidelberg)*; Steffen Wolf (Heidelberg University); Anirudh Goyal (University of Montreal); Roman Remme (Heidelberg University); Yoshua Bengio (Mila)

Scale Steerable Filters for Locally Scale-Invariant Convolutional Neural Networks Rohan Ghosh (National University of Singapore)*; Anupam Gupta (National University of Singapore)

A Mean Field Theory of Quantized Deep Networks: The Quantization-Depth Trade-Off Yaniv Blumenfeld (Technion)*; Dar Gilboa (Columbia University); Daniel Soudry (Technion)

Rethinking Complexity in Deep Learning: A View from Function Space Aristide Baratin (Mila, Université de Montréal)*; Thomas George (MILA, Université de Montréal); César Laurent (Mila, Université de Montréal); Valentin Thomas (MILA); Guillaume Lajoie (Université de Montréal, Mila); Simon Lacoste-Julien (Mila, Université de Montréal)

The Deep Learning Limit: Negative Neural Network eigenvalues just noise? Diego Granziol (Oxford)*; Stefan Zohren (University of Oxford); Stephen Roberts (Oxford); Dmitry P Vetrov (Higher School of Economics); Andrew Gordon Wilson (Cornell University); Timur Garipov (Samsung AI Center in Moscow)

Gradient descent in Gaussian random fields as a toy model for high-dimensional optimisation Mariano Chouza (Tower Research Capital); Stephen Roberts (Oxford); Stefan Zohren

(University of Oxford)*
Deep Learning for Inverse Problems Abhejit Rajagopal (University of California, Santa Barbara)*; Vincent R Radzicki (University of California, Santa Barbara)

Abstract 6: **On the Interplay between Physics and Deep Learning in Theoretical Physics for Deep Learning**, *Cranmer* 11:00 AM

Speaker: Kyle Cranmer (NYU)

Abstract: The interplay between physics and deep learning is typically divided into two themes. The first is “physics for deep learning”, where techniques from physics are brought to bear on understanding dynamics of learning. The second is “deep learning for physics,” which focuses on application of deep learning techniques to physics problems. I will present a more nuanced view of this interplay with examples of how the structure of physics problems have inspired advances in deep learning and how it yields insights on topics such as inductive bias, interpretability, and causality.

Abstract 7: **Why Deep Learning Works: Traditional and Heavy-Tailed Implicit Self-Regularization in Deep Neural Networks in Theoretical Physics for Deep Learning**, *Mahoney* 11:30 AM

Speaker: Michael Mahoney (ICSI and Department of Statistics, University of California at Berkeley)

Abstract:

Random Matrix Theory (RMT) is applied to analyze the weight matrices of Deep Neural Networks (DNNs), including both production quality, pre-trained models and smaller models trained from scratch. Empirical and theoretical results clearly indicate that the DNN training process itself implicitly implements a form of self-regularization, implicitly sculpting a more regularized energy or penalty landscape. In particular, the empirical spectral density (ESD) of DNN layer matrices displays signatures of traditionally-regularized statistical models, even in the absence of exogenously specifying traditional forms of explicit regularization.

Building on relatively recent results in RMT, most notably its extension to Universality classes of Heavy-Tailed matrices, and applying them to these empirical results, we develop a theory to identify 5+1 Phases of Training, corresponding to increasing amounts of implicit self-regularization. For smaller and/or older DNNs, this implicit self-regularization is like traditional Tikhonov regularization, in that there appears to be a ``size scale'' separating signal from noise. For state-of-the-art DNNs, however, we identify a novel form of heavy-tailed self-regularization, similar to the self-organization seen in the statistical physics of disordered systems. This implicit self-regularization can depend strongly on the many knobs of the training process. In particular, by exploiting the generalization gap phenomena, we demonstrate that we can cause a small model to exhibit all 5+1 phases of training simply by changing the batch size. This demonstrates that---all else being equal---DNN optimization with larger batch sizes leads to less-well implicitly-regularized models, and it provides an explanation for the generalization gap phenomena. Coupled with work on energy landscapes and heavy-tailed spin glasses, it also suggests an explanation of why deep learning works. Joint work with Charles Martin of Calculation Consulting, Inc.

Abstract 11: **Is Optimization a sufficient language to understand Deep Learning? in Theoretical Physics for Deep Learning**, *Arora* 02:00 PM

Speaker: Sanjeev Arora (Princeton/IAS)

Abstract: There is an old debate in neuroscience about whether or not learning has to boil down to optimizing a single cost function. This talk will suggest that even to understand mathematical properties of deep learning, we have to go beyond the conventional view of "optimizing a single cost function". The reason is that phenomena occur along the gradient descent trajectory that are not fully captured in the value of the cost function. I will illustrate briefly with three new results that involve such phenomena:

(i) (joint work with Cohen, Hu, and Luo) How deep matrix factorization solves matrix completion better than classical algorithms

https://arxiv.org/abs/1905.13655

(ii) (joint with Du, Hu, Li, Salakhutdinov, and Wang) How to compute (exactly) with an infinitely wide net ("mean field limit", in physics terms) https://arxiv.org/abs/1904.11955

(iii) (joint with Kuditipudi, Wang, Hu, Lee, Zhang, Li, Ge) Explaining mode-connectivity for real-life deep nets (the phenomenon that low-cost solutions found by gradient descent are interconnected in the parameter space via low-cost paths; see Garipov et al'18 and Draxler et al'18)

Abstract 15: **Understanding overparameterized neural networks in Theoretical Physics for Deep Learning**, *Sohl-Dickstein* 03:30 PM

Speaker: Jascha Sohl-Dickstein (Google Brain)

Abstract: As neural networks become highly overparameterized, their accuracy improves, and their behavior becomes easier to analyze theoretically. I will give an introduction to a rapidly growing body of work which examines the learning dynamics and prior over functions induced by infinitely wide, randomly initialized, neural networks. Core results that I will discuss include: that the distribution over functions computed by a wide neural network often corresponds to a Gaussian process with a particular compositional kernel, both before and after training; that the predictions of wide neural networks are linear in their parameters throughout training; and that this perspective enables analytic predictions for how trainability depends on hyperparameters and architecture. These results provide for surprising capabilities -- for instance, the evaluation of test set predictions which would come from an infinitely wide trained neural network without ever instantiating a neural network, or the rapid training of 10,000+ layer convolutional networks. I will argue that this growing understanding of neural networks in the limit of infinite width is foundational for future theoretical and practical understanding of deep learning.

Abstract 20: **Poster discussion in Theoretical Physics for Deep Learning**, *Novak, Gabella, Dreyer, Golkar, Tong, Higgins, Milletari, Antognini, Goldt, Ramírez Rivera, Bondesan, Karakida, Tachet des Combes, Mahoney, Walker, Fort, Smith, Ghosh, Baratin, Granziol, Roberts, Vetrov, Wilson, Laurent, Thomas, Lacoste-Julien, Gilboa, Soudry, Gupta, Goyal, Bengio, Elsen, De, Jastrzebski, Martin, Shabanian, Courville, Akaho, Zdeborova, Dyer, Weiler, de Haan, Cohen, Welling, Luo, peng, Rahaman, Matthey, J. Rezende, Choi, Cranmer, Xiao, Lee, Bahri, Pennington, Yang, Hron, Sohl-Dickstein, Gur-Ari* 05:00 PM

Bayesian Deep Convolutional Networks with Many Channels are Gaussian Processes Roman Novak (Google Brain)*; Lechao Xiao (Google Brain); Jaehoon Lee (Google Brain); Yasaman Bahri (Google Brain); Greg Yang (Microsoft Research AI); Jiri Hron (University of Cambridge); Daniel Abolafia (Google Brain); Jeffrey Pennington (Google Brain); Jascha Sohl-Dickstein (Google Brain)

Topology of Learning in Artificial Neural Networks Maxime Gabella (Magma Learning)*

Jet grooming through reinforcement learning Frederic Dreyer (University of Oxford)*; Stefano Carrazza (University of Milan)

Inferring the quantum density matrix with machine learning Kyle Cranmer (New York University); Siavash Golkar (NYU)*; Duccio Pappadopulo (Bloomberg)

Backdrop: Stochastic Backpropagation Siavash Golkar (NYU)*; Kyle Cranmer (New York University)

Explain pathology in Deep Gaussian Process using Chaos Theory Anh Tong (UNIST)*; Jaesik Choi (Ulsan National Institute of Science and Technology)

Towards a Definition of Disentangled Representations Irina Higgins (DeepMind)*; David Amos (DeepMind); Sebastien Racaniere (DeepMind); David Pfau (DeepMind); Loic Matthey (DeepMind); Danilo Jimenez Rezende (DeepMind)

Towards Understanding Regularization in Batch

Normalization Ping Luo (The Chinese University of Hong Kong); Xinjiang Wang (); Wenqi Shao (The Chinese University of HongKong)*; Zhanglin Peng (SenseTime)

Covariance in Physics and Convolutional Neural Networks Miranda Cheng (University of Amsterdam)*; Vassilis Anagiannis (University of Amsterdam); Maurice Weiler (University of Amsterdam); Pim de Haan (University of Amsterdam); Taco S. Cohen (Qualcomm AI Research); Max Welling (University of Amsterdam)

Meanfield theory of activation functions in Deep Neural Networks Mirco Milletari (Microsoft)*; Thiparat Chotibut (SUTD) ; Paolo E. Trevisanutto (National University of Singapore)

Finite size corrections for neural network Gaussian processes Joseph M Antognini (Whisper AI)*

Analysing the dynamics of online learning in over-parameterised two-layer neural networks Sebastian Goldt (Institut de Physique théorique, Paris)*; Madhu Advani (Harvard University); Andrew Saxe (University of Oxford); Florent Krzakala (École Normale Supérieure); Lenka Zdeborova (CEA Saclay)

A Halo Merger Tree Generation and Evaluation Framework Sandra Robles (Universidad Autónoma de Madrid); Jonathan Gómez (Pontificia Universidad Católica de Chile); Adín Ramírez Rivera (University of Campinas)*; Jenny Gonzáles (Pontificia Universidad Católica de Chile); Nelson Padilla (Pontificia Universidad Católica de Chile); Diego Dujovne (Universidad Diego Portales)

Learning Symmetries of Classical Integrable Systems Roberto Bondesan (Qualcomm AI Research)*, Austen Lamacraft (Cavendish Laboratory, University of Cambridge, UK)

Pathological Spectrum of the Fisher Information Matrix in Deep Neural Networks Ryo Karakida (National Institute of Advanced Industrial Science and Technology)*; Shotaro Akaho (AIST); Shun-ichi Amari (RIKEN)

How Noise during Training Affects the Hessian Spectrum Mingwei Wei (Northwestern University); David Schwab (Facebook AI Research)*

A Quantum Field Theory of Representation Learning Robert Bamler (University of California at Irvine)*; Stephan Mandt (University of California, Irivine)

Convergence Properties of Neural Networks on Separable Data Remi Tachet des Combes (Microsoft Research Montreal)*; Mohammad Pezeshki (Mila & University of Montreal); Samira Shabanian (Microsoft, Canada); Aaron Courville (MILA, Université de Montréal); Yoshua Bengio (Mila)

Universality and Capacity Metrics in Deep Neural Networks Michael Mahoney (University of California, Berkeley)*; Charles Martin (Calculation Consulting)

Asymptotics of Wide Networks from Feynman Diagrams Guy Gur-Ari (Google)*; Ethan Dyer (Google)

Deep Learning on the 2-Dimensional Ising Model to Extract the Crossover Region Nicholas Walker (Louisiana State Univ - Baton Rouge)*

Large Scale Structure of the Loss Landscape of Neural Networks Stanislav Fort (Stanford University)*; Stanislaw Jastrzebski (New York University)

Momentum Enables Large Batch Training Samuel L Smith (DeepMind)*; Erich Elsen (Google); Soham De (DeepMind)

Learning the Arrow of Time Nasim Rahaman (University of Heidelberg)*; Steffen Wolf (Heidelberg University); Anirudh Goyal (University of Montreal); Roman Remme (Heidelberg University); Yoshua Bengio (Mila)

Scale Steerable Filters for Locally Scale-Invariant Convolutional Neural Networks Rohan Ghosh (National University of Singapore)*; Anupam Gupta (National University of Singapore)

A Mean Field Theory of Quantized Deep Networks: The Quantization-Depth Trade-Off Yaniv Blumenfeld (Technion)*; Dar Gilboa (Columbia University); Daniel Soudry (Technion)

Rethinking Complexity in Deep Learning: A View from Function Space Aristide Baratin (Mila, Université de Montréal)*; Thomas George (MILA, Université de Montréal); César Laurent (Mila, Université de Montréal); Valentin Thomas (MILA); Guillaume Lajoie (Université de Montréal, Mila); Simon Lacoste-Julien (Mila, Université de Montréal)

The Deep Learning Limit: Negative Neural Network eigenvalues just noise? Diego Granziol (Oxford)*; Stefan Zohren (University of Oxford); Stephen Roberts (Oxford); Dmitry P Vetrov (Higher School of Economics); Andrew Gordon Wilson (Cornell University); Timur Garipov (Samsung AI Center in Moscow)

Gradient descent in Gaussian random fields as a toy model for high-dimensional optimisation Mariano Chouza (Tower Research Capital); Stephen Roberts (Oxford); Stefan Zohren (University of Oxford)*

Deep Learning for Inverse Problems Abhejit Rajagopal (University of California, Santa Barbara)*; Vincent R Radzicki (University of California, Santa Barbara)

## AI in Finance: Applications and Infrastructure for Multi-Agent Learning

***Prashant Reddy, Tucker Balch, Michael Wellman, Senthil Kumar, Ion Stoica, Edith Elkind***

**201, Fri Jun 14, 08:30 AM**

Finance is a rich domain for AI and ML research. Model-driven strategies for stock trading and risk assessment models for loan approvals are quintessential financial applications that are reasonably well-understood. However, there are a number of other applications that call for attention as well.

In particular, many finance domains involve ecosystems of interacting and competing agents. Consider for instance the detection of financial fraud and money-laundering. This is a challenging multi-agent learning problem, especially because the real world agents involved evolve their strategies constantly. Similarly, in algorithmic trading of stocks, commodities, etc., the actions of any given trading agent affects, and is affected by, other trading agents -- many of these agents are constantly learning in order to adapt to evolving market scenarios. Further, such trading agents operate at such a speed and scale that they must be fully autonomous. They have grown in sophistication to employ advanced ML strategies including deep learning, reinforcement learning, and transfer learning.

Financial institutions have a long history of investing in technology as a differentiator and have been key drivers in advancing computing infrastructure (e.g., low-latency networking). As more financial applications employ deep learning and reinforcement learning, there is consensus now on the need for more advanced computing architectures--for training large machine learning models and simulating large multi-agent learning systems--that balance scale with the stringent privacy requirements of finance.

Historically, financial firms have been highly secretive about their proprietary technology developments. But now, there is also emerging consensus on the need for (1) deeper engagement with academia to advance a shared knowledge of the unique challenges faced in FinTech, and (2) more open collaboration with academic and technology partners through intellectually sophisticated fora such as this proposed workshop.

### Schedule

| | |
|---|---|
| 09:00 AM | **Opening Remarks** *Reddy, Kumar* |
| 09:10 AM | **Invited Talk 1: Adaptive Tolling for Multiagent Traffic Optimization** *Stone* |
| 09:30 AM | **Invited Talk 2: The Strategic Perils of Learning from Historical Data** *Morgenstern* |
| 09:50 AM | **Oral Paper Presentations 1** |
| 10:30 AM | **Coffee Break and Socialization** |
| 11:00 AM | **Invited Talk 3: Trend-Following Trading Strategies and Financial Market Stability** *Wellman* |

| | |
|---|---|
| 11:20 AM | **Poster Highlights - Lightning Round** |
| 12:00 PM | **Lunch** |
| 02:00 PM | **Invited Talk 4: Towards AI Innovation in the Financial Domain** *Veloso* |
| 02:20 PM | **Oral Paper Presentations 2** |
| 03:00 PM | **Coffee Break and Socialization** |
| 03:30 PM | **Invited Talk 5: Intra-day Stock Price Prediction as a Measure of Market Efficiency** *Balch* |
| 03:50 PM | **Invited Talk 6: RLlib: A Platform for Finance Research** *Stoica* |
| 04:10 PM | **Poster Session - All Accepted Papers** |
| 06:00 PM | **ICML Reception** |

Abstracts (3):

Abstract 4: **Oral Paper Presentations 1 in AI in Finance: Applications and Infrastructure for Multi-Agent Learning**, 09:50 AM

09:50-10:03 Risk-Sensitive Compact Decision Trees for Autonomous Execution in presence of Simulated Market Response, Svitlana Vyetrenko (JP Morgan Chase); Kyle Xu (Georgia Institute of Technology)

10:04-10:16 Robust Trading via Adversarial Reinforcement Learning Thomas Spooner (University of Liverpool); Rahul Savani (Univ. of Liverpool)

10:17-10:30 Generating Realistic Stock Market Order Streams, Junyi Li (University of Michigan); Xintong Wang (University of Michigan); Yaoyang Lin (University of Michigan); Arunesh Sinha (University of Michigan); Michael Wellman (University of Michigan)

Abstract 7: **Poster Highlights - Lightning Round in AI in Finance: Applications and Infrastructure for Multi-Agent Learning**, 11:20 AM

Self Organizing Supply Chains for Micro-Prediction; Present and Future Uses of the ROAR Protocol, Peter D Cotton (JP Morgan Chase)

Learning-Based Trading Strategies in the Face of Market Manipulation, Xintong Wang (University of Michigan); Chris Hoang (University of Michigan); Michael Wellman (University of Michigan)

Multi-Agent Simulation for Pricing and Hedging in a Dealer Market, Sumitra Ganesh (JPMorgan AI Research); Nelson Vadori (JPMorgan AI Research); Mengda Xu (JPMorgan AI Research); Hua Zheng (JPMorgan Chase); Prashant Reddy (JPMorgan AI Research); Manuela Veloso (JPMorgan AI Research)

Multi-Agent Reinforcement Learning for Liquidation Strategy Analysis, Wenhang Bao (Columbia University); Xiao-Yang Liu (Columbia University)

Some people aren't worth listening to: periodically retraining classifiers with feedback from a team of end users, Joshua Lockhart (JPMorgan AI Research); Mahmoud Mahfouz (JPMorgan AI Research); Tucker Balch (JPMorgan AI Research); Manuela Veloso (JPMorgan AI Research)

Optimistic Bull or Pessimistic Bear: Adaptive Deep Reinforcement Learning for Stock Portfolio Allocation, Xinyi Li (Columbia University); Yinchuan Li ( Beijing Institute of Technology); Yuancheng Zhan (University of Science and Technology of China); Xiao-Yang Liu (Columbia University)

How to Evaluate Trading Strategies: Backtesting or Agent-based Simulation?, Tucker Balch (JPMorgan AI Research); David Byrd (Georgia Tech); Mahmoud Mahfouz (JPMorgan AI Research)

Deep Reinforcement Learning for Optimal Trade Execution, Siyu Lin (University of Virginia)

Abstract 10: **Oral Paper Presentations 2 in AI in Finance: Applications and Infrastructure for Multi-Agent Learning**, 02:20 PM

02:20-02:33 Towards Inverse Reinforcement Learning for Limit Order Book Dynamics, Jacobo Roa Vicens (University College London); Cyrine Chtourou (JPMorgan Chase); Angelos Filos

(University of Oxford); Francisco Rullan (University College of London); Yarin Gal (University of Oxford);
Ricardo Silva (University College London)

02:34-02:47 An Agent-Based Model of Financial Benchmark Manipulation, Megan J Shearer (University of Michigan); Gabriel Rauterberg (University of Michigan); Michael Wellman (University of Michigan)

02:47-03:00 The sharp, the flat and the shallow: Can weakly interacting agents learn to escape bad minima?, Panos Parpas (Imperial College London); Nikolas Kantas (Imperial College London); Grigorios Pavliotis (Imperial College London)

## The Third Workshop On Tractable Probabilistic Modeling (TPM)

***Daniel Lowd, Antonio Vergari, Alejandro Molina, Tahrima Rahman, Pedro Domingos, Antonio Vergari***

**202, Fri Jun 14, 08:30 AM**

Probabilistic modeling has become the de facto framework to reason about uncertainty in Machine Learning and AI. One of the main challenges in probabilistic modeling is the trade-off between the expressivity of the models and the complexity of performing various types of inference, as well as learning them from data.

This inherent trade-off is clearly visible in powerful -- but intractable -- models like Markov random fields, (restricted) Boltzmann machines, (hierarchical) Dirichlet processes and Variational Autoencoders. Despite these models' recent successes, performing inference on them resorts to approximate routines. Moreover, learning such models from data is generally harder as inference is a sub-routine of learning, requiring simplifying assumptions or further approximations. Having guarantees on tractability at inference and learning time is then a highly desired property in many real-world scenarios.

Tractable probabilistic modeling (TPM) concerns methods guaranteeing exactly this: performing exact (or tractably approximate) inference and/or learning. To achieve this, the following approaches have been proposed: i) low or bounded-treewidth probabilistic graphical models and determinantal point processes, that exchange expressiveness for efficiency; ii) graphical models with high girth or weak potentials, that provide bounds on the performance of approximate inference methods; and iii) exchangeable probabilistic models that exploit symmetries to reduce inference complexity. More recently, models compiling inference routines into efficient computational graphs such as arithmetic circuits, sum-product networks, cutset networks and probabilistic sentential decision diagrams have advanced the state-of-the-art inference performance by exploiting context-specific independence, determinism or by exploiting latent variables. TPMs have been successfully used in numerous real-world applications: image classification, completion and generation, scene understanding, activity recognition, language and speech modeling, bioinformatics, collaborative filtering, verification and diagnosis of physical systems.

The aim of this workshop is to bring together researchers working on the different fronts of tractable probabilistic modeling, highlighting recent trends and open challenges. At the same time, we want to foster the discussion across similar or complementary sub-fields in the broader probabilistic modeling community. In particular, the rising field of neural probabilistic models, such as normalizing flows and autoregressive models that achieve impressive results in generative modeling. It is an interesting open challenge for the TPM community to keep a broad range of inference routines tractable while leveraging these models' expressiveness. Furthermore, the rising field of probabilistic programming promises to be the new lingua franca of model-based learning. This offers the TPM community opportunities to push the expressiveness of the models used for general-purpose universal probabilistic languages, such as Pyro, while maintaining efficiency.

We want to promote discussions and advance the field both by having high quality contributed works, as well as high level invited speakers coming from the aforementioned tangent sub-fields of probabilistic modeling.

## Schedule

09:00 AM **Welcome**

09:10 AM **Testing Arithmetic Circuits** *Darwiche*

09:50 AM **Poster spotlights**

10:30 AM **Coffee Break**

11:00 AM **Tractable Islands Revisited** *Dechter*

11:40 AM **Poster spotlights**

12:00 PM **Sum-Product Networks and Deep Learning: A Love Marriage** *Peharz*

12:40 PM **Lunch**

02:20 PM **Tensor Variable Elimination in Pyro** *Bingham*

03:00 PM **Coffee Break**

03:30 PM **Invertible Residual Networks and a Novel Perspective on Adversarial Examples** *Jacobsen*

04:10 PM **Poster session**

Abstracts (5):

Abstract 2: **Testing Arithmetic Circuits in The Third Workshop On Tractable Probabilistic Modeling (TPM)**, *Darwiche* 09:10 AM

I will discuss Testing Arithmetic Circuits (TACs), which are new tractable probabilistic models that are universal function approximators like neural networks. A TAC represents a piecewise multilinear function and computes a marginal query on the newly introduced Testing Bayesian Network (TBN). The structure of a TAC is automatically compiled from a Bayesian network and its parameters are learned from labeled data using gradient descent. TACs can incorporate background knowledge that is encoded in the Bayesian network, whether conditional independence or domain constraints. Hence, the behavior of a TAC comes with some guarantees that are invariant to how it is trained from data. Moreover, a TAC is amenable to being interpretable since its nodes and parameters have precise meanings by virtue of being compiled from a Bayesian network. This recent work aims to fuse models (Bayesian networks) and functions (DNNs) with the goal of realizing their collective benefits.

Abstract 5: **Tractable Islands Revisited in The Third Workshop On Tractable Probabilistic Modeling (TPM)**, *Dechter* 11:00 AM

"An important component of human problem-solving expertise is the ability to use knowledge about solving easy problems to guide the solution of difficult ones." - Minsky
A longstanding intuition in AI is that intelligent agents should be able to use solutions to easy problems to solve hard problems. This has often been termed the "tractable island paradigm." How do we act on this intuition in the domain of probabilistic reasoning? This talk will describe the status of probabilistic reasoning algorithms that are driven by the tractable islands paradigm when solving optimization, likelihood and mixed (max-sum-product, e.g. marginal map) queries. I will show how heuristics generated via variational relaxation into tractable structures, can guide heuristic search and Monte-Carlo sampling, yielding anytime solvers that produce approximations with confidence bounds that improve with time, and become exact if enough time is allowed.

Abstract 7: **Sum-Product Networks and Deep Learning: A Love Marriage in The Third Workshop On Tractable Probabilistic Modeling (TPM)**, *Peharz* 12:00 PM

Sum-product networks (SPNs) are a prominent class of tractable probabilistic model, facilitating efficient marginalization, conditioning, and other inference routines. However, despite these attractive properties, SPNs have received rather little attention in the (probabilistic) deep learning community, which rather focuses on intractable models such as generative adversarial networks, variational autoencoders, normalizing flows, and autoregressive density estimators. In this talk, I discuss several recent endeavors which demonstrate that i) SPNs can be effectively used as deep learning models, and ii) that hybrid

learning approaches utilizing SPNs and other deep learning models are in fact sensible and beneficial.

Abstract 9: **Tensor Variable Elimination in Pyro in The Third Workshop On Tractable Probabilistic Modeling (TPM)**, *Bingham* 02:20 PM

A wide class of machine learning algorithms can be reduced to variable elimination on factor graphs. While factor graphs provide a unifying notation for these algorithms, they do not provide a compact way to express repeated structure when compared to plate diagrams for directed graphical models. In this talk, I will describe a generalization of undirected factor graphs to plated factor graphs, and a corresponding generalization of the variable elimination algorithm that exploits efficient tensor algebra in graphs with plates of variables. This tensor variable elimination algorithm has been integrated into the Pyro probabilistic programming language, enabling scalable, automated exact inference in a wide variety of deep generative models with repeated discrete latent structure. I will discuss applications of such models to polyphonic music modeling, animal movement modeling, and unsupervised word-level sentiment analysis, as well as algorithmic applications to exact subcomputations in approximate inference and ongoing work on extensions to continuous latent variables.

Abstract 11: **Invertible Residual Networks and a Novel Perspective on Adversarial Examples in The Third Workshop On Tractable Probabilistic Modeling (TPM)**, *Jacobsen* 03:30 PM

In this talk, I will discuss how state-of-the-art discriminative deep networks can be turned into likelihood-based density models. Further, I will discuss how such models give rise to an alternative viewpoint on adversarial examples. Under this viewpoint adversarial examples are a consequence of excessive invariances learned by the classifier, manifesting themselves in striking failures when evaluating the model on out of distribution inputs. I will discuss how the commonly used cross-entropy objective encourages such overly invariant representations. Finally, I will present an extension to cross-entropy that, by exploiting properties of invertible deep networks, enables control of erroneous invariances in theory and practice.

## Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)

***Sujith Ravi, Zornitsa Kozareva, Lixin Fan, Max Welling, Yurong Chen, Werner Bailer, Brian Kulis, Haoji Hu, Jonathan Dekhtiar, Yingyan Lin, Diana Marculescu***

**203, Fri Jun 14, 08:30 AM**

This joint workshop aims to bring together researchers, educators, practitioners who are interested in techniques as well as applications of on-device machine learning and compact, efficient neural network representations. One aim of the workshop discussion is to establish close connection between researchers in the machine learning community and engineers in industry, and to benefit both academic researchers as well as industrial practitioners. The other aim is the evaluation and comparability of resource-efficient machine learning methods and compact and efficient network representations, and their relation to particular target platforms (some of which may be highly optimized for neural network inference). The research community has still to develop established evaluation procedures and metrics.

The workshop also aims at reproducibility and comparability of methods for compact and efficient neural network representations, and on-device machine learning. Contributors are thus encouraged to make their code available. The workshop organizers plan to make some example tasks and datasets available, and invite contributors to use them for testing their work. In order to provide comparable performance evaluation conditions, the use of a common platform (such as Google Colab) is intended.

## Schedule

| | |
|---|---|
| 08:30 AM | **Welcome and Introduction** |
| 08:40 AM | **Hardware Efficiency Aware Neural Architecture Search and Compression** *Han* |
| 09:10 AM | **Structured matrices for efficient deep learning** *Kumar* |
| 09:40 AM | **DeepCABAC: Context-adaptive binary arithmetic coding for deep neural network compression** *Wiedemann* |
| 10:00 AM | **Poster spotlight presentations** |
| 10:30 AM | **Coffee Break AM** |
| 11:00 AM | **Understanding the Challenges of Algorithm and Hardware Co-design for Deep Neural Networks** *Sze* |
| 11:30 AM | **Dream Distillation: A Data-Independent Model Compression Framework** *Bhardwaj* |
| 11:50 AM | **The State of Sparsity in Deep Neural Networks** *Gale* |
| 12:10 PM | **Lunch break** |
| 12:40 PM | **Poster session** *Hao, Lin, Li, Ruthotto, Yang, Karkada* |
| 02:00 PM | **DNN Training and Inference with Hyper-Scaled Precision** *Gopalakrishnan* |
| 02:30 PM | **Mixed Precision Training & Inference** *Dekhtiar* |
| 03:00 PM | **Coffee Break PM** |
| 03:30 PM | **Learning Compact Neural Networks Using Ordinary Differential Equations as Activation Functions** |
| 03:50 PM | **Triplet Distillation for Deep Face Recognition** |
| 04:10 PM | **Single-Path NAS: Device-Aware Efficient ConvNet Design** *Stamoulis* |
| 04:30 PM | **Panel discussion** |
| 05:30 PM | **Wrap-up and Closing** |

Abstracts (10):

Abstract 4: **DeepCABAC: Context-adaptive binary arithmetic coding for deep neural network compression in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, *Wiedemann* 09:40 AM

Simon Wiedemann, Heiner Kirchhoffer, Stefan Matlage, Paul Haase, Arturo Marban Gonzalez, Talmaj Marinc, Heiko Schwarz, Detlev Marpe, Thomas Wiegand, Ahmed Osman and Wojciech Samek

http://arxiv.org/abs/1905.08318

Abstract 5: **Poster spotlight presentations in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, 10:00 AM

2min presentations of the posters presentations during lunch break

Abstract 7: **Understanding the Challenges of Algorithm and Hardware Co-design for Deep Neural Networks in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, *Sze* 11:00 AM

The co-design of algorithm and hardware has become an increasingly important approach for addressing the computational complexity of Deep Neural Networks (DNNs). There are several open problems and challenges in the co-design process and application; for instance, what metrics should be used to drive the algorithm design, how to automate the process in a simple way, how to extend these approaches to tasks beyond image classification, and how to design flexible hardware to support these different approaches. In this talk, we highlight recent and ongoing work that aim to address these challenges, namely energy-aware pruning and NetAdapt that automatically incorporate direct metrics such as latency and energy into the training and design of the DNN; FastDepth that extends the co-design approaches to a depth estimation task; and a flexible hardware accelerator called Eyeriss v2 that is computationally efficient across a wide range of diverse DNNs.

BIO: Vivienne Sze is an Associate Professor at MIT in the Electrical Engineering and Computer Science Department. Her research interests include energy-aware signal processing algorithms, and low-power circuit and system design for portable multimedia applications, including computer vision, deep learning, autonomous navigation, and video process/coding. Prior to joining MIT, she was a Member of Technical Staff in the R&D Center at TI, where she designed low-power algorithms and architectures for video coding. She also represented TI in the JCT-VC committee of ITU-T and ISO/IEC standards body during the development of High Efficiency Video Coding (HEVC), which received a Primetime Engineering Emmy Award. She is a co-editor of the book entitled "High Efficiency Video Coding (HEVC): Algorithms and Architectures" (Springer, 2014).
Prof. Sze received the B.A.Sc. degree from the University of Toronto in 2004, and the S.M. and Ph.D. degree from MIT in 2006 and 2010, respectively. In 2011, she received the Jin-Au Kong Outstanding Doctoral Thesis Prize in Electrical Engineering at MIT. She is a recipient of the 2019 Edgerton Faculty Award, the 2018 Facebook Faculty Award, the 2018 & 2017 Qualcomm Faculty Award, the 2018 & 2016 Google Faculty Research Award, the 2016 AFOSR Young Investigator Research Program (YIP) Award, the 2016 3M Non-Tenured Faculty Award, the 2014 DARPA Young Faculty Award, the 2007 DAC/ISSCC Student Design Contest Award, and a co-recipient of the 2017 CICC Outstanding Invited Paper Award, the 2016 IEEE Micro Top Picks Award and the 2008 A-SSCC Outstanding Design Award.
For more information about research in the Energy-Efficient Multimedia Systems Group at MIT visit: http://www.rle.mit.edu/eems/

Abstract 8: **Dream Distillation: A Data-Independent Model Compression Framework in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, *Bhardwaj* 11:30 AM

Kartikeya Bhardwaj, Naveen Suda and Radu Marculescu

http://arxiv.org/abs/1905.07072

Abstract 9: **The State of Sparsity in Deep Neural Networks in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, *Gale* 11:50 AM

Trevor Gale, Erich Elsen and Sara Hooker

https://arxiv.org/abs/1902.09574 (to be updated)

Abstract 11: **Poster session in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, *Hao, Lin, Li, Ruthotto, Yang, Karkada* 12:40 PM

Xiaofan Zhang, Hao Cong, Yuhong Li, Yao Chen, Jinjun Xiong, Wen-Mei Hwu and Deming Chen. A Bi-Directional Co-Design Approach to Enable Deep Learning on IoT Devices
https://arxiv.org/abs/1905.08369

Kushal Datta, Aishwarya Bhandare, Deepthi Karkada, Vamsi Sripathi, Sun Choi, Vikram Saletore and Vivek Menon. Efficient 8-Bit Quantization of Transformer Neural Machine Language Translation Model

Bin Yang, Lin Yang, Xiaochun Li, Wenhan Zhang, Hua Zhou, Yequn Zhang, Yongxiong Ren and Yinbo Shi. 2-bit Model Compression of Deep Convolutional Neural Network on ASIC Engine for image retrieval
https://arxiv.org/abs/1905.03362

Zhong Qiu Lin, Brendan Chwyl and Alexander Wong. EdgeSegNet: A Compact Network for Semantic Segmentation
https://arxiv.org/abs/1905.04222

Sheng Lin, Xiaolong Ma, Shaokai Ye, Geng Yuan, Kaisheng Ma and Yanzhi Wang. Toward Extremely Low Bit and Lossless Accuracy in DNNs with Progressive ADMM
https://arxiv.org/abs/1905.00789

Chengcheng Li, Zi Wang, Dali Wang, Xiangyang Wang and Hairong Qi. Investigating Channel Pruning through Structural Redundancy Reduction - A Statistical Study

https://arxiv.org/abs/1905.06498

Wei Niu, Yanzhi Wang and Bin Ren. CADNN: Ultra Fast Execution of DNNs on Mobile Devices with Advanced Model Compression and Architecture-Aware Optimization
https://arxiv.org/abs/1905.00571

Jonathan Ephrath, Lars Ruthotto, Eldad Haber and Eran Treister. LeanResNet: A Low-cost yet Effective Convolutional Residual Networks
https://arxiv.org/abs/1904.06952

Dushyant Mehta, Kwang In Kim and Christian Theobalt. Implicit Filter Sparsification In Convolutional Neural Networks
https://arxiv.org/abs/1905.04967

Abstract 12: **DNN Training and Inference with Hyper-Scaled Precision in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, *Gopalakrishnan* 02:00 PM

Kailash Gopalakrishnan

Abstract 15: **Learning Compact Neural Networks Using Ordinary Differential Equations as Activation Functions in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, 03:30 PM

Mohamadali Torkamani, Phillip Wallis, Shiv Shankar and Amirmohammad Rooshenas

https://arxiv.org/abs/1905.07685

Abstract 16: **Triplet Distillation for Deep Face Recognition in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, 03:50 PM

Yushu Feng, Huan Wang and Haoji Hu

https://arxiv.org/abs/1905.04457

Abstract 17: **Single-Path NAS: Device-Aware Efficient ConvNet Design in Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations (ODML-CDNNR)**, *Stamoulis* 04:10 PM

Dimitrios Stamoulis, Ruizhou Ding, Di Wang, Dimitrios Lymberopoulos, Bodhi Priyantha, Jie Liu and Diana Marculescu

https://arxiv.org/abs/1905.04159

## Negative Dependence: Theory and Applications in Machine Learning

***Mike Gartrell, Jennifer Gillenwater, Alex Kulesza, Zelda Mariet***

**204, Fri Jun 14, 08:30 AM**

Models of negative dependence are increasingly important in machine learning. Whether selecting training data, finding an optimal experimental design, exploring in reinforcement learning, or making suggestions with recommender systems, selecting high-quality but diverse items has become a core challenge. This workshop aims to bring together researchers who, using theoretical or applied techniques, leverage negative dependence in their work. We will delve into the rich underlying mathematical theory, understand key applications, and discuss the most promising directions for future research.

## Schedule

| | |
|---|---|
| 08:45 AM | **Opening Remarks** |
| 08:50 AM | **Victor-Emmanuel Brunel: Negative Association and Discrete Determinantal Point Processes** *Brunel* |
| 09:30 AM | **Aarti Singh: Experimental Design** *Singh* |
| 10:10 AM | **On Two Ways to use Determinantal Point Processes for Monte Carlo Integration** *Gautier* |
| 10:30 AM | **[Coffee Break]** |

| | |
|---|---|
| 11:00 AM | **Jeff Bilmes: Deep Submodular Synergies** *Bilmes* |
| 11:40 AM | **Submodular Batch Selection for Training Deep Neural Networks** *N Balasubramanian* |
| 12:00 PM | **[Lunch Break]** |
| 02:00 PM | **Michal Valko: How Negative Dependence Broke the Quadratic Barrier for Learning with Graphs and Kernels** *Valko* |
| 02:40 PM | **Exact Sampling of Determinantal Point Processes with Sublinear Time Preprocessing** *Derezinski* |
| 03:00 PM | **[Coffee Break]** |
| 03:30 PM | **Sergei Levine: Distribution Matching and Mutual Information in Reinforcement Learning** *Levine* |
| 04:10 PM | **Seq2Slate: Re-ranking and Slate Optimization with RNNs** *Meshi* |
| 04:30 PM | **[Mini Break]** |
| 04:40 PM | **Cheng Zhang: Active Mini-Batch Sampling using Repulsive Point Processes** *Zhang* |
| 05:20 PM | **Gaussian Process Optimization with Adaptive Sketching: Scalable and No Regret** *Valko* |
| 05:40 PM | **Towards Efficient Evaluation of Risk via Herding** *Xu* |
| 06:00 PM | **Closing Remarks** |

Abstracts (12):

Abstract 2: **Victor-Emmanuel Brunel: Negative Association and Discrete Determinantal Point Processes in Negative Dependence: Theory and Applications in Machine Learning**, *Brunel* 08:50 AM

Discrete Determinantal Point Processes (DPPs) form a class of probability distributions that can describe the random selection of items from a finite or countable collection. They naturally arise in many problems in probability theory, and they have gained a lot of attention in machine learning, due to both their modeling flexibility and their tractability. In the finite case, a DPP is parametrized by a matrix, whose principal minors are the weights given by the DPP to each possible subset of items. When the matrix is symmetric, the DPP has a very special property, called Negative Association. Thanks to this property, symmetric DPPs enforce diversity within the randomly selected items, which is a feature that is sought for in many applications of Machine Learning, such as recommendation systems.

Abstract 3: **Aarti Singh: Experimental Design in Negative Dependence: Theory and Applications in Machine Learning**, *Singh* 09:30 AM

TBD

Abstract 4: **On Two Ways to use Determinantal Point Processes for Monte Carlo Integration in Negative Dependence: Theory and Applications in Machine Learning**, *Gautier* 10:10 AM

This paper focuses on Monte Carlo integration with determinantal point processes (DPPs) which enforce negative dependence between quadrature nodes. We survey the properties of two unbiased Monte Carlo estimators of the integral of inter- est: a direct one proposed by Bardenet & Hardy (2016) and a less obvious 60-year-old estimator by Ermakov & Zolotukhin (1960) that actually also relies on DPPs. We provide an efficient implementation to sample exactly a particular multidimensional DPP called multivariate Jacobi ensemble. This let us investigate the behavior of both estimators on toy problems in yet unexplored regimes.

Abstract 6: **Jeff Bilmes: Deep Submodular Synergies in Negative Dependence: Theory and Applications in Machine Learning**, *Bilmes* 11:00 AM

Submodularity is an attractive framework in machine learning to model concepts such as diversity, dispersion, and cooperative costs, and is having an ever increasing impact on the field of machine learning. Deep learning is having a bit of success as well. In this talk, we will discuss synergies, where submodular functions and deep neural networks can be used together to their

mutual benefit. First, we'll discuss deep submodular functions (DSFs), an expressive class of functions that include many widely used submodular functions and that are defined analogously to deep neural networks (DNN). We'll show that the class of DSFs strictly increases with depth and discuss applications. Second, we'll see how a modification to DNN autoencoders can produce features that can be used in DSFs. These DSF/DNN hybrids address an open problem which is how best to produce a submodular function for your application. Third, we'll see how submodular functions can speed up the training of models. In one case, submodularity can be used to produce a sequence of mini-batches that speeds up training of DNN systems. In another case, submodularity can produce a training data subset for which we can show faster convergence to the optimal solution in the convex case. Empirically, this method speeds up gradient methods by up to 10x for convex and 3x for non-convex (i.e., deep) functions.

The above discusses various projects that were performed jointly with Wenruo Bai, Shengjie Wang, Chandrashekhar Lavania, Baharan Mirzasoleiman, and Jure Leskovec.

Abstract 7: **Submodular Batch Selection for Training Deep Neural Networks in Negative Dependence: Theory and Applications in Machine Learning**, *N Balasubramanian* 11:40 AM

Mini-batch gradient descent based methods are the de facto algorithms for training neural network architectures today. We introduce a mini-batch selection strategy based on submodular function maximization. Our novel submodular formulation captures the informativeness of each sample and diversity of the whole subset. We design an efficient, greedy algorithm which can give high-quality solutions to this NP-hard combinatorial optimization problem. Our extensive experiments on standard datasets show that the deep models trained using the proposed batch selection strategy provide better generalization than Stochastic Gradient Descent as well as a popular baseline sampling strategy across different learning rates, batch sizes, and distance metrics.

Abstract 9: **Michal Valko: How Negative Dependence Broke the Quadratic Barrier for Learning with Graphs and Kernels in Negative Dependence: Theory and Applications in Machine Learning**, *Valko* 02:00 PM

As we advance with resources, we move from reasoning on entities to reasoning on pairs and groups. We have beautiful frameworks: graphs, kernels, DPPs. However, the methods that work with pairs, relationships, and similarity are just slow. Kernel regression, or second-order gradient methods, or sampling from DPPs do not scale to large data, because of the costly construction and storing of matrix K_n. Prior work showed that sampling points according to their ridge leverage scores (RLS) generates small dictionaries with strong spectral approximation guarantees for K_n. However, computing exact RLS requires building and storing the whole kernel matrix. In this talk, we start with SQUEAK, a new online approach for kernel approximations based on RLS sampling that sequentially processes the data, storing a dictionary with a number of points that only depends on the effective dimension d_eff(gamma) of the dataset. The beauty of negative dependence, that we estimate on the fly, makes it possible to exclude huge portions of dictionary. With the small dictionary, SQUEAK never constructs the whole matrix K_n, runs in linear time O(n*d_eff(gamma)^3) w.r.t. n, and requires only a single pass over the dataset. A distributed version of SQUEAK runs in as little as O(log(n)*d_eff(gamma)^3) time. This online tool opens out a range of possibilities to finally have scalable, adaptive, and provably accurate kernel methods: semi-supervised learning or Laplacian smoothing on large graphs, scalable GP-UCB, efficient second-order kernel online learning, and even fast DPP sampling, some of these being featured in this workshop.

Abstract 10: **Exact Sampling of Determinantal Point Processes with Sublinear Time Preprocessing in Negative Dependence: Theory and Applications in Machine Learning**, *Derezinski* 02:40 PM

We study the complexity of sampling from a distribution over all index subsets of the set {1,...,n} with the probability of a subset S proportional to the determinant of the submatrix L_S of some n x n p.s.d. matrix L, where L_S corresponds to the entries of L indexed by S. Known as a determinantal point process, this distribution is widely used in machine learning to induce diversity in subset selection. In practice, we often wish to sample multiple subsets S with small expected size k = E[|S|] << n from a very large matrix L, so it is important to minimize the preprocessing cost of the procedure (performed once) as well as the sampling cost (performed repeatedly). To that end, we propose a new algorithm which, given access to L, samples exactly from a determinantal point process while satisfying the following two properties: (1) its preprocessing cost is n x poly(k) (sublinear in the size of L) and (2) its sampling cost is poly(k) (independent of the size of L). Prior to this work, state-of-the-art exact samplers required O(n^3) preprocessing time and sampling time linear in n or dependent on the spectral properties of L.

Abstract 12: **Sergei Levine: Distribution Matching and Mutual Information in Reinforcement Learning in Negative Dependence: Theory and Applications in Machine Learning**, *Levine* 03:30 PM

Conventionally, reinforcement learning is considered to be a framework for optimization: the aim for standard reinforcement learning algorithms is to recover an optimal or near-optimal policy that maximizes the reward over time. However, when considering more advanced reinforcement learning problems, from inverse reinforcement learning to unsupervised and hierarchical reinforcement learning, we often encounter settings where it is desirable to learn policies that match target distributions over trajectories or states, covering all modes, or else to simply learn collections of behaviors that are as broad and varied as possible. Information theory and probabilistic inference offer is a powerful set of tools for developing algorithms for these kinds of distribution matching problems. In this talk, I will outline methods that combine reinforcement learning, inference, and information theory to learn policies that match target distributions and acquire diverse behaviors, and discuss the applications of such methods for a variety of problems in artificial intelligence and robotics.

Abstract 13: **Seq2Slate: Re-ranking and Slate Optimization with RNNs in Negative Dependence: Theory and Applications in Machine Learning**, *Meshi* 04:10 PM

Ranking is a central task in machine learning and information retrieval. In this task, it is especially important to present the user with a slate of items that is appealing as a whole. This in turn requires taking into account interactions between items, since intuitively, placing an item on the slate affects the decision of which other items should be placed alongside it. In this work, we propose a sequence-to-sequence model for ranking called seq2slate. At each step, the model predicts the next "best" item to place on the slate given the items already selected. The sequential nature of the model allows complex dependencies between the items to be captured directly in a flexible and scalable way. We show how to learn the model end-to-end from weak supervision in the form of easily obtained click-through data. We further demonstrate the usefulness of our approach in experiments on standard ranking benchmarks as well as in a real-world recommendation system.

Abstract 15: **Cheng Zhang: Active Mini-Batch Sampling using Repulsive Point Processes in Negative Dependence: Theory and Applications in Machine Learning**, *Zhang* 04:40 PM

We explore active mini-batch selection using repulsive point processes for stochastic gradient descent (SGD). Our approach simultaneously introduces active bias and leads to stochastic gradients with lower variance. We show theoretically and empirically that our approach improves over standard SGD both in terms of convergence speed as well as final model performance.

Abstract 16: **Gaussian Process Optimization with Adaptive Sketching: Scalable and No**

**Regret in Negative Dependence: Theory and Applications in Machine Learning**, *Valko* 05:20 PM

Gaussian processes (GP) are a popular Bayesian approach for the optimization of black-box functions. Despite their effectiveness in simple problems, GP-based algorithms hardly scale to complex high-dimensional functions, as their per-iteration time and space cost is at least quadratic in the number of dimensions $d$ and iterations~$t$. Given a set of $A$ alternative to choose from, the overall runtime $O(t^3A)$ quickly becomes prohibitive. In this paper, we introduce BKB (budgeted kernelized bandit), an approximate GP algorithm for optimization under bandit feedback that achieves near-optimal regret (and hence near-optimal convergence rate) with near-constant per-iteration complexity and no assumption on the input space or the GP's covariance.

Combining a kernelized linear bandit algorithm (GP-UCB) with randomized matrix sketching technique (i.e., leverage score sampling), we prove that selecting inducing points based on their posterior variance gives an accurate low-rank approximation of the GP, preserving variance estimates and confidence intervals. As a consequence, BKB does not suffer from variance starvation, an important problem faced by many previous sparse GP approximations. Moreover, we show that our procedure selects at most $ \widetilde{O}(d_{eff})$ points, where $d_{eff}$ is the \emph{effective} dimension of the explored space, which is typically much smaller than both $d$ and $t$. This greatly reduces the dimensionality of the
problem, thus leading to a $\widetilde{O}(TAd_{eff}^2)$ runtime and $\widetilde{O}(A d_{eff})$ space complexity.

Abstract 17: **Towards Efficient Evaluation of Risk via Herding in Negative Dependence: Theory and Applications in Machine Learning**, *Xu* 05:40 PM

We introduce a novel use of herding to address the problem of selecting samples from a large unlabeled dataset to efficiently evaluate the risk of a given model. Herding is an algorithm which elaborately draws samples to approximate the underlying distribution. We use herding to select the most informative samples and show that the loss evaluated on $k$ samples produced by herding converges to the expected loss at a rate $\mathcal{O}(1/k)$, which is much faster than $ \mathcal{O}(1/\sqrt{k})$ for iid random sampling. We validate our analysis on both synthetic data and real data, and further explore the empirical performance of herding-based sampling in different cases of high-dimensional data.

## Understanding and Improving Generalization in Deep Learning

***Dilip Krishnan, Hossein Mobahi, Behnam Neyshabur, Peter Bartlett, Dawn Song, Nati Srebro***

**Grand Ballroom A, Fri Jun 14, 08:30 AM**

The 1st workshop on Generalization in Deep Networks: Theory and Practice will be held as part of ICML 2019. Generalization is one of the fundamental problems of machine learning, and increasingly important as deep networks make their presence felt in domains with big, small, noisy or skewed data. This workshop will consider generalization from both theoretical and practical perspectives. We welcome contributions from paradigms such as representation learning, transfer learning and reinforcement learning. The workshop invites researchers to submit working papers in the following research areas:

Implicit regularization: the role of optimization algorithms in generalization
Explicit regularization methods
Network architecture choices that improve generalization
Empirical approaches to understanding generalization
Generalization bounds; empirical evaluation criteria to evaluate bounds
Robustness: generalizing to distributional shift a.k.a dataset shift
Generalization in the context of representation learning, transfer learning and deep reinforcement learning: definitions and empirical approaches

## Schedule

08:30 AM **Opening Remarks**

08:40 AM **Keynote by Dan Roy: Progress on Nonvacuous Generalization Bounds** *Roy*

09:20 AM **Keynote by Chelsea Finn: Training for Generalization** *Finn*

09:50 AM **A Meta-Analysis of Overfitting in Machine Learning**

10:05 AM **Uniform convergence may be unable to explain generalization in deep learning**

10:20 AM **Break and Poster Session**

10:40 AM **Keynote by Sham Kakade: Prediction, Learning, and Memory** *Kakade*

11:10 AM **Keynote by Mikhail Belkin: A Hard Look at Generalization and its Theories** *Belkin*

11:40 AM **Towards Task and Architecture-Independent Generalization Gap Predictors**

11:55 AM **Data-Dependent Sample Complexity of Deep Neural Networks via Lipschitz Augmentation**

12:10 PM **Lunch and Poster Session**

01:30 PM **Keynote by Aleksander Mądry: Are All Features Created Equal?** *Madry*

02:00 PM **Keynote by Jason Lee: On the Foundations of Deep Learning: SGD, Overparametrization, and Generalization** *Lee*

02:30 PM **Towards Large Scale Structure of the Loss Landscape of Neural Networks**

02:45 PM **Zero-Shot Learning from scratch: leveraging local compositional representations**

03:00 PM **Break and Poster Session**

03:30 PM **Panel Discussion (Nati Srebro, Dan Roy, Chelsea Finn, Mikhail Belkin, Aleksander Mądry, Jason Lee)** *Srebro, Roy, Finn, Belkin, Madry, Lee*

04:30 PM **Overparameterization without Overfitting: Jacobian-based Generalization Guarantees for Neural Networks**

04:45 PM **How Learning Rate and Delay Affect Minima Selection in AsynchronousTraining of Neural Networks: Toward Closing the Generalization Gap**

05:00 PM **Poster Session**

Abstracts (18):

Abstract 2: **Keynote by Dan Roy: Progress on Nonvacuous Generalization Bounds in Understanding and Improving Generalization in Deep Learning**, *Roy* 08:40 AM

Generalization bounds are one of the main tools available for explaining the performance of learning algorithms. At the same time, most bounds in the literature are loose to an extent that raises the question as to whether these bounds actually have any explanatory power in the nonasymptotic regime of actual machine learning practice. I'll report on progress towards developing bounds and techniques---both statistical and computational---aimed at closing the gap between empirical performance and theoretical understanding.

Bio: Daniel Roy is an Assistant Professor in the Department of Statistical Sciences and, by courtesy, Computer Science at the University of Toronto, and a founding faculty member of the Vector Institute for Artificial Intelligence. Daniel is a recent recipient of an Ontario Early Researcher Award and Google Faculty Research Award. Before joining U of T, Daniel held a Newton International Fellowship from the Royal Academy of Engineering and a Research Fellowship at Emmanuel College, University of Cambridge. Daniel earned his S.B., M.Eng., and Ph.D. from the Massachusetts Institute of Technology: his dissertation on probabilistic programming won an MIT EECS Sprowls Dissertation Award. Daniel's group works on foundations of machine learning and statistics.

Abstract 3: **Keynote by Chelsea Finn: Training for Generalization in Understanding and Improving Generalization in Deep Learning**, *Finn* 09:20 AM

TBA.

Bio: Chelsea Finn is a research scientist at Google Brain, a post-doc at Berkeley AI Research Lab (BAIR), and will join the Stanford Computer Science faculty in Fall 2019. Finn's research studies how new algorithms can enable machines to acquire intelligent behavior through learning and interaction, allowing them to perform a variety of complex sensorimotor skills in real-world settings. She has developed deep learning algorithms for concurrently learning visual perception and control in robotic manipulation skills, inverse reinforcement methods for scalable acquisition of nonlinear reward functions, and meta-learning algorithms that can enable fast, few-shot adaptation in both visual perception and deep reinforcement learning. Finn's research has been recognized through an NSF graduate fellowship, the C.V. Ramamoorthy Distinguished Research Award, and the Technology Review 35 Under 35 Award, and her work has been covered by various media outlets, including the New York Times, Wired, and Bloomberg. With Sergey Levine and John Schulman, she also designed and taught a course on deep reinforcement learning, with thousands of followers online.

Finn received a PhD in Computer Science from UC Berkeley and a S.B. in Electrical Engineering and Computer Science from MIT.

Abstract 4: **A Meta-Analysis of Overfitting in Machine Learning in Understanding and Improving Generalization in Deep Learning**, 09:50 AM

Authors: Sara Fridovich-Keil, Moritz Hardt, John Miller, Ben Recht, Rebecca Roelofs, Ludwig Schmidt and Vaishaal Shankar

Abstract: We conduct the first large meta-analysis of overfitting due to test set reuse in the machine learning community. Our analysis is based on over one hundred machine learning competitions hosted on the Kaggle platform over the course of several years. In each competition, numerous practitioners repeatedly evaluated their progress against a holdout set that forms the basis of a public ranking available throughout the competition. Performance on a separate test set used only once determined the final ranking. By systematically comparing the public ranking with the final ranking, we assess how much participants adapted to the holdout set over the course of a competition. Our longitudinal study shows, somewhat surprisingly, little evidence of substantial overfitting. These findings speak to the robustness of the holdout method across different data domains, loss functions, model classes, and human analysts.

Abstract 5: **Uniform convergence may be unable to explain generalization in deep learning in Understanding and Improving Generalization in Deep Learning**, 10:05 AM

Authors: Vaishnavh Nagarajan and J. Zico Kolter

Abstract: We cast doubt on the power of uniform convergence-based generalization bounds to provide a complete picture of why overparameterized deep networks generalize well. While it is well-known that many existing bounds are numerically large, through a variety of experiments, we first bring to light another crucial and more concerning aspect of these bounds: in practice, these bounds can {\em increase} with the dataset size. Guided by our observations, we then present examples of overparameterized linear classifiers and neural networks trained by stochastic gradient descent (SGD) where uniform convergence provably cannot ``explain generalization,'' even if we take into account implicit regularization {\em to the fullest extent possible}. More precisely, even if we consider only the set of classifiers output by SGD that have test errors less than some small $ \epsilon$, applying (two-sided) uniform convergence on this set of classifiers yields a generalization guarantee that is larger than $1- \epsilon$ and is therefore nearly vacuous.

Abstract 6: **Break and Poster Session in Understanding and Improving Generalization in Deep Learning**, 10:20 AM

1- Uniform convergence may be unable to explain generalization in deep learning. Vaishnavh Nagarajan and J. Zico Kolter
2- The effects of optimization on generalization in infinitely wide neural networks. Anastasia Borovykh
3- Generalized Capsule Networks with Trainable Routing Procedure. Zhenhua Chen, Chuhua Wang, David Crandall and Tiancong Zhao
4- Implicit Regularization of Discrete Gradient Dynamics in Deep Linear Neural Networks. Gauthier Gidel, Francis Bach and Simon Lacoste-Julien
5- Stable Rank Normalization for Improved Generalization in Neural Networks. Amartya Sanyal, Philip H Torr and Puneet K Dokania
6- On improving deep learning generalization with adaptive sparse connectivity. Shiwei Liu, Decebal Constantin Mocanu and Mykola Pechenizkiy
7- Identity Connections in Residual Nets Improve Noise Stability. Shuzhi Yu and Carlo Tomasi
8- Factors for the Generalisation of Identity Relations by Neural Networks. Radha Manisha Kopparti and Tillman Weyde
9- Output-Constrained Bayesian Neural Networks. Wanqian Yang, Lars Lorch, Moritz A. Graule, Srivatsan Srinivasan, Anirudh Suresh, Jiayu Yao, Melanie F. Pradier and Finale Doshi-Velez
10- An Empirical Study on Hyperparameters and their Interdependence for RL Generalization. Xingyou Song, Yilun Du and Jacob Jackson
11- Towards Large Scale Structure of the Loss Landscape of Neural Networks. Stanislav Fort and Stanislaw Jastrzebski
12- Detecting Extrapolation with Influence Functions. David Madras, James Atwood and Alex D'Amour
13- Towards Task and Architecture-Independent Generalization Gap Predictors. Scott Yak, Hanna Mazzawi and Javier Gonzalvo
14- SGD Picks a Stable Enough Trajectory. Stanisław Jastrzębski and Stanislav Fort
15- MazeNavigator: A Customisable 3D Benchmark for Assessing Generalisation in Reinforcement Learning. Luke Harries, Sebastian Lee, Jaroslaw Rzepecki, Katja Hofmann and Sam Devlin
16- Utilizing Eye Gaze to Enhance the Generalization of Imitation Network to Unseen Environments. Congcong Liu, Yuying Chen, Lei Tai, Ming Liu and Bertram Shi
17- Investigating Under and Overfitting in Wasserstein Generative Adversarial Networks. Ben Adlam, Charles Weill and Amol Kapoor
18- An Empirical Evaluation of Adversarial Robustness under Transfer Learning. Todor Davchev, Timos Korres, Stathi Fotiadis, Nick Antonopoulos and Subramanian Ramamoorthy
19- On Adversarial Robustness of Small vs Large Batch Training. Sandesh Kamath, Amit Deshpande and K V Subrahmanyam
20- The Principle of Unchanged Optimality in Reinforcement Learning Generalization. Xingyou Song and Alex Irpan
21- On the Generalization Capability of Memory Networks for Reasoning. Monireh Ebrahimi, Md Kamruzzaman Sarker, Federico Bianchi, Ning Xie, Aaron Eberhart, Derek Doran and Pascal Hitzler
22- Visualizing How Embeddings Generalize. Xiaotong Liu, Hong Xuan, Zeyu Zhang, Abby Stylianou and Robert Pless
23- Theoretical Analysis of the Fixup Initialization for Fast Convergence and High Generalization Ability. Yasutaka Furusho and Kazushi Ikeda
24- Data-Dependent Sample Complexity of Deep Neural Networks via Lipschitz Augmentation. Colin Wei and Tengyu Ma
25- Few-Shot Transfer Learning From Multiple Pre-Trained Networks. Joshua Ka-Wing Lee, Prasanna Sattigeri and Gregory Wornell
26- Uniform Stability and High Order Approximation of SGLD in Non-Convex Learning. Mufan Li and Maxime Gazeau
27- Better Generalization with Adaptive Adversarial Training. Amit Despande, Sandesh Kamath and K V Subrahmanyam
28- Adversarial Training Generalizes Spectral Norm Regularization. Kevin Roth, Yannic Kilcher and Thomas Hofmann
29- A Causal View on Robustness of Neural Networks. Cheng Zhang and Yingzhen Li
30- Improving PAC-Bayes bounds for neural networks using geometric properties of the training method. Anirbit Mukherjee, Dan Roy, Pushpendre Rastogi and Jun Yang
31- An Analysis of the Effect of Invariance on Generalization in Neural Networks. Clare Lyle, Marta Kwiatkowska and Yarin Gal
32- Data-Dependent Mututal Information Bounds for SGLD. Jeffrey Negrea, Daniel Roy, Gintare Karolina Dziugaite, Mahdi Haghifam and Ashish Khisti
33- Comparing normalization in conditional computation tasks. Vincent Michalski, Vikram Voleti, Samira Ebrahimi Kahou, Anthony Ortiz,

Pascal Vincent, Chris Pal and Doina Precup
34- Weight and Batch Normalization implement Classical Generalization Bounds. Andrzej Banburski, Qianli Liao, Brando Miranda, Lorenzo Rosasco, Jack Hidary and Tomaso Poggio
35- Increasing batch size through instance repetition improves generalization. Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler and Daniel Soudry
36- Zero-Shot Learning from scratch: leveraging local compositional representations. Tristan Sylvain, Linda Petrini and Devon Hjelm
37- Circuit-Based Intrinsic Methods to Detect Overfitting. Satrajit Chatterjee and Alan Mishchenko
38- Dimension Reduction Approach for Interpretability of Sequence to SequenceRecurrent Neural Networks. Kun Su and Eli Shlizerman
39- Tight PAC-Bayesian generalization error bounds for deep learning. Guillermo Valle Perez, Chico Q. Camargo and Ard A. Louis
40- How Learning Rate and Delay Affect Minima Selection in AsynchronousTraining of Neural Networks: Toward Closing the Generalization Gap. Niv Giladi, Mor Shpigel Nacson, Elad Hoffer and Daniel Soudry
41- Making Convolutional Networks Shift-Invariant Again. Richard Zhang
42- A Meta-Analysis of Overfitting in Machine Learning. Sara Fridovich-Keil, Moritz Hardt, John Miller, Ben Recht, Rebecca Roelofs, Ludwig Schmidt and Vaishaal Shankar
43- Kernelized Capsule Networks. Taylor Killian, Justin Goodwin, Olivia Brown and Sung Son
44- Model similarity mitigates test set overuse. Moritz Hardt, Horia Mania, John Miller, Ben Recht and Ludwig Schmidt
45- Understanding Generalization of Deep Neural Networks Trained with Noisy Labels. Wei Hu, Zhiyuan Li and Dingli Yu
46- Domainwise Classification Network for Unsupervised Domain Adaptation. Seonguk Seo, Yumin Suh, Bohyung Han, Taeho Lee, Tackgeun You, Woong-Gi Chang and Suha Kwak
47- The Generalization-Stability Tradeoff in Neural Network Pruning. Brian Bartoldson, Ari Morcos, Adrian Barbu and Gordon Erlebacher
48- Information matrices and generalization. Valentin Thomas, Fabian Pedregosa, Bart van Merriënboer, Pierre-Antoine Manzagol, Yoshua Bengio and Nicolas Le Roux
49- Adaptively Preconditioned Stochastic Gradient Langevin Dynamics. Chandrasekaran Anirudh Bhardwaj
50- Additive or Concatenating Skip-connections Improve Data Separability. Yasutaka Furusho and Kazushi Ikeda
51- On the Inductive Bias of Neural Tangent Kernels. Alberto Bietti and Julien Mairal
52- PAC Bayes Bound Minimization via Kronecker Normalizing Flows. Chin-Wei Huang, Ahmed Touati, Pascal Vincent, Gintare Karolina Dziugaite, Alexandre Lacoste and Aaron Courville
53- SGD on Neural Networks Learns Functions of Increasing Complexity. Preetum Nakkiran, Gal Kaplun, Dimitris Kalimeris, Tristan Yang, Ben Edelman, Fred Zhang and Boaz Barak
54- Overparameterization without Overfitting: Jacobian-based Generalization Guarantees for Neural Networks. Samet Oymak, Mingchen Li, Zalan Fabian and Mahdi Soltanolkotabi
55- Incorrect gradients and regularization: a perspective of loss landscapes. Mehrdad Yazdani
56- Divide and Conquer: Leveraging Intermediate Feature Representations for Quantized Training of Neural Networks. Ahmed Youssef, Prannoy Pilligundla and Hadi Esmaeilzadeh
57- SinReQ: Generalized Sinusoidal Regularization for Low-Bitwidth Deep Quantized Training. Ahmed Youssef, Prannoy Pilligundla and Hadi Esmaeilzadeh
58- Natural Adversarial Examples. Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt and Dawn Song
59- On the Properties of the Objective Landscapes and Generalization of Gradient-Based Meta-Learning. Simon Guiroy, Vikas Verma and Christopher Pal
60- Angular Visual Hardness. Beidi Chen, Weiyang Liu, Animesh Garg, Zhiding Yu, Anshumali Shrivastava and Animashree Anandkumar
61- Luck Matters: Understanding Training Dynamics of Deep ReLU Networks. Yuandong Tian, Tina Jiang, Qucheng Gong and Ari Morcos
62- Understanding of Generalization in Deep Learning via Tensor Methods. Jingling Li, Yanchao Sun, Ziyin Liu, Taiji Suzuki and Furong Huang
63- Learning from Rules Performs as Implicit Regularization. Hossein Hosseini, Ramin Moslemi, Ali Hooshmand and Ratnesh Sharma
64- Stochastic Mirror Descent on Overparameterized Nonlinear Models: Convergence, Implicit Regularization, and Generalization. Navid Azizan, Sahin Lale and Babak Hassibi

65- Scaling Characteristics of Sequential Multitask Learning: Networks Naturally Learn to Learn. Guy Davidson and Michael Mozer
66- Size-free generalization bounds for convolutional neural networks. Phillip Long and Hanie Sedghi

Abstract 7: **Keynote by Sham Kakade: Prediction, Learning, and Memory in Understanding and Improving Generalization in Deep Learning**, *Kakade* 10:40 AM

Building accurate language models that capture meaningful long-term dependencies is a core challenge in language processing. We consider the problem of predicting the next observation given a sequence of past observations, specifically focusing on the question of how to make accurate predictions that explicitly leverage long-range dependencies. Empirically, and perhaps surprisingly, we show that state-of-the-art language models, including LSTMs and Transformers, do not capture even basic properties of natural language: the entropy rates of their generations drift dramatically upward over time. We also provide provable methods to mitigate this phenomenon: specifically, we provide a calibration-based approach to improve an estimated model based on any measurable long-term mismatch between the estimated model and the true underlying generative distribution. More generally, we will also present fundamental information theoretic and computational limits of sequential prediction with a memory.

Bio: Sham Kakade is a Washington Research Foundation Data Science Chair, with a joint appointment in the Department of Computer Science and the Department of Statistics at the University of Washington. He works on the theoretical foundations of machine learning, focusing on designing provable and practical statistically and computationally efficient algorithms. Amongst his contributions, with a diverse set of collaborators, are: establishing principled approaches in reinforcement learning (including the natural policy gradient, conservative policy iteration, and the PAC-MDP framework); optimal algorithms in the stochastic and non-stochastic multi-armed bandit problems (including the widely used linear bandit and the Gaussian process bandit models); computationally and statistically efficient tensor decomposition methods for estimation of latent variable models (including estimation of mixture of Gaussians, latent Dirichlet allocation, hidden Markov models, and overlapping communities in social networks); faster algorithms for large scale convex and nonconvex optimization (including how to escape from saddle points efficiently). He is the recipient of the IBM Goldberg best paper award (in 2007) for contributions to fast nearest neighbor search and the best paper, INFORMS Revenue Management and Pricing Section Prize (2014). He has been program chair for COLT 2011.

Sham completed his Ph.D. at the Gatsby Computational Neuroscience Unit at University College London, under the supervision of Peter Dayan, and he was a postdoc at the Dept. of Computer Science, University of Pennsylvania , under the supervision of Michael Kearns. Sham was an undergraduate at Caltech , studying in physics under the supervision of John Preskill. Sham has been a Principal Research Scientist at Microsoft Research, New England, an associate professor at the Department of Statistics, Wharton, UPenn, and an assistant professor at the Toyota Technological Institute at Chicago.

Abstract 8: **Keynote by Mikhail Belkin: A Hard Look at Generalization and its Theories in Understanding and Improving Generalization in Deep Learning**, *Belkin* 11:10 AM

"A model with zero training error is overfit to the training data and will typically generalize poorly" goes statistical textbook wisdom. Yet in modern practice over-parametrized deep networks with near perfect (interpolating) fit on training data still show excellent test performance. This fact is difficult to reconcile with most modern theories of generalization that rely on bounding the difference between the empirical and expected error. Indeed, as we will discuss, bounds of that type cannot be expected to explain generalization of interpolating models. I will proceed to show how classical and modern models can be unified within a new "double descent" risk curve that extends the usual U-

shaped bias-variance trade-off curve beyond the point of interpolation. This curve delimits the regime of applicability of classical bounds and the regime where new analyses are required. I will give examples of first theoretical analyses in that modern regime and discuss the (considerable) gaps in our knowledge. Finally I will briefly discuss some implications for optimization.

Bio: Mikhail Belkin is a Professor in the departments of Computer Science and Engineering and Statistics at the Ohio State University. He received a PhD in mathematics from the University of Chicago in 2003. His research focuses on understanding the fundamental structure in data, the principles of recovering these structures and their computational, mathematical and statistical properties. This understanding, in turn, leads to algorithms for dealing with real-world data. His work includes algorithms such as Laplacian Eigenmaps and Manifold Regularization based on ideas of classical differential geometry, which have been widely used for analyzing non-linear high-dimensional data. He has done work on spectral methods, Gaussian mixture models, kernel methods and applications. Recently his work has been focussed on understanding generalization and optimization in modern over-parametrized machine learning. Prof. Belkin is a recipient of an NSF Career Award and a number of best paper and other awards and has served on the editorial boards of the Journal of Machine Learning Research and IEEE PAMI.

Abstract 9: **Towards Task and Architecture-Independent Generalization Gap Predictors in Understanding and Improving Generalization in Deep Learning**, 11:40 AM

Authors: Scott Yak, Hanna Mazzawi and Javier Gonzalvo

Abstract: Can we use deep learning to predict when deep learning works? Our results suggest the affirmative. We created a dataset by training 13,500 neural networks with different architectures, on different variations of spiral datasets, and using different optimization parameters. We used this dataset to train task-independent and architecture-independent generalization gap predictors for those neural networks. We extend Jiang et al.(2018) to also use DNNs and RNNs and show that they outperform the linear model, obtaining R^2=0.965. We also show results for architecture-independent, task-independent, and out-of-distribution generalization gap prediction tasks. Both DNNs and RNNs consistently and significantly outperform linear models, with RNNs obtaining R^2=0.584.

Abstract 10: **Data-Dependent Sample Complexity of Deep Neural Networks via Lipschitz Augmentation in Understanding and Improving Generalization in Deep Learning**, 11:55 AM

Authors: Colin Wei and Tengyu Ma

Abstract: Existing Rademacher complexity bounds for neural networks rely only on norm control of the weight matrices and depend exponentially on depth via a product of the matrix norms. Lower bounds show that this exponential dependence on depth is unavoidable when no additional properties of the training data are considered. We suspect that this conundrum comes from the fact that these bounds depend on the training data only through the margin. In practice, many data-dependent techniques such as Batchnorm improve the generalization performance. For feedforward neural nets as well as RNNs, we obtain tighter Rademacher complexity bounds by considering additional data-dependent properties of the network: the norms of the hidden layers of the network, and the norms of the Jacobians of each layer with respect to the previous layers. Our bounds scale polynomially in depth when these empirical quantities are small, as is usually the case in practice. To obtain these bounds, we develop general tools for augmenting a sequence of functions to make their composition Lipschitz and then covering the augmented functions. Inspired by our theory, we directly regularize the network's Jacobians during training and empirically demonstrate that this improves test performance.

Abstract 11: **Lunch and Poster Session in Understanding and Improving Generalization in Deep Learning**, 12:10 PM

1- Uniform convergence may be unable to explain generalization in deep learning. Vaishnavh Nagarajan and J. Zico Kolter
2- The effects of optimization on generalization in infinitely wide neural networks. Anastasia Borovykh
3- Generalized Capsule Networks with Trainable Routing Procedure. Zhenhua Chen, Chuhua Wang, David Crandall and Tiancong Zhao
4- Implicit Regularization of Discrete Gradient Dynamics in Deep Linear Neural Networks. Gauthier Gidel, Francis Bach and Simon Lacoste-Julien
5- Stable Rank Normalization for Improved Generalization in Neural Networks. Amartya Sanyal, Philip H Torr and Puneet K Dokania
6- On improving deep learning generalization with adaptive sparse connectivity. Shiwei Liu, Decebal Constantin Mocanu and Mykola Pechenizkiy
7- Identity Connections in Residual Nets Improve Noise Stability. Shuzhi Yu and Carlo Tomasi
8- Factors for the Generalisation of Identity Relations by Neural Networks. Radha Manisha Kopparti and Tillman Weyde
9- Output-Constrained Bayesian Neural Networks. Wanqian Yang, Lars Lorch, Moritz A. Graule, Srivatsan Srinivasan, Anirudh Suresh, Jiayu Yao, Melanie F. Pradier and Finale Doshi-Velez
10- An Empirical Study on Hyperparameters and their Interdependence for RL Generalization. Xingyou Song, Yilun Du and Jacob Jackson
11- Towards Large Scale Structure of the Loss Landscape of Neural Networks. Stanislav Fort and Stanislaw Jastrzebski
12- Detecting Extrapolation with Influence Functions. David Madras, James Atwood and Alex D'Amour
13- Towards Task and Architecture-Independent Generalization Gap Predictors. Scott Yak, Hanna Mazzawi and Javier Gonzalvo
14- SGD Picks a Stable Enough Trajectory. Stanisław Jastrzębski and Stanislav Fort
15- MazeNavigator: A Customisable 3D Benchmark for Assessing Generalisation in Reinforcement Learning. Luke Harries, Sebastian Lee, Jaroslaw Rzepecki, Katja Hofmann and Sam Devlin
16- Utilizing Eye Gaze to Enhance the Generalization of Imitation Network to Unseen Environments. Congcong Liu, Yuying Chen, Lei Tai, Ming Liu and Bertram Shi
17- Investigating Under and Overfitting in Wasserstein Generative Adversarial Networks. Ben Adlam, Charles Weill and Amol Kapoor
18- An Empirical Evaluation of Adversarial Robustness under Transfer Learning. Todor Davchev, Timos Korres, Stathi Fotiadis, Nick Antonopoulos and Subramanian Ramamoorthy
19- On Adversarial Robustness of Small vs Large Batch Training. Sandesh Kamath, Amit Deshpande and K V Subrahmanyam
20- The Principle of Unchanged Optimality in Reinforcement Learning Generalization. Xingyou Song and Alex Irpan
21- On the Generalization Capability of Memory Networks for Reasoning. Monireh Ebrahimi, Md Kamruzzaman Sarker, Federico Bianchi, Ning Xie, Aaron Eberhart, Derek Doran and Pascal Hitzler
22- Visualizing How Embeddings Generalize. Xiaotong Liu, Hong Xuan, Zeyu Zhang, Abby Stylianou and Robert Pless
23- Theoretical Analysis of the Fixup Initialization for Fast Convergence and High Generalization Ability. Yasutaka Furusho and Kazushi Ikeda
24- Data-Dependent Sample Complexity of Deep Neural Networks via Lipschitz Augmentation. Colin Wei and Tengyu Ma
25- Few-Shot Transfer Learning From Multiple Pre-Trained Networks. Joshua Ka-Wing Lee, Prasanna Sattigeri and Gregory Wornell
26- Uniform Stability and High Order Approximation of SGLD in Non-Convex Learning. Mufan Li and Maxime Gazeau
27- Better Generalization with Adaptive Adversarial Training. Amit Despande, Sandesh Kamath and K V Subrahmanyam
28- Adversarial Training Generalizes Spectral Norm Regularization. Kevin Roth, Yannic Kilcher and Thomas Hofmann
29- A Causal View on Robustness of Neural Networks. Cheng Zhang and Yingzhen Li
30- Improving PAC-Bayes bounds for neural networks using geometric properties of the training method. Anirbit Mukherjee, Dan Roy, Pushpendre Rastogi and Jun Yang
31- An Analysis of the Effect of Invariance on Generalization in Neural Networks. Clare Lyle, Marta Kwiatkowska and Yarin Gal
32- Data-Dependent Mututal Information Bounds for SGLD. Jeffrey Negrea, Daniel Roy, Gintare Karolina Dziugaite, Mahdi Haghifam and Ashish

Khisti
33- Comparing normalization in conditional computation tasks. Vincent Michalski, Vikram Voleti, Samira Ebrahimi Kahou, Anthony Ortiz, Pascal Vincent, Chris Pal and Doina Precup
34- Weight and Batch Normalization implement Classical Generalization Bounds. Andrzej Banburski, Qianli Liao, Brando Miranda, Lorenzo Rosasco, Jack Hidary and Tomaso Poggio
35- Increasing batch size through instance repetition improves generalization. Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler and Daniel Soudry
36- Zero-Shot Learning from scratch: leveraging local compositional representations. Tristan Sylvain, Linda Petrini and Devon Hjelm
37- Circuit-Based Intrinsic Methods to Detect Overfitting. Satrajit Chatterjee and Alan Mishchenko
38- Dimension Reduction Approach for Interpretability of Sequence to SequenceRecurrent Neural Networks. Kun Su and Eli Shlizerman
39- Tight PAC-Bayesian generalization error bounds for deep learning. Guillermo Valle Perez, Chico Q. Camargo and Ard A. Louis
40- How Learning Rate and Delay Affect Minima Selection in AsynchronousTraining of Neural Networks: Toward Closing the Generalization Gap. Niv Giladi, Mor Shpigel Nacson, Elad Hoffer and Daniel Soudry
41- Making Convolutional Networks Shift-Invariant Again. Richard Zhang
42- A Meta-Analysis of Overfitting in Machine Learning. Sara Fridovich-Keil, Moritz Hardt, John Miller, Ben Recht, Rebecca Roelofs, Ludwig Schmidt and Vaishaal Shankar
43- Kernelized Capsule Networks. Taylor Killian, Justin Goodwin, Olivia Brown and Sung Son
44- Model similarity mitigates test set overuse. Moritz Hardt, Horia Mania, John Miller, Ben Recht and Ludwig Schmidt
45- Understanding Generalization of Deep Neural Networks Trained with Noisy Labels. Wei Hu, Zhiyuan Li and Dingli Yu
46- Domainwise Classification Network for Unsupervised Domain Adaptation. Seonguk Seo, Yumin Suh, Bohyung Han, Taeho Lee, Tackgeun You, Woong-Gi Chang and Suha Kwak
47- The Generalization-Stability Tradeoff in Neural Network Pruning. Brian Bartoldson, Ari Morcos, Adrian Barbu and Gordon Erlebacher
48- Information matrices and generalization. Valentin Thomas, Fabian Pedregosa, Bart van Merriënboer, Pierre-Antoine Manzagol, Yoshua Bengio and Nicolas Le Roux
49- Adaptively Preconditioned Stochastic Gradient Langevin Dynamics. Chandrasekaran Anirudh Bhardwaj
50- Additive or Concatenating Skip-connections Improve Data Separability. Yasutaka Furusho and Kazushi Ikeda
51- On the Inductive Bias of Neural Tangent Kernels. Alberto Bietti and Julien Mairal
52- PAC Bayes Bound Minimization via Kronecker Normalizing Flows. Chin-Wei Huang, Ahmed Touati, Pascal Vincent, Gintare Karolina Dziugaite, Alexandre Lacoste and Aaron Courville
53- SGD on Neural Networks Learns Functions of Increasing Complexity. Preetum Nakkiran, Gal Kaplun, Dimitris Kalimeris, Tristan Yang, Ben Edelman, Fred Zhang and Boaz Barak
54- Overparameterization without Overfitting: Jacobian-based Generalization Guarantees for Neural Networks. Samet Oymak, Mingchen Li, Zalan Fabian and Mahdi Soltanolkotabi
55- Incorrect gradients and regularization: a perspective of loss landscapes. Mehrdad Yazdani
56- Divide and Conquer: Leveraging Intermediate Feature Representations for Quantized Training of Neural Networks. Ahmed Youssef, Prannoy Pilligundla and Hadi Esmaeilzadeh
57- SinReQ: Generalized Sinusoidal Regularization for Low-Bitwidth Deep Quantized Training. Ahmed Youssef, Prannoy Pilligundla and Hadi Esmaeilzadeh
58- Natural Adversarial Examples. Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt and Dawn Song
59- On the Properties of the Objective Landscapes and Generalization of Gradient-Based Meta-Learning. Simon Guiroy, Vikas Verma and Christopher Pal
60- Angular Visual Hardness. Beidi Chen, Weiyang Liu, Animesh Garg, Zhiding Yu, Anshumali Shrivastava and Animashree Anandkumar
61- Luck Matters: Understanding Training Dynamics of Deep ReLU Networks. Yuandong Tian, Tina Jiang, Qucheng Gong and Ari Morcos
62- Understanding of Generalization in Deep Learning via Tensor Methods. Jingling Li, Yanchao Sun, Ziyin Liu, Taiji Suzuki and Furong Huang
63- Learning from Rules Performs as Implicit Regularization. Hossein Hosseini, Ramin Moslemi, Ali Hooshmand and Ratnesh Sharma
64- Stochastic Mirror Descent on

Overparameterized Nonlinear Models: Convergence, Implicit Regularization, and Generalization. Navid Azizan, Sahin Lale and Babak Hassibi
65- Scaling Characteristics of Sequential Multitask Learning: Networks Naturally Learn to Learn. Guy Davidson and Michael Mozer
66- Size-free generalization bounds for convolutional neural networks. Phillip Long and Hanie Sedghi

Abstract 12: **Keynote by Aleksander Mądry: Are All Features Created Equal? in Understanding and Improving Generalization in Deep Learning**, *Madry* 01:30 PM

TBA.

Bio: Aleksander Mądry is an Associate Professor of Computer Science in the MIT EECS Department and a principal investigator in the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL). He received his PhD from MIT in 2011 and, prior to joining the MIT faculty, he spent some time at Microsoft Research New England and on the faculty of EPFL.

Aleksander's research interests span algorithms, continuous optimization, science of deep learning and understanding machine learning from a robustness perspective.

Abstract 13: **Keynote by Jason Lee: On the Foundations of Deep Learning: SGD, Overparametrization, and Generalization in Understanding and Improving Generalization in Deep Learning**, *Lee* 02:00 PM

Deep Learning has had phenomenal empirical successes in many domains including computer vision, natural language processing, and speech recognition. To consolidate and boost the empirical success, we need to develop a more systematic and deeper understanding of the elusive principles of deep learning. In this talk, I will provide analysis of several elements of deep learning including non-convex optimization, overparametrization, and generalization error. First, we show that gradient descent and many other algorithms are guaranteed to converge to a local minimizer of the loss. For several interesting problems including the matrix completion problem, this guarantees that we converge to a global minimum. Then we will show that gradient descent converges to a global minimizer for deep overparametrized networks. Finally, we analyze the generalization error by showing that a subtle combination of SGD, logistic loss, and architecture combine to promote large margin classifiers, which are guaranteed to have low generalization error. Together, these results show that on overparametrized deep networks SGD finds solution of both low train and test error.

Bio: Jason Lee is an assistant professor in Data Sciences and Operations at the University of Southern California. Prior to that, he was a postdoctoral researcher at UC Berkeley working with Michael Jordan. Jason received his PhD at Stanford University advised by Trevor Hastie and Jonathan Taylor. His research interests are in statistics, machine learning, and optimization. Lately, he has worked on high dimensional statistical inference, analysis of non-convex optimization algorithms, and theory for deep learning.

Abstract 14: **Towards Large Scale Structure of the Loss Landscape of Neural Networks in Understanding and Improving Generalization in Deep Learning**, 02:30 PM

Authors: Stanislav Fort and Stanislaw Jastrzebski

Abstract: There are many surprising and perhaps counter-intuitive properties of optimization of deep neural networks. We propose and experimentally verify a unified phenomenological model of the loss landscape that incorporates many of them. Our core idea is to model the loss landscape as a set of high dimensional \emph{sheets} that together form a distributed, large-scale, inter-connected structure. For instance, we predict an existence of low loss subspaces connecting a set (not only a pair) of solutions, and verify it experimentally. We conclude by showing that hyperparameter choices such as learning rate, batch size, dropout and $L_2$ regularization, affect the path optimizer takes through the landscape in a similar way.

Abstract 15: **Zero-Shot Learning from scratch: leveraging local compositional representations in Understanding and Improving Generalization in Deep Learning**, 02:45 PM

Authors: Tristan Sylvain, Linda Petrini and Devon Hjelm

Abstract: Zero-shot classification is a task focused on generalization where no instance from the target classes is seen during training. To allow for test-time transfer, each class is annotated with semantic information, commonly in the form of attributes or text descriptions. While classical zero-shot learning does not specify how this problem should be solved, the most successful approaches rely on features extracted from encoders pre-trained on large datasets, most commonly Imagenet. This approach raises important questions that might otherwise distract researchers from answering fundamental questions about representation learning and generalization. For instance, one should wonder to what extent these methods actually learn representations robust with respect to the task, rather than simply exploiting information stored in the encoder. To remove these distractors, we propose a more challenging setting: Zero-Shot Learning from scratch, which effectively forbids the use encoders fine-tuned on other datasets. Our analysis on this setting highlights the importance of local information, and compositional representations.

Abstract 16: **Break and Poster Session in Understanding and Improving Generalization in Deep Learning**, 03:00 PM

1- Uniform convergence may be unable to explain generalization in deep learning. Vaishnavh Nagarajan and J. Zico Kolter
2- The effects of optimization on generalization in infinitely wide neural networks. Anastasia Borovykh
3- Generalized Capsule Networks with Trainable Routing Procedure. Zhenhua Chen, Chuhua Wang, David Crandall and Tiancong Zhao
4- Implicit Regularization of Discrete Gradient Dynamics in Deep Linear Neural Networks. Gauthier Gidel, Francis Bach and Simon Lacoste-Julien
5- Stable Rank Normalization for Improved Generalization in Neural Networks. Amartya Sanyal, Philip H Torr and Puneet K Dokania
6- On improving deep learning generalization with adaptive sparse connectivity. Shiwei Liu, Decebal Constantin Mocanu and Mykola Pechenizkiy
7- Identity Connections in Residual Nets Improve Noise Stability. Shuzhi Yu and Carlo Tomasi
8- Factors for the Generalisation of Identity Relations by Neural Networks. Radha Manisha Kopparti and Tillman Weyde
9- Output-Constrained Bayesian Neural Networks. Wanqian Yang, Lars Lorch, Moritz A. Graule, Srivatsan Srinivasan, Anirudh Suresh, Jiayu Yao, Melanie F. Pradier and Finale Doshi-Velez
10- An Empirical Study on Hyperparameters and their Interdependence for RL Generalization. Xingyou Song, Yilun Du and Jacob Jackson
11- Towards Large Scale Structure of the Loss Landscape of Neural Networks. Stanislav Fort and Stanislaw Jastrzebski
12- Detecting Extrapolation with Influence Functions. David Madras, James Atwood and Alex D'Amour
13- Towards Task and Architecture-Independent Generalization Gap Predictors. Scott Yak, Hanna Mazzawi and Javier Gonzalvo
14- SGD Picks a Stable Enough Trajectory. Stanisław Jastrzębski and Stanislav Fort
15- MazeNavigator: A Customisable 3D Benchmark for Assessing Generalisation in Reinforcement Learning. Luke Harries, Sebastian Lee, Jaroslaw Rzepecki, Katja Hofmann and Sam Devlin
16- Utilizing Eye Gaze to Enhance the Generalization of Imitation Network to Unseen Environments. Congcong Liu, Yuying Chen, Lei Tai, Ming Liu and Bertram Shi
17- Investigating Under and Overfitting in Wasserstein Generative Adversarial Networks. Ben Adlam, Charles Weill and Amol Kapoor
18- An Empirical Evaluation of Adversarial Robustness under Transfer Learning. Todor Davchev, Timos Korres, Stathi Fotiadis, Nick Antonopoulos and Subramanian Ramamoorthy
19- On Adversarial Robustness of Small vs Large Batch Training. Sandesh Kamath, Amit Deshpande and K V Subrahmanyam
20- The Principle of Unchanged Optimality in Reinforcement Learning Generalization. Xingyou

Song and Alex Irpan
21- On the Generalization Capability of Memory Networks for Reasoning. Monireh Ebrahimi, Md Kamruzzaman Sarker, Federico Bianchi, Ning Xie, Aaron Eberhart, Derek Doran and Pascal Hitzler
22- Visualizing How Embeddings Generalize. Xiaotong Liu, Hong Xuan, Zeyu Zhang, Abby Stylianou and Robert Pless
23- Theoretical Analysis of the Fixup Initialization for Fast Convergence and High Generalization Ability. Yasutaka Furusho and Kazushi Ikeda
24- Data-Dependent Sample Complexity of Deep Neural Networks via Lipschitz Augmentation. Colin Wei and Tengyu Ma
25- Few-Shot Transfer Learning From Multiple Pre-Trained Networks. Joshua Ka-Wing Lee, Prasanna Sattigeri and Gregory Wornell
26- Uniform Stability and High Order Approximation of SGLD in Non-Convex Learning. Mufan Li and Maxime Gazeau
27- Better Generalization with Adaptive Adversarial Training. Amit Despande, Sandesh Kamath and K V Subrahmanyam
28- Adversarial Training Generalizes Spectral Norm Regularization. Kevin Roth, Yannic Kilcher and Thomas Hofmann
29- A Causal View on Robustness of Neural Networks. Cheng Zhang and Yingzhen Li
30- Improving PAC-Bayes bounds for neural networks using geometric properties of the training method. Anirbit Mukherjee, Dan Roy, Pushpendre Rastogi and Jun Yang
31- An Analysis of the Effect of Invariance on Generalization in Neural Networks. Clare Lyle, Marta Kwiatkowska and Yarin Gal
32- Data-Dependent Mututal Information Bounds for SGLD. Jeffrey Negrea, Daniel Roy, Gintare Karolina Dziugaite, Mahdi Haghifam and Ashish Khisti
33- Comparing normalization in conditional computation tasks. Vincent Michalski, Vikram Voleti, Samira Ebrahimi Kahou, Anthony Ortiz, Pascal Vincent, Chris Pal and Doina Precup
34- Weight and Batch Normalization implement Classical Generalization Bounds. Andrzej Banburski, Qianli Liao, Brando Miranda, Lorenzo Rosasco, Jack Hidary and Tomaso Poggio
35- Increasing batch size through instance repetition improves generalization. Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler and Daniel Soudry
36- Zero-Shot Learning from scratch: leveraging local compositional representations. Tristan Sylvain, Linda Petrini and Devon Hjelm
37- Circuit-Based Intrinsic Methods to Detect Overfitting. Satrajit Chatterjee and Alan Mishchenko
38- Dimension Reduction Approach for Interpretability of Sequence to SequenceRecurrent Neural Networks. Kun Su and Eli Shlizerman
39- Tight PAC-Bayesian generalization error bounds for deep learning. Guillermo Valle Perez, Chico Q. Camargo and Ard A. Louis
40- How Learning Rate and Delay Affect Minima Selection in AsynchronousTraining of Neural Networks: Toward Closing the Generalization Gap. Niv Giladi, Mor Shpigel Nacson, Elad Hoffer and Daniel Soudry
41- Making Convolutional Networks Shift-Invariant Again. Richard Zhang
42- A Meta-Analysis of Overfitting in Machine Learning. Sara Fridovich-Keil, Moritz Hardt, John Miller, Ben Recht, Rebecca Roelofs, Ludwig Schmidt and Vaishaal Shankar
43- Kernelized Capsule Networks. Taylor Killian, Justin Goodwin, Olivia Brown and Sung Son
44- Model similarity mitigates test set overuse. Moritz Hardt, Horia Mania, John Miller, Ben Recht and Ludwig Schmidt
45- Understanding Generalization of Deep Neural Networks Trained with Noisy Labels. Wei Hu, Zhiyuan Li and Dingli Yu
46- Domainwise Classification Network for Unsupervised Domain Adaptation. Seonguk Seo, Yumin Suh, Bohyung Han, Taeho Lee, Tackgeun You, Woong-Gi Chang and Suha Kwak
47- The Generalization-Stability Tradeoff in Neural Network Pruning. Brian Bartoldson, Ari Morcos, Adrian Barbu and Gordon Erlebacher
48- Information matrices and generalization. Valentin Thomas, Fabian Pedregosa, Bart van Merriënboer, Pierre-Antoine Manzagol, Yoshua Bengio and Nicolas Le Roux
49- Adaptively Preconditioned Stochastic Gradient Langevin Dynamics. Chandrasekaran Anirudh Bhardwaj
50- Additive or Concatenating Skip-connections Improve Data Separability. Yasutaka Furusho and Kazushi Ikeda
51- On the Inductive Bias of Neural Tangent Kernels. Alberto Bietti and Julien Mairal
52- PAC Bayes Bound Minimization via Kronecker Normalizing Flows. Chin-Wei Huang, Ahmed Touati, Pascal Vincent, Gintare Karolina Dziugaite, Alexandre Lacoste and Aaron Courville

53- SGD on Neural Networks Learns Functions of Increasing Complexity. Preetum Nakkiran, Gal Kaplun, Dimitris Kalimeris, Tristan Yang, Ben Edelman, Fred Zhang and Boaz Barak
54- Overparameterization without Overfitting: Jacobian-based Generalization Guarantees for Neural Networks. Samet Oymak, Mingchen Li, Zalan Fabian and Mahdi Soltanolkotabi
55- Incorrect gradients and regularization: a perspective of loss landscapes. Mehrdad Yazdani
56- Divide and Conquer: Leveraging Intermediate Feature Representations for Quantized Training of Neural Networks. Ahmed Youssef, Prannoy Pilligundla and Hadi Esmaeilzadeh
57- SinReQ: Generalized Sinusoidal Regularization for Low-Bitwidth Deep Quantized Training. Ahmed Youssef, Prannoy Pilligundla and Hadi Esmaeilzadeh
58- Natural Adversarial Examples. Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt and Dawn Song
59- On the Properties of the Objective Landscapes and Generalization of Gradient-Based Meta-Learning. Simon Guiroy, Vikas Verma and Christopher Pal
60- Angular Visual Hardness. Beidi Chen, Weiyang Liu, Animesh Garg, Zhiding Yu, Anshumali Shrivastava and Animashree Anandkumar
61- Luck Matters: Understanding Training Dynamics of Deep ReLU Networks. Yuandong Tian, Tina Jiang, Qucheng Gong and Ari Morcos
62- Understanding of Generalization in Deep Learning via Tensor Methods. Jingling Li, Yanchao Sun, Ziyin Liu, Taiji Suzuki and Furong Huang
63- Learning from Rules Performs as Implicit Regularization. Hossein Hosseini, Ramin Moslemi, Ali Hooshmand and Ratnesh Sharma
64- Stochastic Mirror Descent on Overparameterized Nonlinear Models: Convergence, Implicit Regularization, and Generalization. Navid Azizan, Sahin Lale and Babak Hassibi
65- Scaling Characteristics of Sequential Multitask Learning: Networks Naturally Learn to Learn. Guy Davidson and Michael Mozer
66- Size-free generalization bounds for convolutional neural networks. Phillip Long and Hanie Sedghi

Abstract 18: **Overparameterization without Overfitting: Jacobian-based Generalization Guarantees for Neural Networks in Understanding and Improving Generalization in Deep Learning**, 04:30 PM

Authors: Samet Oymak, Mingchen Li, Zalan Fabian and Mahdi Soltanolkotabi

Abstract: Many modern neural network architectures contain many more parameters than the size of the training data. Such networks can easily overfit to training data, hence it is crucial to understand the fundamental principles that facilitate good test accuracy. This paper explores the generalization capabilities of neural networks trained via gradient descent. We show that the Jacobian matrix associated with the network dictates the directions where learning is generalizable and fast versus directions where overfitting occurs and learning is slow. We develop a bias-variance theory which provides a control knob to split the Jacobian spectum into ``information" and ``nuisance" spaces associated with the large and small singular values of the Jacobian. We show that (i) over the information space learning is fast and we can quickly train a model with zero training loss that can also generalize well, (ii) over the nuisance subspace overfitting might result in higher variance hence early stopping can help with generalization at the expense of some bias. We conduct numerical experiments on deep networks that corroborate out theory and demonstrate that: (i) the Jacobian of typical networks exhibit a bimodal structure with a few large singular values and many small ones leading to a low-dimensional information space (ii) most of the useful information lies on the information space where learning happens quickly.

Abstract 19: **How Learning Rate and Delay Affect Minima Selection in AsynchronousTraining of Neural Networks: Toward Closing the Generalization Gap in Understanding and Improving Generalization in Deep Learning**, 04:45 PM

Authors: Niv Giladi, Mor Shpigel Nacson, Elad Hoffer and Daniel Soudry

Abstract: Background: Recent developments have made it possible to accelerate neural networks training significantly using large batch sizes and

data parallelism. Training in an asynchronous fashion, where delay occurs, can make training even more scalable. However, asynchronous training has its pitfalls, mainly a degradation in generalization, even after convergence of the algorithm. This gap remains not well understood, as theoretical analysis so far mainly focused on the convergence rate of asynchronous methods. Contributions: We examine asynchronous training from the perspective of dynamical stability. We find that the degree of delay interacts with the learning rate, to change the set of minima accessible by an asynchronous stochastic gradient descent algorithm. We derive closed-form rules on how the hyperparameters could be changed while keeping the accessible set the same. Specifically, for high delay values, we find that the learning rate should be decreased inversely with the delay, and discuss the effect of momentum. We provide empirical experiments to validate our theoretical findings

Abstract 20: **Poster Session in Understanding and Improving Generalization in Deep Learning**, 05:00 PM

1- Uniform convergence may be unable to explain generalization in deep learning. Vaishnavh Nagarajan and J. Zico Kolter
2- The effects of optimization on generalization in infinitely wide neural networks. Anastasia Borovykh
3- Generalized Capsule Networks with Trainable Routing Procedure. Zhenhua Chen, Chuhua Wang, David Crandall and Tiancong Zhao
4- Implicit Regularization of Discrete Gradient Dynamics in Deep Linear Neural Networks. Gauthier Gidel, Francis Bach and Simon Lacoste-Julien
5- Stable Rank Normalization for Improved Generalization in Neural Networks. Amartya Sanyal, Philip H Torr and Puneet K Dokania
6- On improving deep learning generalization with adaptive sparse connectivity. Shiwei Liu, Decebal Constantin Mocanu and Mykola Pechenizkiy
7- Identity Connections in Residual Nets Improve Noise Stability. Shuzhi Yu and Carlo Tomasi
8- Factors for the Generalisation of Identity Relations by Neural Networks. Radha Manisha Kopparti and Tillman Weyde
9- Output-Constrained Bayesian Neural Networks. Wanqian Yang, Lars Lorch, Moritz A. Graule, Srivatsan Srinivasan, Anirudh Suresh, Jiayu Yao, Melanie F. Pradier and Finale Doshi-Velez
10- An Empirical Study on Hyperparameters and their Interdependence for RL Generalization. Xingyou Song, Yilun Du and Jacob Jackson
11- Towards Large Scale Structure of the Loss Landscape of Neural Networks. Stanislav Fort and Stanislaw Jastrzebski
12- Detecting Extrapolation with Influence Functions. David Madras, James Atwood and Alex D'Amour
13- Towards Task and Architecture-Independent Generalization Gap Predictors. Scott Yak, Hanna Mazzawi and Javier Gonzalvo
14- SGD Picks a Stable Enough Trajectory. Stanisław Jastrzębski and Stanislav Fort
15- MazeNavigator: A Customisable 3D Benchmark for Assessing Generalisation in Reinforcement Learning. Luke Harries, Sebastian Lee, Jaroslaw Rzepecki, Katja Hofmann and Sam Devlin
16- Utilizing Eye Gaze to Enhance the Generalization of Imitation Network to Unseen Environments. Congcong Liu, Yuying Chen, Lei Tai, Ming Liu and Bertram Shi
17- Investigating Under and Overfitting in Wasserstein Generative Adversarial Networks. Ben Adlam, Charles Weill and Amol Kapoor
18- An Empirical Evaluation of Adversarial Robustness under Transfer Learning. Todor Davchev, Timos Korres, Stathi Fotiadis, Nick Antonopoulos and Subramanian Ramamoorthy
19- On Adversarial Robustness of Small vs Large Batch Training. Sandesh Kamath, Amit Deshpande and K V Subrahmanyam
20- The Principle of Unchanged Optimality in Reinforcement Learning Generalization. Xingyou Song and Alex Irpan
21- On the Generalization Capability of Memory Networks for Reasoning. Monireh Ebrahimi, Md Kamruzzaman Sarker, Federico Bianchi, Ning Xie, Aaron Eberhart, Derek Doran and Pascal Hitzler
22- Visualizing How Embeddings Generalize. Xiaotong Liu, Hong Xuan, Zeyu Zhang, Abby Stylianou and Robert Pless
23- Theoretical Analysis of the Fixup Initialization for Fast Convergence and High Generalization Ability. Yasutaka Furusho and Kazushi Ikeda
24- Data-Dependent Sample Complexity of Deep Neural Networks via Lipschitz Augmentation. Colin Wei and Tengyu Ma
25- Few-Shot Transfer Learning From Multiple Pre-

Trained Networks. Joshua Ka-Wing Lee, Prasanna Sattigeri and Gregory Wornell
26- Uniform Stability and High Order Approximation of SGLD in Non-Convex Learning. Mufan Li and Maxime Gazeau
27- Better Generalization with Adaptive Adversarial Training. Amit Despande, Sandesh Kamath and K V Subrahmanyam
28- Adversarial Training Generalizes Spectral Norm Regularization. Kevin Roth, Yannic Kilcher and Thomas Hofmann
29- A Causal View on Robustness of Neural Networks. Cheng Zhang and Yingzhen Li
30- Improving PAC-Bayes bounds for neural networks using geometric properties of the training method. Anirbit Mukherjee, Dan Roy, Pushpendre Rastogi and Jun Yang
31- An Analysis of the Effect of Invariance on Generalization in Neural Networks. Clare Lyle, Marta Kwiatkowska and Yarin Gal
32- Data-Dependent Mututal Information Bounds for SGLD. Jeffrey Negrea, Daniel Roy, Gintare Karolina Dziugaite, Mahdi Haghifam and Ashish Khisti
33- Comparing normalization in conditional computation tasks. Vincent Michalski, Vikram Voleti, Samira Ebrahimi Kahou, Anthony Ortiz, Pascal Vincent, Chris Pal and Doina Precup
34- Weight and Batch Normalization implement Classical Generalization Bounds. Andrzej Banburski, Qianli Liao, Brando Miranda, Lorenzo Rosasco, Jack Hidary and Tomaso Poggio
35- Increasing batch size through instance repetition improves generalization. Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler and Daniel Soudry
36- Zero-Shot Learning from scratch: leveraging local compositional representations. Tristan Sylvain, Linda Petrini and Devon Hjelm
37- Circuit-Based Intrinsic Methods to Detect Overfitting. Satrajit Chatterjee and Alan Mishchenko
38- Dimension Reduction Approach for Interpretability of Sequence to SequenceRecurrent Neural Networks. Kun Su and Eli Shlizerman
39- Tight PAC-Bayesian generalization error bounds for deep learning. Guillermo Valle Perez, Chico Q. Camargo and Ard A. Louis
40- How Learning Rate and Delay Affect Minima Selection in AsynchronousTraining of Neural Networks: Toward Closing the Generalization Gap. Niv Giladi, Mor Shpigel Nacson, Elad Hoffer and Daniel Soudry
41- Making Convolutional Networks Shift-Invariant Again. Richard Zhang
42- A Meta-Analysis of Overfitting in Machine Learning. Sara Fridovich-Keil, Moritz Hardt, John Miller, Ben Recht, Rebecca Roelofs, Ludwig Schmidt and Vaishaal Shankar
43- Kernelized Capsule Networks. Taylor Killian, Justin Goodwin, Olivia Brown and Sung Son
44- Model similarity mitigates test set overuse. Moritz Hardt, Horia Mania, John Miller, Ben Recht and Ludwig Schmidt
45- Understanding Generalization of Deep Neural Networks Trained with Noisy Labels. Wei Hu, Zhiyuan Li and Dingli Yu
46- Domainwise Classification Network for Unsupervised Domain Adaptation. Seonguk Seo, Yumin Suh, Bohyung Han, Taeho Lee, Tackgeun You, Woong-Gi Chang and Suha Kwak
47- The Generalization-Stability Tradeoff in Neural Network Pruning. Brian Bartoldson, Ari Morcos, Adrian Barbu and Gordon Erlebacher
48- Information matrices and generalization. Valentin Thomas, Fabian Pedregosa, Bart van Merriënboer, Pierre-Antoine Manzagol, Yoshua Bengio and Nicolas Le Roux
49- Adaptively Preconditioned Stochastic Gradient Langevin Dynamics. Chandrasekaran Anirudh Bhardwaj
50- Additive or Concatenating Skip-connections Improve Data Separability. Yasutaka Furusho and Kazushi Ikeda
51- On the Inductive Bias of Neural Tangent Kernels. Alberto Bietti and Julien Mairal
52- PAC Bayes Bound Minimization via Kronecker Normalizing Flows. Chin-Wei Huang, Ahmed Touati, Pascal Vincent, Gintare Karolina Dziugaite, Alexandre Lacoste and Aaron Courville
53- SGD on Neural Networks Learns Functions of Increasing Complexity. Preetum Nakkiran, Gal Kaplun, Dimitris Kalimeris, Tristan Yang, Ben Edelman, Fred Zhang and Boaz Barak
54- Overparameterization without Overfitting: Jacobian-based Generalization Guarantees for Neural Networks. Samet Oymak, Mingchen Li, Zalan Fabian and Mahdi Soltanolkotabi
55- Incorrect gradients and regularization: a perspective of loss landscapes. Mehrdad Yazdani
56- Divide and Conquer: Leveraging Intermediate Feature Representations for Quantized Training of Neural Networks. Ahmed Youssef, Prannoy Pilligundla and Hadi Esmaeilzadeh
57- SinReQ: Generalized Sinusoidal Regularization

for Low-Bitwidth Deep Quantized Training. Ahmed Youssef, Prannoy Pilligundla and Hadi Esmaeilzadeh
58- Natural Adversarial Examples. Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt and Dawn Song
59- On the Properties of the Objective Landscapes and Generalization of Gradient-Based Meta-Learning. Simon Guiroy, Vikas Verma and Christopher Pal
60- Angular Visual Hardness. Beidi Chen, Weiyang Liu, Animesh Garg, Zhiding Yu, Anshumali Shrivastava and Animashree Anandkumar
61- Luck Matters: Understanding Training Dynamics of Deep ReLU Networks. Yuandong Tian, Tina Jiang, Qucheng Gong and Ari Morcos
62- Understanding of Generalization in Deep Learning via Tensor Methods. Jingling Li, Yanchao Sun, Ziyin Liu, Taiji Suzuki and Furong Huang
63- Learning from Rules Performs as Implicit Regularization. Hossein Hosseini, Ramin Moslemi, Ali Hooshmand and Ratnesh Sharma
64- Stochastic Mirror Descent on Overparameterized Nonlinear Models: Convergence, Implicit Regularization, and Generalization. Navid Azizan, Sahin Lale and Babak Hassibi
65- Scaling Characteristics of Sequential Multitask Learning: Networks Naturally Learn to Learn. Guy Davidson and Michael Mozer
66- Size-free generalization bounds for convolutional neural networks. Phillip Long and Hanie Sedghi

## 6th ICML Workshop on Automated Machine Learning (AutoML 2019)

***Frank Hutter, Joaquin Vanschoren, Katharina Eggensperger, Matthias Feurer***

**Grand Ballroom B, Fri Jun 14, 08:30 AM**

Machine learning has achieved considerable successes in recent years, but this success often relies on human experts, who construct appropriate features, design learning architectures, set their hyperparameters, and develop new learning algorithms. Driven by the demand for off-the-shelf machine learning methods from an ever-growing community, the research area of AutoML targets the progressive automation of machine learning aiming to make effective methods available to everyone. The workshop targets a broad audience ranging from core machine learning researchers in different fields of ML connected to AutoML, such as neural architecture search, hyperparameter optimization, meta-learning, and learning to learn, to domain experts aiming to apply machine learning to new types of problems.

### Schedule

09:00 AM **Welcome** *Hutter*

09:05 AM **Keynote by Peter Frazier: Grey-box Bayesian Optimization for AutoML** *Frazier*

09:40 AM **Poster Session 1 (all papers)** *Gargiani, Zur, Baskin, Zheltonozhskii, Li, Talwalkar, Shang, Behl, Baydin, Couckuyt, Dhaene, Lin, Wei, Sun, Majumder, Donini, Ozaki, Adams, Geißler, Luo, peng, , Zhang, Langford, Caruana, Dey, Weill, Gonzalvo, Yang, Yak, Hotaj, Macko, Mohri, Cortes, Webb, Chen, Jankowiak, Goodman, Klein, Hutter, Javaheripi, Samragh, Lim, Kim, KIM, Volpp, Drori, Krishnamurthy, Cho, Jastrzebski, de Laroussilhe, Tan, Ma, Houlsby, Gesmundo, Borsos, Maziarz, Petroski Such, Lehman, Stanley, Clune, Gijsbers, Vanschoren, Mohr, Hüllermeier, Xiong, Zhang, zhu, Shao, Faust, Valko, Li, Escalante, Wever, Khorlin, Javidi, Francis, Mukherjee, Kim, McCourt, Kim, You, Choi, Knudde, Tornede, Jerfel*

11:00 AM **Keynote by Rachel Thomas: Lessons Learned from Helping 200,000 non-ML experts use ML** *Thomas*

11:35 AM **Contributed Talk 1: A Boosting Tree Based AutoML System for Lifelong Machine Learning** *Xiong*

12:00 PM **Poster Session 2 (all papers)**

12:50 PM **Lunch Break**

02:00 PM **Keynote by Jeff Dean: An Overview of Google's Work on AutoML and Future Directions** *Dean*

| | |
|---|---|
| 02:35 PM | **Contributed Talk 2: Transfer NAS: Knowledge Transfer between Search Spaces with Transformer Agents** *Borsos* |
| 03:00 PM | **Poster Session 3 (all papers)** |
| 04:00 PM | **Contributed Talk 3: Random Search and Reproducibility for Neural Architecture Search** *Li* |
| 04:25 PM | **Keynote by Charles Sutton: Towards Semi-Automated Machine Learning** *Sutton* |
| 05:00 PM | **Panel Discussion** *Zhang, Sutton, Li, Thomas, LeDell* |
| 06:00 PM | **Closing Remarks** *Hutter* |

Abstracts (4):

Abstract 2: **Keynote by Peter Frazier: Grey-box Bayesian Optimization for AutoML in 6th ICML Workshop on Automated Machine Learning (AutoML 2019)**, *Frazier* 09:05 AM

Bayesian optimization is a powerful and flexible tool for AutoML. While BayesOpt was first deployed for AutoML simply as a black-box optimizer, recent approaches perform grey-box optimization: they leverage capabilities and problem structure specific to AutoML such as freezing and thawing training, early stopping, treating cross-validation error minimization as multi-task learning, and warm starting from previously tuned models. We provide an overview of this area and describe recent advances for optimizing sampling-based acquisition functions that make grey-box BayesOpt significantly more efficient.

Abstract 4: **Keynote by Rachel Thomas: Lessons Learned from Helping 200,000 non-ML experts use ML in 6th ICML Workshop on Automated Machine Learning (AutoML 2019)**, *Thomas* 11:00 AM

The mission of AutoML is to make ML available for non-ML experts and to accelerate research on ML. We have a very similar mission at fast.ai and have helped over 200,000 non-ML experts use state-of-the-art ML (via our research, software, & teaching), yet we do not use methods from the AutoML literature. I will share several insights we've learned through this work, with the hope that they may be helpful to AutoML researchers.

Abstract 8: **Keynote by Jeff Dean: An Overview of Google's Work on AutoML and Future Directions in 6th ICML Workshop on Automated Machine Learning (AutoML 2019)**, *Dean* 02:00 PM

In this talk I'll survey work by Google researchers over the past several years on the topic of AutoML, or learning-to-learn. The talk will touch on basic approaches, some successful applications of AutoML to a variety of domains, and sketch out some directions for future AutoML systems that can leverage massively multi-task learning systems for automatically solving new problems.

Abstract 12: **Keynote by Charles Sutton: Towards Semi-Automated Machine Learning in 6th ICML Workshop on Automated Machine Learning (AutoML 2019)**, *Sutton* 04:25 PM

The practical work of deploying a machine learning system is dominated by issues outside of training a model: data preparation, data cleaning, understanding the data set, debugging models, and so on. What does it mean to apply ML to this “grunt work” of machine learning and data science? I will describe first steps towards tools in these directions, based on the idea of semi-automating ML: using unsupervised learning to find patterns in the data that can be used to guide data analysts. I will also describe a new notebook system for pulling these tools together: if we augment Jupyter-style notebooks with data-flow and provenance information, this enables a new class of data-aware notebooks which are much more natural for data manipulation.

## Generative Modeling and Model-Based Reasoning for Robotics and AI

***Aravind Rajeswaran, Emanuel Todorov, Igor Mordatch, William Agnew, Amy Zhang, Joelle Pineau,***

***Michael Chang, Dumitru Erhan, Sergey Levine, Kimberly Stachenfeld, Marvin Zhang***

**Hall A, Fri Jun 14, 08:30 AM**

Workshop website: https://sites.google.com/view/mbrl-icml2019

In the recent explosion of interest in deep RL, “model-free” approaches based on Q-learning and actor-critic architectures have received the most attention due to their flexibility and ease of use. However, this generality often comes at the expense of efficiency (statistical as well as computational) and robustness. The large number of required samples and safety concerns often limit direct use of model-free RL for real-world settings.

Model-based methods are expected to be more efficient. Given accurate models, trajectory optimization and Monte-Carlo planning methods can efficiently compute near-optimal actions in varied contexts. Advances in generative modeling, unsupervised, and self-supervised learning provide methods for learning models and representations that support subsequent planning and reasoning. Against this backdrop, our workshop aims to bring together researchers in generative modeling and model-based control to discuss research questions at their intersection, and to advance the state of the art in model-based RL for robotics and AI. In particular, this workshops aims to make progress on questions related to:

1. How can we learn generative models efficiently? Role of data, structures, priors, and uncertainty.
2. How to use generative models efficiently for planning and reasoning? Role of derivatives, sampling, hierarchies, uncertainty, counterfactual reasoning etc.
3. How to harmoniously integrate model-learning and model-based decision making?
4. How can we learn compositional structure and environmental constraints? Can this be leveraged for better generalization and reasoning?

## Schedule

| | |
|---|---|
| 08:45 AM | **Welcome and Introduction** *Rajeswaran* |
| 09:00 AM | **Yann LeCun** |
| 09:30 AM | **Jessica Hamrick** |
| 10:00 AM | **Spotlight Session 1** |
| 11:00 AM | **Stefan Schaal** |
| 02:00 PM | **David Silver** |
| 03:30 PM | **Byron Boots** |
| 04:20 PM | **Chelsea Finn** |
| 04:50 PM | **Abhinav Gupta** |
| 05:20 PM | **Discussion Panel** |

## Uncertainty and Robustness in Deep Learning

***Sharon Yixuan Li, Balaji Lakshminarayanan, Dan Hendrycks, Tom Dietterich, Justin Gilmer***

**Hall B, Fri Jun 14, 08:30 AM**

There has been growing interest in rectifying deep neural network vulnerabilities. Challenges arise when models receive samples drawn from outside the training distribution. For example, a neural network tasked with classifying handwritten digits may assign high confidence predictions to cat images. Anomalies are frequently encountered when deploying ML models in the real world. Well-calibrated predictive uncertainty estimates are indispensable for many machine learning applications, such as self-driving cars and medical diagnosis systems. Generalization to unseen and worst-case inputs is also essential for robustness to distributional shift. In order to have ML models reliably predict in open environment, we must deepen technical understanding in the following areas: (1) learning algorithms that are robust to changes in input data distribution (e.g.,

detect out-of-distribution examples); (2) mechanisms to estimate and calibrate confidence produced by neural networks and (3) methods to improve robustness to adversarial and common corruptions, and (4) key applications for uncertainty such as in artificial intelligence (e.g., computer vision, robotics, self-driving cars, medical imaging) as well as broader machine learning tasks.

This workshop will bring together researchers and practitioners from the machine learning communities, and highlight recent work that contribute to address these challenges. Our agenda will feature contributed papers with invited speakers. Through the workshop we hope to help identify fundamentally important directions on robust and reliable deep learning, and foster future collaborations.

## Schedule

08:30 AM **Welcome** *Li*

08:40 AM **Spotlight** *Scott, Koshy, Aigrain, Bidart, Panda, Yap, Yacoby, Gontijo Lopes, Marchisio, Englesson, Yang, Graule, Sun, Kang, Dusenberry, Du, Maennel, Menda, Edupuganti, Metz, Stutz, Srinivasan, Sämann, N Balasubramanian, Mohseni, Cornish, Butepage, Wang, Li, Han, Li, Andriushchenko, Ruff, Vadera, Ovadia, Thulasidasan, Ji, Niu, Mahloujifar, Kumar, CHUN, Yin, Xu, Gomes, Rohekar*

09:30 AM **Keynote by Max Welling: A Nonparametric Bayesian Approach to Deep Learning (without GPs)** *Welling*

10:00 AM **Poster Session 1 (all papers)**

11:00 AM **Keynote by Kilian Weinberger: On Calibration and Fairness** *Weinberger*

11:30 AM **Why ReLU networks yield high-confidence predictions far away from the training data and how to mitigate the problem** *Andriushchenko*

11:40 AM **Detecting Extrapolation with Influence Functions** *Madras*

11:50 AM **How Can We Be So Dense? The Robustness of Highly Sparse Representations** *Ahmad*

12:00 PM **Keynote by Suchi Saria: Safety Challenges with Black-Box Predictors and Novel Learning Approaches for Failure Proofing** *Saria*

02:00 PM **Subspace Inference for Bayesian Deep Learning** *Kirichenko, Izmailov, Wilson*

02:10 PM **Quality of Uncertainty Quantification for Bayesian Neural Network Inference** *Yao*

02:20 PM **'In-Between' Uncertainty in Bayesian Neural Networks** *Foong*

02:30 PM **Keynote by Dawn Song: Adversarial Machine Learning: Challenges, Lessons, and Future Directions** *Song*

03:30 PM **Keynote by Terrance Boult: The Deep Unknown: on Open-set and Adversarial Examples in Deep Learning** *Boult*

04:00 PM **Panel Discussion (moderator: Tom Dietterich)** *Welling, Weinberger, Boult, Song, Dietterich*

05:00 PM **Poster Session 2 (all papers)**

Abstracts (9):

Abstract 3: **Keynote by Max Welling: A Nonparametric Bayesian Approach to Deep Learning (without GPs) in Uncertainty and Robustness in Deep Learning**, *Welling* 09:30 AM

We present a new family of exchangeable stochastic processes suitable for deep learning. Our nonparametric Bayesian method models distributions over functions by learning a graph of dependencies on top of latent representations of the points in the given dataset. In doing so, they define a Bayesian model without explicitly positing a prior distribution over latent global parameters; they instead adopt priors over the relational structure of the given dataset, a task that is much simpler. We show how we can learn such models from data, demonstrate that they are scalable to large datasets through mini-batch optimization and describe how we can make predictions for new points via their posterior predictive distribution. We experimentally evaluate FNPs on the tasks of toy regression and image classification and show that, when compared to baselines that employ global latent

parameters, they offer both competitive predictions as well as more robust uncertainty estimates.

Abstract 5: **Keynote by Kilian Weinberger: On Calibration and Fairness in Uncertainty and Robustness in Deep Learning**, *Weinberger* 11:00 AM

We investigate calibration for deep learning algorithms in classification and regression settings. Although we show that typically deep networks tend to be highly mis-calibrated, we demonstrate that this is easy to fix - either to obtain more trustworthy confidence estimates or to detect outliers in the data. Finally, we relate calibration with the recently raised tension between minimizing error disparity across different population groups while maintaining calibrated probability estimates. We show that calibration is compatible only with a single error constraint (i.e. equal false-negatives rates across groups), and show that any algorithm that satisfies this relaxation is no better than randomizing a percentage of predictions for an existing classifier. These unsettling findings, which extend and generalize existing results, are empirically confirmed on several datasets.

Abstract 6: **Why ReLU networks yield high-confidence predictions far away from the training data and how to mitigate the problem in Uncertainty and Robustness in Deep Learning**, *Andriushchenko* 11:30 AM

Classifiers used in the wild, in particular for safety-critical systems, should know when they don’t know, in particular make low confidence predictions far away from the train- ing data. We show that ReLU type neural networks fail in this regard as they produce almost always high confidence predictions far away from the training data. For bounded domains we propose a new robust optimization technique similar to adversarial training which enforces low confidence pre- dictions far away from the training data. We show that this technique is surprisingly effective in reducing the confidence of predictions far away from the training data while maintaining high confidence predictions and test error on the original classification task com- pared to standard training. This is a short version of the corresponding CVPR paper.

Abstract 7: **Detecting Extrapolation with Influence Functions in Uncertainty and Robustness in Deep Learning**, *Madras* 11:40 AM

In this work, we explore principled methods for extrapolation detection. We define extrapolation as occurring when a model’s conclusion at a test point is underdetermined by the training data. Our metrics for detecting extrapolation are based on influence functions, inspired by the intuition that a point requires extrapolation if its inclusion in the training set would significantly change the model’s learned parameters. We provide interpretations of our methods in terms of the eigendecomposition of the Hessian. We present experimental evidence that our method is capable of identifying extrapolation to out-of-distribution points.

Abstract 8: **How Can We Be So Dense? The Robustness of Highly Sparse Representations in Uncertainty and Robustness in Deep Learning**, *Ahmad* 11:50 AM

Neural networks can be highly sensitive to noise and perturbations. In this paper we suggest that high dimensional sparse representations can lead to increased robustness to noise and interference. A key intuition we develop is that the ratio of the match volume around a sparse vector divided by the total representational space decreases exponentially with dimensionality, leading to highly robust matching with low interference from other patterns. We analyze efficient sparse networks containing both sparse weights and sparse activations. Simulations on MNIST, the Google Speech Command Dataset, and CIFAR-10 show that such networks demonstrate improved robustness to random noise compared to dense networks, while maintaining competitive accuracy. We propose that sparsity should be a core design constraint for creating highly robust networks.

Abstract 10: **Subspace Inference for Bayesian Deep Learning in Uncertainty and Robustness in Deep Learning**, *Kirichenko, Izmailov, Wilson* 02:00 PM

Bayesian inference was once a gold standard for learning with neural networks, providing accurate full predictive distributions and well calibrated uncertainty. However, scaling Bayesian inference techniques to deep neural networks is challenging due to the high dimensionality of the parameter space. In this pa- per, we construct low-dimensional subspaces of parameter space that contain diverse sets of models, such as the first principal components of the stochastic gradient descent (SGD) trajectory. In these subspaces, we are able to apply elliptical slice sampling and variational inference, which struggle in the full parameter space. We show that Bayesian model averaging over the induced posterior in these subspaces produces high accurate predictions and well-calibrated predictive uncertainty for both regression and image classification.

Abstract 11: **Quality of Uncertainty Quantification for Bayesian Neural Network Inference in Uncertainty and Robustness in Deep Learning**, *Yao* 02:10 PM

Bayesian Neural Networks (BNNs) place priors over the parameters in a neural network. Inference in BNNs, however, is difficult; all inference methods for BNNs are approximate. In this work, we empirically compare the quality of predictive uncertainty estimates for 10 common inference methods on both regression and classification tasks. Our experiments demonstrate that commonly used metrics (e.g. test log-likelihood) can be misleading. Our experiments also indicate that inference innovations designed to capture structure in the posterior do not necessarily produce high quality posterior approximations.

Abstract 12: **'In-Between' Uncertainty in Bayesian Neural Networks in Uncertainty and Robustness in Deep Learning**, *Foong* 02:20 PM

We describe a limitation in the expressiveness of the predictive uncertainty estimate given by mean- field variational inference (MFVI), a popular approximate inference method for Bayesian neural networks. In particular, MFVI fails to give calibrated uncertainty estimates in between separated regions of observations. This can lead to catastrophically overconfident predictions when testing on out-of-distribution data. Avoiding such over-confidence is critical for active learning, Bayesian optimisation and out-of-distribution robustness. We instead find that a classical technique, the linearised Laplace approximation, can handle ‘in- between’ uncertainty much better for small network architectures.

Abstract 14: **Keynote by Terrance Boult: The Deep Unknown: on Open-set and Adversarial Examples in Deep Learning in Uncertainty and Robustness in Deep Learning**, *Boult* 03:30 PM

The first part of the talk will explore issues with deep networks dealing with "unknowns" inputs, and the general problems of open-set recognition in deep networks. We review the core of open-set recognition theory and its application in our first attempt at open-set deep networks, "OpenMax" We discuss is successes and limitations and why classic "open-set" approaches don't really solve the problem of deep unknowns. We then present our recent work from NIPS2018, on a new model we call the ObjectoSphere. Using ObjectoSphere loss begins to address the learning of deep features that can handle unknown inputs. We present examples of its use first on simple datasets sets (MNIST/CFAR) and then onto unpublished work applying it to the real-world problem of open-set face recognition. We discuss of the relationship between open set recognition theory and adversarial image generation, showing how our deep-feature adversarial approach, called LOTS can attack the first OpenMax solution, as well as successfully attack even open-set face recognition systems. We end with a discussion of how open set theory can be applied to improve network robustness.

## Reinforcement Learning for Real Life

***Yuxi Li, Alborz Geramifard, Lihong Li, Csaba Szepesvari, Tao Wang***

**Seaside Ballroom, Fri Jun 14, 08:30 AM**

Reinforcement learning (RL) is a general learning, predicting, and decision making paradigm. RL provides solution methods for sequential decision making problems as well as those can be transformed into sequential ones. RL connects deeply with optimization, statistics, game theory, causal inference, sequential experimentation, etc., overlaps largely with approximate dynamic programming and optimal control, and applies broadly in science, engineering and arts.

RL has been making steady progress in academia recently, e.g., Atari games, AlphaGo, visuomotor policies for robots. RL has also been applied to real world scenarios like recommender systems and neural architecture search. See a recent collection about RL applications at https://medium.com/@yuxili/rl-applications-73ef685c07eb. It is desirable to have RL systems that work in the real world with real benefits. However, there are many issues for RL though, e.g. generalization, sample efficiency, and exploration vs. exploitation dilemma. Consequently, RL is far from being widely deployed. Common, critical and pressing questions for the RL community are then: Will RL have wide deployments? What are the issues? How to solve them?

The goal of this workshop is to bring together researchers and practitioners from industry and academia interested in addressing practical and/or theoretical issues in applying RL to real life scenarios, review state of the arts, clarify impactful research problems, brainstorm open challenges, share first-hand lessons and experiences from real life deployments, summarize what has worked and what has not, collect tips for people from industry looking to apply RL and RL experts interested in applying their methods to real domains, identify potential opportunities, generate new ideas for future lines of research and development, and promote awareness and collaboration. This is not "yet another RL workshop": it is about how to successfully apply RL to real life applications. This is a less addressed issue in the RL/ML/AI community, and calls for immediate attention for sustainable prosperity of RL research and development.

## Schedule

| | |
|---|---|
| 08:30 AM | **optional early-bird posters** |
| 08:50 AM | **opening remarks by organizers** |
| 09:00 AM | **invited talk by David Silver (Deepmind): AlphaStar: Mastering the Game of StarCraft II** *Silver* |
| 09:20 AM | **invited talk by John Langford (Microsoft Research): How do we make Real World Reinforcement Learning revolution?** *Langford* |
| 09:40 AM | **invited talk by Craig Boutilier (Google Research): Reinforcement Learning in Recommender Systems: Some Challenges** *Boutilier* |
| 10:00 AM | **posters** *Chen, Garau Luis, Albert Smet, Modi, Tomkins, Simmons-Edler, Mao, Irpan, Lu, Wang, Mukherjee, Raghu, Shihab, Ahn, Fakoor, Chaudhari, Smirnova, Oh, Tang, Qin, Li, Brittain, Fox, Paul, Gao, Chow, Dulac-Arnold, Nachum, Karampatziakis, Balaji, Paul, Davody, Bouneffouf, Sahni, Kim, Kolobov, Amini, Liu, Chen, , Boutilier* |
| 10:30 AM | **coffee break** |
| 11:00 AM | **panel discussion with Craig Boutilier (Google Research), Emma Brunskill (Stanford), Chelsea Finn (Google Brain, Stanford, UC Berkeley), Mohammad Ghavamzadeh (Facebook AI), John Langford (Microsoft Research) and David Silver (Deepmind)** *Stone, Boutilier, Brunskill, Finn, Langford, Silver, Ghavamzadeh* |
| 12:00 PM | **optional posters** |

Abstracts (3):

Abstract 3: **invited talk by David Silver (Deepmind): AlphaStar: Mastering the**

**Game of StarCraft II in Reinforcement Learning for Real Life**, *Silver* 09:00 AM

In recent years, the real-time strategy game of StarCraft has emerged by consensus as an important challenge for AI research. It combines several major difficulties that are intractable for many existing algorithms: a large, structured action space; imperfect information about the opponent; a partially observed map; and cycles in the strategy space. Each of these challenges represents a major difficulty faced by real-world applications, for example those based on internet-scale action spaces, game theory in e.g. security, point-and-click interfaces, or robust AI in the presence of diverse and potentially exploitative user strategies. Here, we introduce AlphaStar: a novel combination of deep learning and reinforcement learning that mastered this challenging domain and defeated human professional players for the first time.

Abstract 4: **invited talk by John Langford (Microsoft Research): How do we make Real World Reinforcement Learning revolution? in Reinforcement Learning for Real Life**, *Langford* 09:20 AM

Abstract: Doing Real World Reinforcement Learning implies living with steep constraints on the sample complexity of solutions. Where is this viable? Where might it be viable in the near future? In the far future? How can we design a research program around identifying and building such solutions? In short, what are the missing elements we need to really make reinforcement learning more mundane and commonly applied than Supervised Learning? The potential is certainly there given the naturalness of RL compared to supervised learning, but the present is manifestly different. https://en.wikipedia.org/wiki/John_Langford_(computer_scientist)

Abstract 5: **invited talk by Craig Boutilier (Google Research): Reinforcement Learning in Recommender Systems: Some Challenges in Reinforcement Learning for Real Life**, *Boutilier* 09:40 AM

Abstract: I'll present a brief overview of some recent work on reinforcement learning motivated by practical issues that arise in the application of RL to online, user-facing applications like recommender systems. These include stochastic action sets, long-term cumulative effects, and combinatorial action spaces. I'll provide some detail on the last of these, describing SlateQ, a novel decomposition technique that allows value-based RL (e.g., Q-learning) in slate-based recommender to scale to commercial production systems, and briefly describe both small-scale simulation and a large-scale experiment with YouTube.

Bio: Craig is Principal Scientist at Google, working on various aspects of decision making under uncertainty (e.g., reinforcement learning, Markov decision processes, user modeling, preference modeling and elicitation) and recommender systems. He received his Ph.D. from the University of Toronto in 1992, and has held positions at the University of British Columbia, University of Toronto, CombineNet, and co-founded Granata Decision Systems.

Craig was Editor-in-Chief of JAIR; Associate Editor with ACM TEAC, JAIR, JMLR, and JAAMAS; Program Chair for IJCAI-09 and UAI-2000. Boutilier is a Fellow of the Royal Society of Canada (RSC), the Association for Computing Machinery (ACM) and the Association for the Advancement of Artificial Intelligence (AAAI). He was recipient of the 2018 ACM/SIGAI Autonomous Agents Research Award and a Tier I Canada Research Chair; and has received (with great co-authors) a number of Best Paper awards including: the 2009 IJCAI-JAIR Best Paper Prize; the 2014 AIJ Prominent Paper Award; and the 2018 NeurIPS Best Paper Award.

## Real-world Sequential Decision Making: Reinforcement Learning and Beyond

***Hoang Le, Yisong Yue, Adith Swaminathan, Byron Boots, Ching-An Cheng***

**Seaside Ballroom, Fri Jun 14, 14:00 PM**

Workshop website:

This workshop aims to bring together researchers from industry and academia in order to describe recent advances and discuss future research

directions pertaining to real-world sequential decision making, broadly construed. We aim to highlight new and emerging research opportunities for the machine learning community that arise from the evolving needs for making decision making theoretically and practically relevant for realistic applications.

Research interest in reinforcement and imitation learning has surged significantly over the past several years, with the empirical successes of self-playing in games and availability of increasingly realistic simulation environments. We believe the time is ripe for the research community to push beyond simulated domains and start exploring research directions that directly address the real-world need for optimal decision making. We are particularly interested in understanding the current theoretical and practical challenges that prevent broader adoption of current policy learning and evaluation algorithms in high-impact applications, across a broad range of domains.

This workshop welcomes both theory and application contributions.

## Schedule

| | |
|---|---|
| 02:00 PM | **Emma Brunskill (Stanford) - Minimizing & Understanding the Data Needed to Learn to Make Good Sequences of Decisions** *Brunskill* |
| 02:30 PM | **Miro Dudík (Microsoft Research) - Doubly Robust Off-policy Evaluation with Shrinkage** *Dudik* |
| 03:00 PM | **Poster Session Part 1 and Coffee Break** |
| 04:00 PM | **Suchi Saria (John Hopkins) - Link between Causal Inference and Reinforcement Learning and Applications to Learning from Offline/Observational Data** *Saria* |
| 04:30 PM | **Dawn Woodard (Uber) - Dynamic Pricing and Matching for Ride-Hailing** *Woodard* |
| 05:00 PM | **Panel Discussion with Emma Brunskill, Miro Dudík, Suchi Saria, Dawn Woodard** |
| 05:30 PM | **Poster Session Part 2** |

Abstracts (2):

Abstract 2: **Miro Dudík (Microsoft Research) - Doubly Robust Off-policy Evaluation with Shrinkage in Real-world Sequential Decision Making: Reinforcement Learning and Beyond**, *Dudik* 02:30 PM

Contextual bandits are a learning protocol that encompasses applications such as news recommendation, advertising, and mobile health, where an algorithm repeatedly observes some information about a user, makes a decision what content to present, and accrues a reward if the presented content is successful. In this talk, I will focus on the fundamental task of evaluating a new policy given historic data. I will describe the asymptotically optimal approach of doubly robust (DR) estimation, the reasons for its shortcomings in finite samples, and how to overcome these shortcomings by directly optimizing the bound on finite-sample error. Optimization yields a new family of estimators that, similarly to DR, leverage any direct model of rewards, but shrink importance weights to obtain a better bias-variance tradeoff than DR. Error bounds can also be used to select the best among multiple reward predictors. Somewhat surprisingly, reward predictors that work best with standard DR are not the same as those that work best with our modified DR. Our new estimator and model selection procedure perform extremely well across a wide variety of settings, so we expect they will enjoy broad practical use.

Based on joint work with Yi Su, Maria Dimakopoulou, and Akshay Krishnamurthy.

Abstract 5: **Dawn Woodard (Uber) - Dynamic Pricing and Matching for Ride-Hailing in Real-world Sequential Decision Making: Reinforcement Learning and Beyond**, *Woodard* 04:30 PM

Ride-hailing platforms like Uber, Lyft, Didi Chuxing, and Ola have achieved explosive growth, in part by improving the efficiency of matching between riders and drivers, and by calibrating the balance of supply and demand through dynamic pricing. We survey methods for dynamic pricing and matching in ride-hailing, and

show that these are critical for providing an experience with low waiting time for both riders and drivers. We also discuss approaches used to predict key inputs into those algorithms: demand, supply, and travel time in the road network. Then we link the two levers together by studying a pool-matching mechanism called dynamic waiting that varies rider waiting and walking before dispatch, which is inspired by a recent carpooling product Express Pool from Uber. We show using data from Uber that by jointly optimizing dynamic pricing and dynamic waiting, price variability can be mitigated, while increasing capacity utilization, trip throughput, and welfare. We also highlight several key practical challenges and directions of future research from a practitioner's perspective.

# June 15, 2019

## Workshop on AI for autonomous driving

***Anna Choromanska, Larry Jackel, Li Erran Li, Juan Carlos Niebles, Adrien Gaidon, Ingmar Posner, Wei-Lun (Harry) Chao***

**101, Sat Jun 15, 08:30 AM**

A diverse set of methods have been devised to develop autonomous driving platforms. They range from modular systems, systems that perform manual decomposition of the problem, systems where the components are optimized independently, and a large number of rules are programmed manually, to end-to-end deep-learning frameworks. Today's systems rely on a subset of the following: camera images, HD maps, inertial measurement units, wheel encoders, and active 3D sensors (LIDAR, radar). There is a general agreement that much of the self-driving software stack will continue to incorporate some form of machine learning in any of the above mentioned systems in the future.

Self-driving cars present one of today's greatest challenges and opportunities for Artificial Intelligence (AI). Despite substantial investments, existing methods for building autonomous vehicles have not yet succeeded, i.e., there are no driverless cars on public roads today without human safety drivers. Nevertheless, a few groups have started working on extending the idea of learned tasks to larger functions of autonomous driving. Initial results on learned road following are very promising.

The goal of this workshop is to explore ways to create a framework that is capable of learning autonomous driving capabilities beyond road following, towards fully driverless cars. The workshop will consider the current state of learning applied to autonomous vehicles and will explore how learning may be used in future systems. The workshop will span both theoretical frameworks and practical issues especially in the area of deep learning.

## Schedule

09:00 AM **Opening Remarks**

09:15 AM **Sven Kreiss: "Compositionality, Confidence and Crowd Modeling for Self-Driving Cars"** *Kreiss, Alahi*

09:40 AM **Mayank Bansal: "ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst"** *Bansal*

10:05 AM **Chelsea Finn: "A Practical View on Generalization and Autonomy in the Real World"** *Finn*

10:50 AM **Sergey Levine: "Imitation, Prediction, and Model-Based Reinforcement Learning for Autonomous Driving"** *Levine*

11:15 AM **Wolfram Burgard** *Burgard*

11:40 AM **Dorsa Sadigh: "Influencing Interactive Mixed-Autonomy Systems"** *Sadigh*

01:30 PM **Poster Session** *Jeong, Philion*

02:30 PM **Alexander Amini: "Learning to Drive with Purpose"** *Amini*

02:55 PM **Fisher Yu: "Motion and Prediction for Autonomous Driving"** *Yu, Darrell*

03:20 PM **Alfredo Canziani: "Model-Predictive Policy Learning with Uncertainty Regularization for Driving in Dense Traffic "** *Canziani*

04:05 PM **Jianxiong Xiao: "Self-driving Car: What we can achieve today?"** *Xiao*

04:30 PM **German Ros: "Fostering Autonomous Driving Research with CARLA"** *Ros*

04:55 PM **Venkatraman Narayanan: "The Promise and Challenge of ML in Self-Driving"** *Narayanan, Bagnell*

05:20 PM **Best Paper Award and Panel Discussion**

Abstracts (7):

Abstract 2: **Sven Kreiss: "Compositionality, Confidence and Crowd Modeling for Self-Driving Cars" in Workshop on AI for autonomous driving**, *Kreiss, Alahi* 09:15 AM

I will present our recent works related to the three AI pillars of a self-driving car: perception, prediction, and planning. For the perception pillar, I will present new human pose estimation and monocular distance estimation methods that use a loss that learns its own confidence, the Laplace loss. For prediction, I will show our investigations on interpretable models where we apply deep learning techniques within structured and hand-crafted classical models for path prediction in social contexts. For the third pillar, planning, I will show our crowd-robot interaction module that uses attention-based representation learning suitable for planning in an RL environment with multiple people.

Abstract 3: **Mayank Bansal: "ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst" in Workshop on AI for autonomous driving**, *Bansal* 09:40 AM

Our goal is to train a policy for autonomous driving via imitation learning that is robust enough to drive a real vehicle. We find that standard behavior cloning is insufficient for handling complex driving scenarios, even when we leverage a perception system for preprocessing the input and a controller for executing the output on the car: 30 million examples are still not enough. We propose exposing the learner to synthesized data in the form of perturbations to the expert's driving, which creates interesting situations such as collisions and/or going off the road. Rather than purely imitating all data, we augment the imitation loss with additional losses that penalize undesirable events and encourage progress -- the perturbations then provide an important signal for these losses and lead to robustness of the learned model. We show that the ChauffeurNet model can handle complex situations in simulation, and present ablation experiments that emphasize the importance of each of our proposed changes and show that the model is responding to the appropriate causal factors. Finally, we demonstrate the model driving a real car at our test facility.

Abstract 5: **Sergey Levine: "Imitation, Prediction, and Model-Based Reinforcement Learning for Autonomous Driving" in Workshop on AI for autonomous driving**, *Levine* 10:50 AM

While machine learning has transformed passive perception -- computer vision, speech recognition, NLP -- its impact on autonomous control in real-world robotic systems has been limited due to reservations about safety and reliability. In this talk, I will discuss how end-to-end learning for control can be framed in a way that is data-driven, reliable and, crucially, easy to merge with existing model-based control pipelines based on planning and state estimation. The basic building blocks of this approach to control are generative models that estimate which states are safe and familiar, and model-based reinforcement learning, which can utilize these generative models within a planning and control framework to make decisions. By framing the end-to-end control problem as one of prediction and generation, we can make it possible to use large datasets collected by previous behavioral policies, as well as human operators, estimate confidence or familiarity of new observations to detect "unknown unknowns," and analyze the performance of our end-to-end models on offline data prior to live deployment. I will discuss how model-based RL can enable navigation and obstacle avoidance, how generative models can detect uncertain and unsafe situations, and then discuss how these pieces can be put together into the framework of deep imitative models: generative models trained via imitation of human drivers that can be incorporated into model-based control for autonomous driving, and can reason about future behavior and intentions of other drivers on the road. Finally, I will conclude with a discussion of current research that is likely to make an impact on autonomous driving and safety-critical AI systems in the near future, including meta-learning, off-policy reinforcement learning, and pixel-level video prediction models.

Abstract 9: **Alexander Amini: "Learning to Drive with Purpose" in Workshop on AI for autonomous driving**, *Amini* 02:30 PM

Deep learning has revolutionized the ability to learn "end-to-end" autonomous vehicle control directly from raw sensory data. In recent years,

there have been advances to handle more complex forms of navigational instruction. However, these networks are still trained on biased human driving data (yielding biased models), and are unable to capture the full distribution of possible actions that could be taken. By learning a set of unsupervised latent variables that characterize the training data, we present an online debiasing algorithm for autonomous driving. Additionally, we extend end-to-end driving networks with the ability to drive with purpose and perform point-to-point navigation. We formulate how our model can be used to also localize the robot according to correspondences between the map and the observed visual road topology, inspired by the rough localization that human drivers can perform, even in cases where GPS is noisy or removed all together. Our results highlight the importance of bridging the benefits from end-to-end learning with classical probabilistic reasoning and Bayesian inference to push the boundaries of autonomous driving.

Abstract 11: **Alfredo Canziani: "Model-Predictive Policy Learning with Uncertainty Regularization for Driving in Dense Traffic " in Workshop on AI for autonomous driving**, *Canziani* 03:20 PM

Learning a policy using only observational data is challenging because the distribution of states it induces at execution time may differ from the distribution observed during training. In this work, we propose to train a policy while explicitly penalizing the mismatch between these two distributions over a fixed time horizon. We do this by using a learned model of the environment dynamics which is unrolled for multiple time steps, and training a policy network to minimize a differentiable cost over this rolled-out trajectory. This cost contains two terms: a policy cost which represents the objective the policy seeks to optimize, and an uncertainty cost which represents its divergence from the states it is trained on. We propose to measure this second cost by using the uncertainty of the dynamics model about its own predictions, using recent ideas from uncertainty estimation for deep networks. We evaluate our approach using a large-scale observational dataset of driving behavior recorded from traffic cameras, and show that we are able to learn effective driving policies from purely observational data, with no environment interaction.

Abstract 13: **German Ros: "Fostering Autonomous Driving Research with CARLA" in Workshop on AI for autonomous driving**, *Ros* 04:30 PM

This talk focuses on the relevance of open source solutions to foster autonomous driving research and development. To this end, we present how CARLA has been used within the research community in the last year and what results has it enabled. We will also cover the CARLA Autonomous Driving Challenge and its relevance as an open benchmark for the driving community. We will share with the community new soon-to-be-released features and the future direction of the CARLA simulation platform.

Abstract 14: **Venkatraman Narayanan: "The Promise and Challenge of ML in Self-Driving" in Workshop on AI for autonomous driving**, *Narayanan, Bagnell* 04:55 PM

To deliver the benefits of autonomous driving safely, quickly, and broadly, learnability has to be a key element of the solution. In this talk, I will describe Aurora's philosophy towards building learnability into the self-driving architecture, avoiding the pitfalls of applying vanilla ML to problems involving feedback, and leveraging expert demonstrations for learning decision-making models. I will conclude with our approach to testing and validation.

## Workshop on Multi-Task and Lifelong Reinforcement Learning

***Sarath Chandar, Shagun Sodhani, Khimya Khetarpal, Tom Zahavy, Daniel J. Mankowitz, Shie Mannor, Balaraman Ravindran, Doina Precup, Chelsea Finn, Abhishek Gupta, Amy Zhang, Kyunghyun Cho, Andrei Rusu, Facebook Rob Fergus***

**102, Sat Jun 15, 08:30 AM**

Website link: https://sites.google.com/view/mtlrl/

Significant progress has been made in reinforcement learning, enabling agents to accomplish complex tasks such as Atari games, robotic manipulation, simulated locomotion, and Go. These successes have stemmed from the core reinforcement learning formulation of learning a single policy or value function from scratch. However, reinforcement learning has proven challenging to scale to many practical real world problems due to problems in learning efficiency and objective specification, among many others. Recently, there has been emerging interest and research in leveraging structure and information across multiple reinforcement learning tasks to more efficiently and effectively learn complex behaviors. This includes:

1. curriculum and lifelong learning, where the problem requires learning a sequence of tasks, leveraging their shared structure to enable knowledge transfer
2. goal-conditioned reinforcement learning techniques that leverage the structure of the provided goal space to learn many tasks significantly faster
3. meta-learning methods that aim to learn efficient learning algorithms that can learn new tasks quickly
4. hierarchical reinforcement learning, where the reinforcement learning problem might entail a compositions of subgoals or subtasks with shared structure

Multi-task and lifelong reinforcement learning has the potential to alter the paradigm of traditional reinforcement learning, to provide more practical and diverse sources of supervision, while helping overcome many challenges associated with reinforcement learning, such as exploration, sample efficiency and credit assignment. However, the field of multi-task and lifelong reinforcement learning is still young, with many more developments needed in terms of problem formulation, algorithmic and theoretical advances as well as better benchmarking and evaluation.

The focus of this workshop will be on both the algorithmic and theoretical foundations of multi-task and lifelong reinforcement learning as well as the practical challenges associated with building multi-tasking agents and lifelong learning benchmarks. Our goal is to bring together researchers that study different problem domains (such as games, robotics, language, and so forth), different optimization approaches (deep learning, evolutionary algorithms, model-based control, etc.), and different formalisms (as mentioned above) to discuss the frontiers, open problems and meaningful next steps in multi-task and lifelong reinforcement learning.

## Schedule

| | |
|---|---|
| 08:45 AM | **Opening Remarks** |
| 09:00 AM | **Sergey Levine: Unsupervised Reinforcement Learning and Meta-Learning** *Levine* |
| 09:25 AM | **Spotlight Presentations** |
| 09:50 AM | **Peter Stone: Learning Curricula for Transfer Learning in RL** *Stone* |
| 10:15 AM | **Contributed Talks** |
| 10:30 AM | **Posters and Break** |
| 11:00 AM | **Jacob Andreas: Linguistic Scaffolds for Policy Learning** *Andreas* |
| 11:25 AM | **Karol Hausman: Skill Representation and Supervision in Multi-Task Reinforcement Learning** *Hausman* |
| 11:50 AM | **Contributed Talks** |
| 12:20 PM | **Posters and Lunch Break** |
| 02:00 PM | **Martha White: Learning Representations for Continual Learning** *White* |
| 02:25 PM | **Natalia Diaz-Rodriguez: Continual Learning and Robotics: an overview** *Diaz Rodriguez* |
| 02:50 PM | **Posters and Break** |
| 03:30 PM | **Jeff Clune: Towards Solving Catastrophic Forgetting with Neuromodulation & Learning Curricula by Generating Environments** *Clune* |
| 03:55 PM | **Contributed Talks** |

04:15 PM **Nicolas Heess: TBD** *Heess*
04:40 PM **Benjamin Rosman: Exploiting Structure For Accelerating Reinforcement Learning** *Rosman*
05:05 PM **Panel Discussion**

Abstracts (3):

Abstract 5: **Contributed Talks in Workshop on Multi-Task and Lifelong Reinforcement Learning**, 10:15 AM

10:15 Meta-Learning via Learned Loss Yevgen Chebotar*, Artem Molchanov*, Sarah Bechtle*, Ludovic Righetti, Franziska Meier, Gaurav Sukhatme
10:23 MCP: Learning Composable Hierarchical Control with Multiplicative Compositional Policies Xue Bin Peng, Michael Chang, Grace Zhang Pieter Abbeel, Sergey Levine

Abstract 9: **Contributed Talks in Workshop on Multi-Task and Lifelong Reinforcement Learning**, 11:50 AM

11:50 Which Tasks Should Be Learned Together in Multi-task Learning? Trevor Standley, Amir R. Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, Silvio Savarese
11:58 Finetuning Subpolicies for Hierarchical Reinforcement Learning Carlos Florensa, Alexander Li and Pieter Abbeel
12:06 Online Learning for Auxiliary Task Weighting for Reinforcement Learning Xingyu Lin*, Harjatin Singh Baweja*, George Kantor, David Held
12:14 Guided Meta-Policy Search Russell Mendonca, Abhishek Gupta, Rosen Kralev, Pieter Abbeel, Sergey Levine, Chelsea Finn

Abstract 15: **Contributed Talks in Workshop on Multi-Task and Lifelong Reinforcement Learning**, 03:55 PM

03:55 Online Continual Learning with Maximally Inferred Retrieval Rahaf Alundji*, Lucas Caccia*, Eugene Belilovsky*, Massimo Caccia*, Min Lin, Laurent Charlin, Tinne Tuytelaars

04:05 Skew-Fit: State-Covering Self-Supervised Reinforcement Learning Vitchyr H. Pong *, Murtaza Dalal*, Steven Lin, Ashvin Nair, Shikhar Bahl, Sergey Levine

## Invertible Neural Networks and Normalizing Flows

***Chin-Wei Huang, David Krueger, Rianne Van den Berg, George Papamakarios, Aidan Gomez, Chris Cremer, Ricky T. Q. Chen, Aaron Courville, Danilo J. Rezende***

**103, Sat Jun 15, 08:30 AM**

Invertible neural networks have been a significant thread of research in the ICML community for several years. Such transformations can offer a range of unique benefits:

(1) They preserve information, allowing perfect reconstruction (up to numerical limits) and obviating the need to store hidden activations in memory for backpropagation.
(2) They are often designed to track the changes in probability density that applying the transformation induces (as in normalizing flows).
(3) Like autoregressive models, normalizing flows can be powerful generative models which allow exact likelihood computations; with the right architecture, they can also allow for much cheaper sampling than autoregressive models.

While many researchers are aware of these topics and intrigued by several high-profile papers, few are familiar enough with the technical details to easily follow new developments and contribute. Many may also be unaware of the wide range of applications of invertible neural networks, beyond generative modelling and variational inference.

## Schedule

09:30 AM **Tutorial on normalizing flows** *Jang*
10:30 AM **Poster Spotlights**

| | |
|---|---|
| 10:50 AM | **poster session I** *Rhinehart, Tang, Prabhu, Yap, Wang, Finzi, Kumar, Lu, Kumar, Lei, Przystupa, De Cao, Kirichenko, Izmailov, Wilson, Kruse, Mesquita, Lezcano Casado, Müller, Simmons, Atanov* |
| 11:30 AM | **Building a tractable generator network** |
| 11:50 AM | **Glow: Generative Flow with Invertible 1x1 Convolutions** *Dhariwal* |
| 12:10 PM | **Contributed talk** |
| 02:00 PM | **Householder meets Sylvester: Normalizing flows for variational inference** |
| 02:20 PM | **Neural Ordinary Differential Equations for Continuous Normalizing Flows** |
| 02:40 PM | **Contributed talk** |
| 03:00 PM | **poster session II** |
| 04:00 PM | **The Bijector API: An Invertible Function Library for TensorFlow** |
| 04:20 PM | **Invertible Neural Networks for Understanding and Controlling Learned Representations** |
| 04:40 PM | **Contributed talk** |
| 05:00 PM | **Panel Session** |

## Stein’s Method for Machine Learning and Statistics

***Francois-Xavier Briol, Lester Mackey, Chris Oates, Qiang Liu, Larry Goldstein***

**104 A, Sat Jun 15, 08:30 AM**

Stein's method is a technique from probability theory for bounding the distance between probability measures using differential and difference operators. Although the method was initially designed as a technique for proving central limit theorems, it has recently caught the attention of the machine learning (ML) community and has been used for a variety of practical tasks. Recent applications include generative modeling, global non-convex optimisation, variational inference, de novo sampling, constructing powerful control variates for Monte Carlo variance reduction, and measuring the quality of (approximate) Markov chain Monte Carlo algorithms. Stein's method has also been used to develop goodness-of-fit tests and was the foundational tool in one of the NeurIPS 2017 Best Paper awards.

Although Stein's method has already had significant impact in ML, most of the applications only scratch the surface of this rich area of research in probability theory. There would be significant gains to be made by encouraging both communities to interact directly, and this inaugural workshop would be an important step in this direction. More precisely, the aims are: (i) to introduce this emerging topic to the wider ML community, (ii) to highlight the wide range of existing applications in ML, and (iii) to bring together experts in Stein's method and ML researchers to discuss and explore potential further uses of Stein's method.

### Schedule

| | |
|---|---|
| 08:30 AM | **Overview of the day** *Briol* |
| 08:45 AM | **Tutorial - Larry Goldstein: The Many Faces of a Simple Identity** *Goldstein* |
| 09:45 AM | **Invited Talk - Anima Anandkumar: Stein’s method for understanding optimization in neural networks.** *Anandkumar* |
| 10:30 AM | **Break** |
| 11:00 AM | **Invited Talk - Arthur Gretton: Relative goodness-of-fit tests for models with latent variables.** *Gretton* |
| 11:45 AM | **Invited Talk - Andrew Duncan** *Duncan* |
| 12:30 PM | **Lunch and Poster Session** |
| 01:45 PM | **Invited Talk - Yingzhen Li: Gradient estimation for implicit models with Stein's method.** *Li* |
| 02:30 PM | **Invited Talk - Ruiyi Zhang: On Wasserstein Gradient Flows and Particle-Based Variational Inference** *ZHANG* |

03:15 PM **Break**

03:45 PM **Invited Talk - Paul Valiant: How the Ornstein-Uhlenbeck process drives generalization for deep learning.**

04:30 PM **Invited Talk - Louis Chen: Palm theory, random measures and Stein couplings.** *Chen*

05:15 PM **Panel Discussion - All speakers**

Abstracts (7):

Abstract 2: **Tutorial - Larry Goldstein: The Many Faces of a Simple Identity in Stein's Method for Machine Learning and Statistics**, *Goldstein* 08:45 AM

Stein's identity states that a mean zero, variance one random variable W has the standard normal distribution if and only if E[Wf(W)] = E[f'(W)] for all f in F, where F is the class of functions such that both sides of the equality exist. Stein (1972) used this identity as the key element in his novel method for proving the Central Limit Theorem. The method generalizes to many distributions beyond the normal, allows one to operate on random variables directly rather than through their transforms, provides non-asymptotic bounds to their limit, and handles a variety of dependence. In addition to distributional approximation, Stein's identity has connections to concentration of measure, Malliavin calculus, and statistical procedures including shrinkage and unbiased risk estimation.

Abstract 3: **Invited Talk - Anima Anandkumar: Stein's method for understanding optimization in neural networks. in Stein's Method for Machine Learning and Statistics**, *Anandkumar* 09:45 AM

Training neural networks is a challenging non-convex optimization problem. Stein's method provides a novel way to change optimization problem to a tensor decomposition problem for guaranteed training of two-layer neural networks. We provide risk bounds for our proposed method, with a polynomial sample complexity in the relevant parameters, such as input dimension and number of neurons. Our training method is based on tensor decomposition, which provably converges to the global optimum, under a set of mild non-degeneracy conditions. This provides insights into role of generative process for tractability of supervised learning.

Abstract 5: **Invited Talk - Arthur Gretton: Relative goodness-of-fit tests for models with latent variables. in Stein's Method for Machine Learning and Statistics**, *Gretton* 11:00 AM

I will describe a nonparametric, kernel-based test to assess the relative goodness of fit of latent variable models with intractable unnormalized densities. The test generalises the kernel Stein discrepancy (KSD) tests of (Liu et al., 2016, Chwialkowski et al., 2016, Yang et al., 2018, Jitkrittum et al., 2018) which require exact access to unnormalized densities. We will rely on the simple idea of using an approximate observed-variable marginal in place of the exact, intractable one. As the main theoretical contribution, the new test has a well-controlled type-I error, once we have properly corrected the threshold. In the case of models with low-dimensional latent structure and high-dimensional observations, our test significantly outperforms the relative maximum mean discrepancy test, which cannot exploit the latent structure.

Abstract 8: **Invited Talk - Yingzhen Li: Gradient estimation for implicit models with Stein's method. in Stein's Method for Machine Learning and Statistics**, *Li* 01:45 PM

Implicit models, which allow for the generation of samples but not for point-wise evaluation of probabilities, are omnipresent in real-world problems tackled by machine learning and a hot topic of current research. Some examples include data simulators that are widely used in engineering and scientific research, generative adversarial networks (GANs) for image synthesis, and hot-off-the-press approximate inference techniques relying on implicit distributions. Gradient based optimization/sampling methods are often applied to train these models, however without tractable densities, the objective functions often need to be approximated. In this

talk I will motivate gradient estimation as another approximation approach for training implicit models and perform Monte Carlo based approximate inference. Based on this view, I will then present the Stein gradient estimator which estimates the score function of an implicit model density. I will discuss connections of this approach to score matching, kernel methods, denoising auto-encoders, etc., and show application cases including entropy regularization for GANs, and meta-learning for stochastic gradient MCMC algorithms.

Abstract 9: **Invited Talk - Ruiyi Zhang: On Wasserstein Gradient Flows and Particle-Based Variational Inference in Stein's Method for Machine Learning and Statistics**, *ZHANG* 02:30 PM

Particle-based variational inference methods (ParVIs), such as models associated with Stein variational gradient descent, have gained attention in the Bayesian inference literature, for their capacity to yield flexible and accurate approximations. We explore ParVIs from the perspective of Wasserstein gradient flows, and make both theoretical and practical contributions. We unify various finite-particle approximations that existing ParVIs use, and recognize that the approximation is essentially a compulsory smoothing treatment, in either of two equivalent forms. This novel understanding reveals the assumptions and relations of existing ParVIs, and also inspires new ParVIs. We propose an acceleration framework and a principled bandwidth-selection method for general ParVIs; these are based on the developed theory and leverage the geometry of the Wasserstein space. Experimental results show the improved convergence by the acceleration framework and enhanced sample accuracy by the bandwidth-selection method.

Abstract 11: **Invited Talk - Paul Valiant: How the Ornstein-Uhlenbeck process drives generalization for deep learning. in Stein's Method for Machine Learning and Statistics**, 03:45 PM

After discussing the Ornstein-Uhlenbeck process and how it arises in the context of Stein's method, we turn to an analysis of the stochastic gradient descent method that drives deep learning. We show that certain noise that often arises in the training process induces an Ornstein-Uhlenbeck process on the learned parameters. This process is responsible for a weak regularization effect on the training that, once it reaches stable points, will have provable found the "simplest possible" hypothesis consistent with the training data. At a higher level, we argue how some of the big mysteries of the success of deep learning may be revealed by analyses of the subtle stochastic processes which govern deep learning training, a natural focus for this community.

Abstract 12: **Invited Talk - Louis Chen: Palm theory, random measures and Stein couplings. in Stein's Method for Machine Learning and Statistics**, *Chen* 04:30 PM

In this talk, we present a general Kolmogorov bound for normal approximation, proved using Stein's method. We will apply this bound to random measures using Palm theory and coupling, with applications to stochastic geometry. We will also apply this bound to Stein couplings, which include local dependence, exchangeable pairs, size-bias couplings and more as special cases. This talk is based on joint work with Adrian Roellin and Aihua Xia.

## AI For Social Good (AISG)

***Margaux Luck, Kris Sankaran, Tristan Sylvain, Sean McGregor, Jonnie Penn, Girmaw Abebe Tadesse, Virgile Sylvain, Myriam Côté, Lester Mackey, Rayid Ghani, Yoshua Bengio***

**104 B, Sat Jun 15, 08:30 AM**

#AI for Social Good
##Important information
**Contact information:**
[aisg2019.icml.contact@gmail.com]
(mailto:aisg2019.icml.contact@gmail.com)

**Submission deadline:** ***EXTENDED to April 26th 2019 11:59PM ET***

**[Workshop website](https://aiforsocialgood.github.io/icml2019/index.htm)**

**[Submission website](https://cmt3.research.microsoft.com/AISGW2019/)**

**Poster Information:**

* Poster Size - ** 36W x 48H inches or 90 x 122 cm**

* Poster Paper - **lightweight paper - not laminated**

##Abstract
This workshop builds on our AI for Social Good workshop at [NeurIPS 2018](https://aiforsocialgood.github.io/2018/) and [ICLR 2019](https://aiforsocialgood.github.io/iclr2019/).

**Introduction:** The rapid expansion of AI research presents two clear conundrums:

- the comparative lack of incentives for researchers to address social impact issues and
- the dearth of conferences and journals centered around the topic. Researchers motivated to help often find themselves without a clear idea of which fields to delve into.

**Goals:** Our workshop address both these issues by bringing together machine learning researchers, social impact leaders, stakeholders, policy leaders, and philanthropists to discuss their ideas and applications for social good. To broaden the impact beyond the convening of our workshop, we are partnering with [AI Commons](http://www.aicommons.com) to expose accepted projects and papers to the broader community of machine learning researchers and engineers. The projects/research may be at varying degrees of development, from formulation as a data problem to detailed requirements for effective deployment. We hope that this gathering of talent and information will inspire the creation of new approaches and tools by the community, help scientists access the data they need, involve social and policy stakeholders in the framing of machine learning applications, and attract interest from philanthropists invited to the event to make a dent in our shared goals.

**Topics:** The UN Sustainable Development Goals (SDGs), a set of seventeen objectives whose completion is set to lead to a more equitable, prosperous, and sustainable world. In this light, our main areas of focus are the following: health, education, the protection of democracy, urban planning, assistive technology, agriculture, environmental protection and sustainability, social welfare and justice, developing world. Each of these themes presents unique opportunities for AI to reduce human suffering and allow citizens and democratic institutions to thrive.

Across these topics, we have dual goals: recognizing high-quality work in machine learning motivated by or applied to social applications, and creating meaningful connections between communities dedicated to solving technical and social problems. To this extent, we propose two research tracks:

- **Short Papers Track (Up to four page papers + unlimited pages for citations)** for oral and/or poster presentation. The short papers should focus on past and current research work, showcasing actual results and demonstrating beneficial effects on society. We also accept short papers of recently published or submitted journal contributions to give authors the opportunity to present their work and obtain feedback from conference attendees.
- **Problem Introduction Track (Application form, up to five page responses + unlimited pages for citations)** which will present a specific solution that will be shared with stakeholders, scientists, and funders. The workshop will provide a suite of questions designed to: (1) estimate the feasibility and impact of the proposed solutions, and (2) estimate the importance of data in their implementation. The application responses should highlight ideas that have not yet been implemented in practice but can lead to real impact. The projects may be at varying degrees of development, from formulation as a data problem to structure for effective deployment. The workshop provides a supportive platform for developing these early-stage or hobby proposals into real projects. This process is designed to foster sharing different points of view ranging from the scientific assessment of feasibility, discussion of practical constraints that may be encountered, and attracting interest from philanthropists invited to the event. Accepted

submissions may be promoted to the wider AI solutions community following the workshop via the [AI Commons](http://www.aicommons.com), with whom we are partnering to promote the longer-term development of projects.

## Schedule

08:45 AM **Welcoming and Poster set-up**

09:00 AM **Opening remarks** *Bengio*

09:05 AM **Solving societal challenges with AI through partnerships** *Anandan*

09:45 AM **AI Commons** *Bengio*

09:50 AM **Detecting Waterborne Debris with Sim2Real and Randomization** *Sankaran*

10:00 AM **Conversational agents to address abusive online behaviors** *Beauxis-Aussalet*

10:10 AM **Computer Vision For Food Quality: The Case of Injera**

10:20 AM **AI for Mitigating Effects of Climate and Weather Changes in Agriculture**

10:30 AM **Break / Poster Session 1**

11:00 AM **AI for Ecology and Conservation** *Sheldon*

11:40 AM **Using AI for Economic Upliftment of Handicraft Industry** *Damani*

11:50 AM **Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening** *Geras, Wu*

12:00 PM **Lunch - on your own**

02:00 PM **AI for Whose Social Good?**

02:30 PM **How to Volunteer as an AI Researcher** *High*

03:00 PM **Break / Poster Session 2**

03:30 PM **Poster Session** *Fang, Balashankar, Damani, Beauxis-Aussalet, Wu, Bondi, Rußwurm, Ruhe, Saxena, Spoon*

03:50 PM **Creating constructive change and avoiding unintended consequences from machine learning**

04:20 PM **Learning Global Variations in Outdoor PM_2.5 Concentrations with Satellite Images**

04:30 PM **Pareto Efficient Fairness for Skewed Subgroup Data** *Balashankar*

04:40 PM **Crisis Sub-Events on Social Media: A Case Study of Wildfires**

04:50 PM **Towards Detecting Dyslexia in Children's Handwriting Using Neural Networks**

05:00 PM **Assisting Vulnerable Communities through AI and OR: from Data to Deployed Decisions**

05:30 PM **Open announcement and Best Paper Award**

Abstracts (18):

Abstract 2: **Opening remarks in AI For Social Good (AISG)**, *Bengio* 09:00 AM

Speaker bio:
Yoshua Bengio is Full Professor of the Department of Computer Science and Operations Research, scientific director of Mila, CIFAR Program co-director of the CIFAR Learning in Machines and Brains program (formerly Neural Computation and Adaptive Perception), scientific director of IVADO and Canada Research Chair in Statistical Learning Algorithms. His main research ambition is to understand principles of learning that yield intelligence. He supervises a large group of graduate students and post-docs. His research is widely cited (over 130000 citations found by Google Scholar in August 2018, with an H-index over 120, and rising fast).

Abstract 3: **Solving societal challenges with AI through partnerships in AI For Social Good (AISG)**, *Anandan* 09:05 AM

Wadhwani AI was inaugurated a little more than a year ago with the mission of bringing the power of AI to address societal challenges, especially among underserved communities throughout the world. We aim to address problems all major domains including health, agriculture, education, infrastructure, and financial inclusion. We are currently working on three solutions (two in health and one in agriculture) and are exploring more areas where we can apply AI for social

good.The most important lesson that we have learned during our short stint is the importance of working in close partnership with other stakeholders and players in the social sectors, especially NGOs and Government organizations. In this talk, I will use one case, namely that of developing an AI based approach for Integrated Pest Management (IPM) in Cotton Farming, to describe how this partnership based approach has evolved and been critical to our solution development and implementation.

Speaker bio:
Dr. P. Anandan is the CEO of Wadhwani Institute of Artificial Intelligence. His prior experience includes - Adobe Research Lab India (2016-2017) as a VP for Research and a Distinguished Scientist and Managing Director at Microsoft Research (1997-2014). He was also the founding director of Microsoft Research India which he ran from 2005-2014. Earlier stint was at Sarnoff Corporation (1991-1997) as a researcher and an Assistant Professor of Computer Science at Yale University (1987-1991). His primary research area is Computer vision where he is well known for his fundamental and lasting contributions to the problem of visual motion analysis. He received his PhD in Computer Science from University of Massachusetts, Amherst in 1987, a Masters in Computer Science from University of Nebraska, Lincoln in 1979 and his B.Tech in Electrical Engineering from IIT Madras, India in 1977. He is a distinguished alumnus of IIT Madras, and UMass, Amherst and is on the Nebraska Hall of Computing. His hobbies include playing African drums, writing poems (in Tamil) and travel which makes his work related travel interesting.

Abstract 4: **AI Commons in AI For Social Good (AISG)**, *Bengio* 09:45 AM

AI Commons is a collective project whose goal is to make the benefits of AI available to all. Since AI research can benefit from the input of a large range of talents across the world, the project seeks to develop ways for developers and organizations to collaborate more easily and effectively. As a community operating in an environment of trust and problem-solving, AI Commons can empower researchers to tackle the world's important problems using all the possibilities of cutting-edge AI.

Speaker bio:
Yoshua Bengio is Full Professor of the Department of Computer Science and Operations Research, scientific director of Mila, CIFAR Program co-director of the CIFAR Learning in Machines and Brains program (formerly Neural Computation and Adaptive Perception), scientific director of IVADO and Canada Research Chair in Statistical Learning Algorithms. His main research ambition is to understand principles of learning that yield intelligence. He supervises a large group of graduate students and post-docs. His research is widely cited (over 130000 citations found by Google Scholar in August 2018, with an H-index over 120, and rising fast).

Abstract 5: **Detecting Waterborne Debris with Sim2Real and Randomization in AI For Social Good (AISG)**, *Sankaran* 09:50 AM

Marine debris pollution is one of the most ubiquitous and pressing environmental issues affecting our oceans today. Clean up efforts such as the Great Pacific Garbage Patch project have been implemented across the planet to combat this problem. However, resources to accomplish this goal are limited, and the afflicted area is vast. To this end, unmanned vehicles that are capable of automatically detecting and removing small-sized debris would be a great complementary approach to existing large-scale garbage collectors. Due to the complexity of fully functioning unmanned vehicles for both detecting and removing debris, in this project, we focus on the detection task as a first step. From the perspective of machine learning, there is an unfortunate lack of sufficient labeled data for training a specialized detector, e.g., a classifier that can distinguish debris from other objects like wild animals. Moreover, pre-trained detectors on other domains would be ineffective while creating such datasets manually would be very costly. Due to the recent progress of training deep models with synthetic data and domain randomization, we propose to train a debris detector based on a mixture of real and synthetic images.

Speaker bio:
Kris is a postdoc at Mila working with Yoshua Bengio on problems related to Humanitarian AI.

He is generally interested in ways to broaden the scope of problems studied by the machine learning community and am curious about the ways to bridge statistical and computational thinking.

Abstract 6: **Conversational agents to address abusive online behaviors in AI For Social Good (AISG)**, *Beauxis-Aussalet* 10:00 AM

Technologies to address cyber bullying are limited to detecting and hiding abusive messages. We propose to investigate the potential of conversational technologies for addressing abusers. We will outline directions for studying the effectiveness dialog strategies (e.g., to educate or deter abusers, or keep them busy with chatbots rather than their victims) and for initiating new research on chatbot-mediated mitigation of online abuse.

Speaker bio:
Emma Beauxis-Aussalet is a Senior Track Associate at the Digital Society School of Amsterdam University of Applied Science, where she investigates how data-driven technologies can be applied for the best interests of society. She holds a PhD on classification errors and biases from Utrecht University. Her interests include ethical and explainable AI, data literacy in the general public, and the synergy between human & artificial intelligence to tackle job automation.

Abstract 7: **Computer Vision For Food Quality: The Case of Injera in AI For Social Good (AISG)**, 10:10 AM

The use of Teff as an exclusive crop for making Injera, Ethiopian national staple, has changed overtime.Driven by the ever increasing price of Teff, producers have added other ingredients, of which some are good (maze and rice), while others are not. Hence, households opting for the industrial solution of Injera, are disturbed by the fact that hey can not figure out what exactly is contained in their Injera. Thousands of local producers and local shopkeepers work together to make fresh Injera available to millions around the country. However, consumers are finding it more and more difficult to find a safe Injera for purchase. This injera is usually sold unpacked, unlabeled and in an unsafe way through local shops. This being so, consumers face more and more health risks, all the more as it is impossible to evaluate the ingredients contained in the Injera they are buying There are two kinds of risks: (a) the local producers might try to reduce the cost by using cheap ingredients, including risky additives, and (b) the shops might sell expired Injera warmed up. We discuss here the growing food safety problem faced by millions of Injera consumers in Ethiopia, and the possibility of using AI to solve this problem.

Speaker bio:
Wondimagegnehu is a master’s student in Information Science at Addis Ababa University. He is working on a master's thesis in learning an optimal representation of word structure for morphological complex languages under a constrained settings: limited training data and human supervision. He is interested in exploring research challenges in using AI on a social setting.

Abstract 8: **AI for Mitigating Effects of Climate and Weather Changes in Agriculture in AI For Social Good (AISG)**, 10:20 AM

In recent years, floods, landslides and droughts have become an annual occurrence in Sri Lanka. Despite the efforts made by the government and other entities, these natural disasters remain challenging mainly to the people who live in high risk areas. It is also crucial to predict such disasters early on to facilitate evacuation of people living in these areas. Furthermore, Sri Lankan economy largely depends on agriculture, yet this sector still remains untouched by recent advancements of AI and other predictive analytics techniques. The solution is to develop an AI based platform that generates insights from emerging data sources. It will be modular, extensible and open source. Similar to any other real world AI system, the end solution will consist of multiple data pipelines to extract data, analyze and present results through APIs. The presentation layer will be a public API that can be consumed through a portal such as Disaster Management Centre of Sri Lanka.

Speaker bio:

Narmada is research engineer at ConscientAI Labs based in Sri Lanka. She is also a visiting research student at the Memorial University of Newfoundland, Canada. She is interested in research on climate change and effects of it on human lifestyle and Deep Learning for Computer Vision.

Abstract 10: **AI for Ecology and Conservation in AI For Social Good (AISG)**, *Sheldon* 11:00 AM

AI can help solve big data and decision-making problems to understand and protect the environment. I'll survey several projects the area and discuss how to approach environmental problems using AI. The Dark Ecology project uses weather radar and machine learning to unravel mysteries of bird migration. A surprising probabilistic inference problem arises when analyzing animal survey data to monitor populations. Novel optimization algorithms can help reason about dams, hydropower, and the ecology of river networks.

Speaker bio:
Daniel Sheldon is an Assistant Professor of Computer Science at the University of Massachusetts Amherst and Mount Holyoke College. His research investigates fundamental problems in machine learning and AI motived by large-scale environmental data, dynamic ecological processes, and real-world network phenomena.

Abstract 11: **Using AI for Economic Upliftment of Handicraft Industry in AI For Social Good (AISG)**, *Damani* 11:40 AM

The handicraft industry is a strong pillar of Indian economy which provides large-scale employment opportunities to artisans in rural and underprivileged communities. However, in this era of globalization, diverse modern designs have rendered traditional designs old and monotonous, causing an alarming decline of handicraft sales. In this talk, we will discuss our approach leveraging techniques like GANs, Color Transfer, Pattern Generation etc. to generate contemporary designs for two popular Indian handicrafts - Ikat and Block Print. The resultant designs are evaluated to be significantly more likeable and marketable than the current designs used by artisans.

Speaker bio:
Sonam Damani is an Applied Scientist in Microsoft, India where she has worked on several projects in the field of AI and Deep Learning, including Microsoft's human-like-chatbot Ruuh, Cortana personality, novel art generation using AI, Bing search relevance, among others. In the past year, she has co-authored a bunch of publications in the field of conversational AI and AI creativity that were presented in NeurIPS, WWW and CODS-COMAD.

Abstract 12: **Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening in AI For Social Good (AISG)**, *Geras, Wu* 11:50 AM

We present a deep CNN for breast cancer screening exam classification, trained and evaluated on over 200,000 exams (over 1,000,000 images). Our network achieves an AUC of 0.895 in predicting whether there is a cancer in the breast, when tested on the screening population. We attribute the high accuracy of our model to a two-stage training procedure, which allows us to use a very high-capacity patch-level network to learn from pixel-level labels alongside a network learning from macroscopic breast-level labels. To validate our model, we conducted a reader study with 14 readers, each reading 720 screening mammogram exams, and find our model to be as accurate as experienced radiologists when presented with the same data. Finally, we show that a hybrid model, averaging probability of malignancy predicted by a radiologist with a prediction of our neural network, is more accurate than either of the two separately.

Speaker bio:
- Krzysztof Geras is an assistant professor at NYU School of Medicine and an affiliated faculty at NYU Center for Data Science. His main interests are in unsupervised learning with neural networks, model compression, transfer learning, evaluation of machine learning models and applications of these techniques to medical imaging. He previously did a postdoc at NYU with

Kyunghyun Cho, a PhD at the University of Edinburgh with Charles Sutton and an MSc as a visiting student at the University of Edinburgh with Amos Storkey. His BSc is from the University of Warsaw.

- Nan Wu is a PhD student at NYU Center for Data Science. She is interested in data science with application to healthcare and currently working on medical image analysis. Before joining NYU, she graduated from School for Gifted Young, University of Science and Technology of China, receiving B.S in Statistics and B.A. in Business Administration.

Abstract 14: **AI for Whose Social Good? in AI For Social Good (AISG)**, 02:00 PM

Luke Stark will discuss two recent papers (Greene, Hoffmann & Stark 2019; Stark & Hoffmann 2019) that use discursive analysis to examine a) recent high-profile value statements endorsing ethical design for artificial intelligence and machine learning and b) professional ethics codes in computer science, statistics, and other fields. Guided by insights from Science and Technology Studies, values in design, and the sociology of business ethics, he will discuss the grounding assumptions and terms of debate that shape current conversations about ethical design in data science and AI. He will also advocate for an expanded view of expertise in understanding what ethical AI/ML/AI for Social Good should mean.

Speaker bio:
Luke Stark is a Postdoctoral Researcher in the Fairness, Accountability, Transparency and Ethics (FATE) Group at Microsoft Research Montreal, and an Affiliate of the Berkman Klein Center for Internet & Society at Harvard University. Luke holds a PhD from the Department of Media, Culture, and Communication at New York University, and an Honours BA and MA in History from the University of Toronto. Trained as a media historian, his scholarship centers on the interconnected histories of artificial intelligence (AI) and behavioral science, and on the ways the social and ethical contexts of AI are changing how we work, communicate, and participate in civic life.

Abstract 15: **How to Volunteer as an AI Researcher in AI For Social Good (AISG)**, *High* 02:30 PM

Over the past six years, Will High has volunteered his expertise as a data scientist to various nonprofits and civic causes. He's contributed to work on homelessness, improving charter schools and optimizing water distribution. Will will talk about his experience doing pro-bono work with DataKind, a global nonprofit based in New York that connects leading social change organizations with data science talent to collaborate on cutting-edge analytics and advanced algorithms developed to maximize social impact. He'll comment on DataKind's mission, how to structure effective pro-bono engagements, and broader principles of the pro bono model applied to machine learning, analytics and engineering.

Speaker bio:
Will is a data science executive at Joymode in Los Angeles and works with DataKind as Data Ambassador, consultant and facilitator. Will was previously a Senior Data Scientist at Netflix. He holds a PhD in physics from Harvard.

Abstract 18: **Creating constructive change and avoiding unintended consequences from machine learning in AI For Social Good (AISG)**, 03:50 PM

This talk will give an overview of some of the known failure modes that are leading to unintended consequences in AI development, as well as research agendas and initiatives to mitigate them, including a number that are underway at the Partnership on AI (PAI). Important case studies include the use of algorithmic risk assessment tools in the US criminal justice system, and the side-effects that are caused by using deliberate or unintended optimization processes to design high-stakes technical and bureaucratic system. These are important in their own right, but they are also important contributors to conversations about social good applications of AI, which are also subject to significant potential for unintended consequences.

Speaker bio:
Peter Eckersley is Director of Research at the Partnership on AI, a collaboration between the major technology companies, civil society and academia to ensure that AI is designed and used to benefit humanity. He leads PAI's research on machine learning policy and ethics, including projects within PAI itself and projects in collaboration with the Partnership's extensive membership. Peter's AI research interests are broad, including measuring progress in the field, figuring out how to translate ethical and safety concerns into mathematical constraints, finding the right metaphors and ways of thinking about AI development, and setting sound policies around high-stakes applications such as self-driving vehicles, recidivism prediction, cybersecurity, and military applications of AI. Prior to joining PAI, Peter was Chief Computer Scientist for the Electronic Frontier Foundation. At EFF he lead a team of technologists that launched numerous computer security and privacy projects including Let's Encrypt and Certbot, Panopticlick, HTTPS Everywhere, the SSL Observatory and Privacy Badger; they also worked on diverse Internet policy issues including campaigning to preserve open wireless networks; fighting to keep modern computing platforms open; helping to start the campaign against the SOPA/PIPA Internet blacklist legislation; and running the first controlled tests to confirm that Comcast was using forged reset packets to interfere with P2P protocols. Peter holds a PhD in computer science and law from the University of Melbourne; his research focused on the practicality and desirability of using alternative compensation systems to legalize P2P file sharing and similar distribution tools while still paying authors and artists for their work. He currently serves on the board of the Internet Security Research Group and the Advisory Council of the Open Technology Fund; he is an Affiliate of the Center for International Security and Cooperation at Stanford University and a Distinguished Technology Fellow at EFF.

Abstract 19: **Learning Global Variations in Outdoor PM_2.5 Concentrations with Satellite Images in AI For Social Good (AISG)**, 04:20 PM

The World Health Organization identifies outdoor fine particulate air pollution (PM2.5) as a leading risk factor for premature mortality globally. As such, understanding the global distribution of PM2.5 is an essential precursor towards implementing pollution mitigation strategies and modelling global public health. Here, we present a convolutional neural network based approach for estimating annual average outdoor PM2.5 concentrations using only satellite images. The resulting model achieves comparable performance to current state-of-the-art statistical models.

Speaker bio:
- Kris Y Hong is a research assistant and prospective PhD student in the Weichenthal Lab at McGill University, in Montreal, Canada. His interests lie in applying current statistical and machine learning techniques towards solving humanitarian and environmental challenges. Prior to joining McGill, he was a data analyst at the British Columbia Centre for Disease Control while receiving his B.Sc. in Statistics from the University of British Columbia.

Abstract 20: **Pareto Efficient Fairness for Skewed Subgroup Data in AI For Social Good (AISG)**, *Balashankar* 04:30 PM

As awareness of the potential for learned models to amplify existing societal biases increases, the field of ML fairness has developed mitigation techniques. A prevalent method applies constraints, including equality of performance, with respect to subgroups defined over the intersection of sensitive attributes such as race and gender. Enforcing such constraints when the subgroup populations are considerably skewed with respect to a target can lead to unintentional degradation in performance, without benefiting any individual subgroup, counter to the United Nations Sustainable Development goals of reducing inequalities and promoting growth. In order to avoid such performance degradation while ensuring equitable treatment to all groups, we propose Pareto-Efficient Fairness (PEF), which identifies the operating point on the Pareto curve of subgroup performances closest to the fairness hyperplane. Specifically, PEF finds a Pareto Optimal point which maximizes multiple subgroup accuracy measures. The algorithm

scalarizes using the adaptive weighted metric norm by iteratively searching the Pareto region of all models enforcing the fairness constraint. PEF is backed by strong theoretical results on discoverability and provides domain practitioners finer control in navigating both convex and non-convex accuracy-fairness trade-offs. Empirically, we show that PEF increases performance of all subgroups in skewed synthetic data and UCI datasets.

Speaker bio:
Ananth Balashnkar is a 2nd year Ph.D student in Computer Science advised by Prof. Lakshminarayanan Subramanian at NYU's Courant Institute of Mathematical Sciences. He is currently interested in Interpretable Machine Learning and the challenges involved in applying machine perception for the domains of policy, privacy, economics and healthcare.

Abstract 21: **Crisis Sub-Events on Social Media: A Case Study of Wildfires in AI For Social Good (AISG)**, 04:40 PM

Social media has been extensively used for crisis management. Recent work examines possible sub-events as a major crisis unfolds. In this project, we first propose a framework to identify sub-events from tweets. Then, leveraging 4 California wildfires in 2018-2019 as a case study, we investigate how sub-events cascade based on existing hypotheses drawn from the disaster management literature, and find that most hypotheses are supported on social media, e.g., fire induces smoke, which causes air pollution, which later harms health and eventually affects the healthcare system. In addition, we discuss other unexpected sub-events that emerge from social media.

Speaker bio:
Alejandro (Alex) Jaimes is Chief Scientist and SVP of AI at Dataminr. Alex has 15+ years of intl. experience in research and product impact at scale. He has published 100+ technical papers in top-tier conferences and journals in diverse topics in AI and has been featured widely in the press (MIT Tech review, CNBC, Vice, TechCrunch, Yahoo! Finance, etc.). He has given 80+ invited talks (AI for Good Global Summit (UN, Geneva), the Future of Technology Summit, O'Reilly (AI, Strata, Velocity), Deep Learning Summit, etc.). Alex is also an Endeavor Network mentor (which leads the high-impact entrepreneurship movement around the world), and was an early voice in Human-Centered AI (Computing). He holds a Ph.D. from Columbia U.

Abstract 22: **Towards Detecting Dyslexia in Children's Handwriting Using Neural Networks in AI For Social Good (AISG)**, 04:50 PM

Dyslexia is a learning disability that hinders a person's ability to read. Dyslexia needs to be caught early, however, teachers are not trained to detect dyslexia and screening tests are used inconsistently. We propose (1) two new data sets of handwriting collected from children with and without dyslexia amounting to close to 500 handwriting samples, and (2) an automated early screening technique to be used in conjunction with current approaches, to accelerate the detection process. Preliminary results suggest our system out-performs teachers.

Speaker bio:
Katie Spoon recently completed her B.S./M.S. in computer science from Indiana University with minors in math and statistics, and with research interests in anomaly detection, computer vision, data visualization, and applications of computer vision to health and education, like her senior thesis detecting dyslexia with neural networks. She worked at IBM Research in the summer of 2018 on neuromorphic computing, and will be returning there full-time. She hopes to potentially get a PhD and become a corporate research scientist.

Abstract 23: **Assisting Vulnerable Communities through AI and OR: from Data to Deployed Decisions in AI For Social Good (AISG)**, 05:00 PM

N/A

Speaker bio:
Phebe Vayanos is Assistant Professor of Industrial & Systems Engineering and Computer Science at the University of Southern California, and Associate Director of the CAIS Center for Artificial

Intelligence in Society. Her research aims to address fundamental questions in data-driven optimization (aka prescriptive analytics) with aim to tackle real-world decision- and policy-making problems in uncertain and adversarial environments.

## Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes

***Battista Biggio, Pavel Korshunov, Thomas Mensink, Giorgio Patrini, Arka Sadhu, Delip Rao***

**104 C, Sat Jun 15, 08:30 AM**

With the latest advances of deep generative models, synthesis of images and videos as well as of human voices have achieved impressive realism. In many domains, synthetic media are already difficult to distinguish from real by the human eye and ear. The potential of misuses of these technologies is seldom discussed in academic papers; instead, vocal concerns are rising from media and security organizations as well as from governments. Researchers are starting to experiment on new ways to integrate deep learning with traditional media forensics and security techniques as part of a technological solution. This workshop will bring together experts from the communities of machine learning, computer security and forensics in an attempt to highlight recent work and discuss future effort to address these challenges. Our agenda will alternate contributed papers with invited speakers. The latter will emphasize connections among the interested scientific communities and the standpoint of institutions and media organizations.

### Schedule

09:00 AM **Welcome Remarks** *Patrini*

09:10 AM **Invited Talk by Professor Alexei Efros (UC Berkeley)** *Efros*

09:40 AM **Invited Talk by Dr. Matt Turek (DARPA)** *Turek*

10:10 AM **Contributed Talk: Limits of Deepfake Detection: A Robust Estimation Viewpoint**

10:30 AM **Poster session 1 and Coffee break**

11:30 AM **Invited Talk by Professor Pawel Korus (NYU) Neural Imaging Pipelines - the Scourge or Hope of Forensics?**

12:00 PM **Contributed Talk: We Need No Pixels: Video Manipulation Detection Using Stream Descriptors** *Güera*

12:15 PM **Contributed Talk: A Utility-Preserving GAN for Face Obscuration** *Hao*

12:30 PM **Lunch Break**

02:00 PM **Invited Talk by Professor Luisa Verdoliva (University Federico II Naples)** *Verdoliva*

02:30 PM **Contributed Talk: Tampered Speaker Inconsistency Detection with Phonetically Aware Audio-visual Features** *Korshunov*

02:45 PM **Contributed Talk: Measuring the Effectiveness of Voice Conversion on Speaker Identification and Automatic Speech Recognition Systems** *Keskin*

03:00 PM **Poster session 2 and Coffee break**

03:45 PM **CVPR19 Media Forensics workshop: a Preview**

04:15 PM **Invited Talk by Tom Van de Weghe (Stanford & VRT)**

04:45 PM **Invited Talk by Aviv Ovadya (Thoughtful Technology Project)**

05:00 PM **Panel Discussion moderated by Delip Rao**

Abstracts (5):

Abstract 4: **Contributed Talk: Limits of Deepfake Detection: A Robust Estimation Viewpoint in Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes**, 10:10 AM

Deepfake detection is formulated as a hypothesis testing problem to classify an image as genuine or GAN-generated. A robust statistics view of

GANs is considered to bound the error probability for various GAN implementations in terms of their performance. The bounds are further simplified using a Euclidean approximation for the low error regime. Lastly, relationships between error probability and epidemic thresholds for spreading processes in networks are established.

Abstract 7: **Contributed Talk: We Need No Pixels: Video Manipulation Detection Using Stream Descriptors in Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes**, *Güera* 12:00 PM

Manipulating video content is easier than ever. Due to the misuse potential of manipulated content, multiple detection techniques that analyze the pixel data from the videos have been proposed. However, clever manipulators should also carefully forge the metadata and auxiliary header information, which is harder to do for videos than images. In this paper, we propose to identify forged videos by analyzing their multimedia stream descriptors with simple binary classifiers, completely avoiding the pixel space. Using well-known datasets, our results show that this scalable approach can achieve a high manipulation detection score if the manipulators have not done a careful data sanitization of the multimedia stream descriptors.

Abstract 8: **Contributed Talk: A Utility-Preserving GAN for Face Obscuration in Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes**, *Hao* 12:15 PM

From TV news to Google StreetView, face obscuration has been used for privacy protection. Due to recent advances in the field of deep learning, ob-
scuration methods such as Gaussian blurring and pixelation are not guaranteed to conceal identity. In this paper, we propose a utility-preserving gen-

erative model, UP-GAN, that is able to provide an effective face obscuration, while preserving facial utility. By utility-preserving we mean preserving facial features that do not reveal identity,

such as age, gender, skin tone, pose, and expression. We show that the proposed method achieves
a better performance than the common obscuration methods in terms of obscuration and utility preservation.

Abstract 11: **Contributed Talk: Tampered Speaker Inconsistency Detection with Phonetically Aware Audio-visual Features in Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes**, *Korshunov* 02:30 PM

The recent increase in social media based propaganda, i.e., 'fake news', calls for automated methods to detect tampered content. In this paper,
we focus on detecting tampering in a video with a person speaking to a camera. This form of manipulation is easy to perform, since one can just replace a part of the audio, dramatically changing the meaning of the video. We consider several detection approaches based on phonetic fea-

tures and recurrent networks. We demonstrate that by replacing standard MFCC features with embeddings from a DNN trained for automatic speech recognition, combined with mouth landmarks (visual features), we can achieve a significant performance improvement on several challenging publicly available databases of speakers (VidTIMIT, AMI, and GRID), for which we generated sets of tampered data. The evaluations demonstrate a relative equal error rate reduction of 55% (to 4.5% from 10.0%) on the large GRID corpus based dataset and a satisfying generaliza-

tion of the model on other datasets.

Abstract 12: **Contributed Talk: Measuring the Effectiveness of Voice Conversion on Speaker Identification and Automatic Speech Recognition Systems in Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes**, *Keskin* 02:45 PM

This paper evaluates the effectiveness of a Cycle-

GAN based voice converter (VC) on four speaker identification (SID) systems and an automated speech recognition (ASR) system for various pur-

poses. Audio samples converted by the VC model

are classified by the SID systems as the intended target at up to 46% top-1 accuracy among more than 250 speakers. This encouraging result in imitating the target styles led us to investigate if converted (synthetic) samples can be used to improve ASR training. Unfortunately, adding synthetic data to the ASR training set only marginally

improves word and character error rates. Our results indicate that even though VC models can successfully mimic the style of target speakers as

measured by SID systems, improving ASR training with synthetic data from VC systems needs further research to establish its efficacy.

## ICML Workshop on Imitation, Intent, and Interaction (I3)

***Nicholas Rhinehart, Sergey Levine, Chelsea Finn, He He, Ilya Kostrikov, Justin Fu, Siddharth Reddy***

**201, Sat Jun 15, 08:30 AM**

**Website**: [https://sites.google.com/view/icml-i3](https://sites.google.com/view/icml-i3)

**Abstract:** A key challenge for deploying interactive machine learning systems in the real world is the ability for machines to understand human intent. Techniques such as imitation learning and inverse reinforcement learning are popular data-driven paradigms for modeling agent intentions and controlling agent behaviors, and have been applied to domains ranging from robotics and autonomous driving to dialogue systems. Such techniques provide a practical solution to specifying objectives to machine learning systems when they are difficult to program by hand.

While significant progress has been made in these areas, most research effort has concentrated on modeling and controlling single agents from dense demonstrations or feedback. However, the real world has multiple agents, and dense expert data collection can be prohibitively expensive. Surmounting these obstacles requires progress in frontiers such as:

1) the ability to infer intent from multiple modes of data, such as language or observation, in addition to traditional demonstrations.
2) the ability to model multiple agents and their intentions, both in cooperative and adversarial settings.
3) handling partial or incomplete information from the expert, such as demonstrations that lack dense action annotations, raw videos, etc..

The workshop on Imitation, Intention, and Interaction (I3) seeks contributions at the interface of these frontiers, and will bring together researchers from multiple disciplines such as robotics, imitation and reinforcement learning, cognitive science, AI safety, and natural language understanding. Our aim will be to reexamine the assumptions in standard imitation learning problem statements (e.g., inverse reinforcement learning) and connect distinct application disciplines, such as robotics and NLP, with researchers developing core imitation learning algorithms. In this way, we hope to arrive at new problem formulations, new research directions, and the development of new connections across distinct disciplines that interact with imitation learning methods.

### Schedule

| | |
|---|---|
| 08:45 AM | **Welcoming Remarks** |
| 09:00 AM | **Hal Daumé III** |
| 09:20 AM | **Joyce Chai** |
| 09:40 AM | **Stefano Ermon** |
| 10:00 AM | **Iris R. Seaman** |
| 10:20 AM | **Poster session and coffee** |
| 11:30 AM | **Changyou Chen** |
| 11:50 AM | **Faraz Torabi** |
| 12:10 PM | **Seyed Kamyar Seyed Ghasemipour** |
| 12:30 PM | **Lunch break** |

02:05 PM **Natasha Jaques**
02:25 PM **Pierre Sermanet**
02:45 PM **Nicholas R Waytowich**
03:05 PM **Poster session and coffee**
04:30 PM **Kris Kitani**
04:50 PM **Abhishek Das**

Abstracts (7):

Abstract 2: **Hal Daumé III in ICML Workshop on Imitation, Intent, and Interaction (I3)**, 09:00 AM

Title: Beyond demonstrations: Learning behavior from higher-level supervision

Abstract 3: **Joyce Chai in ICML Workshop on Imitation, Intent, and Interaction (I3)**, 09:20 AM

Title: Collaboration in Situated Language Communication

Abstract 4: **Stefano Ermon in ICML Workshop on Imitation, Intent, and Interaction (I3)**, 09:40 AM

Multi-agent Imitation and Inverse Reinforcement Learning

Abstract 5: **Iris R. Seaman in ICML Workshop on Imitation, Intent, and Interaction (I3)**, 10:00 AM

Title: Nested Reasoning About Autonomous Agents Using Probabilistic Programs

Abstract 11: **Natasha Jaques in ICML Workshop on Imitation, Intent, and Interaction (I3)**, 02:05 PM

Title: Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning

We propose a unified mechanism for achieving coordination and communication in Multi-Agent Reinforcement Learning (MARL), through rewarding agents for having causal influence over other agents' actions. Causal influence is assessed using counterfactual reasoning. At each timestep, an agent simulates alternate actions that it could have taken, and computes their effect on the behavior of other agents. Actions that lead to bigger changes in other agents' behavior are considered influential and are rewarded. We show that this is equivalent to rewarding agents for having high mutual information between their actions. Empirical results demonstrate that influence leads to enhanced coordination and communication in challenging social dilemma environments, dramatically increasing the learning curves of the deep RL agents, and leading to more meaningful learned communication protocols. The influence rewards for all agents can be computed in a decentralized way by enabling agents to learn a model of other agents using deep neural networks. In contrast, key previous works on emergent communication in the MARL setting were unable to learn diverse policies in a decentralized manner and had to resort to centralized training. Consequently, the influence reward opens up a window of new opportunities for research in this area."

Abstract 12: **Pierre Sermanet in ICML Workshop on Imitation, Intent, and Interaction (I3)**, 02:25 PM

Title: Self-Supervision and Play

Abstract: Real-world robotics is too complex to supervise with labels or through reward functions. While some amount of supervision is necessary, a more scalable approach instead is to bootstrap learning through self-supervision by first learning general task-agnostic representations. Specifically, we argue that we should learn from large amounts of unlabeled play data. Play serves as a way to explore and learn the breadth of what is possible in an undirected way. This strategy is widely used in nature to prepare oneself to achieve future tasks without knowing in advance

which ones. In this talk, we present methods for learning vision and control representations entirely from unlabeled sequences. We demonstrate these representations self-arrange semantically and functionally and can be used for downstream tasks, without ever using labels or rewards.

Abstract 15: **Kris Kitani in ICML Workshop on Imitation, Intent, and Interaction (I3)**, 04:30 PM

Title: Multi-modal trajectory forecasting

## Coding Theory For Large-scale Machine Learning

***Viveck Cadambe, Pulkit Grover, Dimitris Papailiopoulos, Gauri Joshi***

**202, Sat Jun 15, 08:30 AM**

# Coding Theory For Large-scale Machine Learning
Coding theory involves the art and science of how to add redundancy to data to ensure that a desirable output is obtained at despite deviations from ideal behavior from the system components that interact with the data. Through a rich, mathematically elegant set of techniques, coding theory has come to significantly influence the design of modern data communications, compression and storage systems. The last few years have seen a rapidly growing interest in coding theory based approaches for the development of efficient machine learning algorithms towards robust, large-scale, distributed computational pipelines.

The CodML workshop brings together researchers developing coding techniques for machine learning, as well as researchers working on systems implementations for computing, with cutting-edge presentations from both sides. The goal is to learn about non-idealities in system components as well as approaches to obtain reliable and robust learning despite these non-idealities, and identify problems of future interest.

The workshop is co-located with ICML 2019, and will be held in Long Beach, California, USA on June 14th or 15th, 2019.

Please see the [website](https://sites.google.com/view/codml2019) for more details:

## Call for Posters

### Scope of the Workshop

In this workshop we solicit research papers focused on the application of coding and information-theoretic techniques for distributed machine learning. More broadly, we seek papers that address the problem of making machine learning more scalable, efficient, and robust. Both theoretical as well as experimental contributions are welcome. We invite authors to submit papers on topics including but not limited to:

* Asynchronous Distributed Training Methods
* Communication-Efficient Training
* Model Compression and Quantization
* Gradient Coding, Compression and Quantization

* Erasure Coding Techniques for Straggler Mitigation
* Data Compression in Large-scale Machine Learning
* Erasure Coding Techniques for ML Hardware Acceleration
* Fast, Efficient and Scalable Inference
* Secure and Private Machine Learning
* Data Storage/Access for Machine Learning Jobs
* Performance evaluation of coding techniques

### Submission Format and Instructions

The authors should prepare extended abstracts in the ICML paper format and submit via [CMT](https://cmt3.research.microsoft.com/CODMLW2019/). Submitted papers may not exceed three (3) single-spaced double-column pages excluding references. All results, proofs, figures, tables must be included in the 3 pages. The submitted manuscripts should include author names and affiliations, and an abstract that does not exceed 250 words. The authors may include a link to an extended version of the paper that includes supplementary material (proofs, experimental details, etc.) but the reviewers are not required to read the extended version.

### Dual Submission Policy

Accepted submissions will be considered non-archival and can be submitted elsewhere without modification, as long as the other conference allows it. Moreover, submissions to CodML based on work recently accepted to other venues are also acceptable (though authors should explicitly make note of this in their submissions).

### Key Dates

**Paper Submission:** May 3rd, 2019, 11:59 PM anywhere on earth

**Decision Notification:** May 12th, 2019.

**Workshop date:** June 14 or 15, 2019

## Schedule

| | |
|---|---|
| 08:45 AM | **Opening Remarks** |
| 09:00 AM | **Salman Avestimehr: Lagrange Coded Computing: Optimal Design for Resilient, Secure, and Private Distributed Learning** *Avestimehr* |
| 09:30 AM | **Vivienne Sze: Exploiting redundancy for efficient processing of DNNs and beyond** *Sze* |
| 10:00 AM | **"Locality Driven Coded Computation" Michael Rudow, Rashmi Vinayak and Venkat Guruswami** *Rudow* |
| 10:10 AM | **"CodeNet: Training Large-Scale Neural Networks in Presence of Soft-Errors," Sanghamitra Dutta, Ziqian Bai, Tze Meng Low and Pulkit Grover** *Dutta* |
| 10:20 AM | **"Reliable Clustering with Redundant Data Assignment" Venkat Gandikota, Arya Mazumdar and Ankit Singh Rawat** *Gandikota* |
| 10:30 AM | **Poster Session I** *Draper, Aktas, Guler, Wang, Gandikota, Park, So, Tauz, Narra, Lin, Maddahali, Yang, Dutta, Reisizadeh, Wang, Balevi, Jain, McVay, Rudow, Soto, Li, Subramaniam, Demirhan, Gupta, Oktay, Barnes, Ballé, Haddadpour, Jeong, Chen, Fahim* |
| 11:30 AM | **Rashmi Vinayak: Resilient ML inference via coded computation: A learning-based approach** *Vinayak* |
| 12:00 PM | **Lunch Break** |
| 01:30 PM | **Markus Weimer: A case for coded computing on elastic compute** *Weimer* |
| 02:00 PM | **Alex Dimakis: Coding Theory for Distributed Learning** *Dimakis* |
| 02:30 PM | **Wei Zhang: Distributed deep learning system building at IBM: Scale-up and Scale-out case studies** *Zhang* |
| 03:30 PM | **"OverSketched Newton: Fast Convex Optimization for Serverless Systems," Vipul Gupta, Swanand Kadhe, Thomas Courtade, Michael Mahoney and Kannan Ramchandran** *Gupta* |
| 03:40 PM | **"Cooperative SGD: A Unified Framework for the Design and Analysis of Communication-Efficient SGD Algorithms", Jianyu Wang and Gauri Joshi** *Wang* |
| 03:50 PM | **"Secure Coded Multi-Party Computation for Massive Matrices with Adversarial Nodes," Seyed Reza, Mohammad Ali Maddah-Ali and Mohammad Reza Aref** *Maddahali* |
| 04:00 PM | **Poster Session II** |

## The How2 Challenge: New Tasks for Vision & Language

***Florian Metze, Lucia Specia, Desmond Elliot, Loic Barrault, Ramon Sanabria, Shruti Palaskar***

**203, Sat Jun 15, 08:30 AM**

Research at the intersection of vision and language has been attracting a lot of attention in

recent years. Topics include the study of multi-modal representations, translation between modalities, bootstrapping of labels from one modality into another, visually-grounded question answering, segmentation and storytelling, and grounding the meaning of language in visual data. An ever-increasing number of tasks and datasets are appearing around this recently-established field.

At NeurIPS 2018, we released the How2 data-set, containing more than 85,000 (2000h) videos, with audio, transcriptions, translations, and textual summaries. We believe it presents an ideal resource to bring together researchers working on the previously mentioned separate tasks around a single, large dataset. This rich dataset will facilitate the comparison of tools and algorithms, and hopefully foster the creation of additional annotations and tasks. We want to foster discussion about useful tasks, metrics, and labeling techniques, in order to develop a better understanding of the role and value of multi-modality in vision and language. We seek to create a venue to encourage collaboration between different sub-fields, and help establish new research directions and collaborations that we believe will sustain machine learning research for years to come.

[Workshop Homepage](https://srvk.github.io/how2-challenge/)

## Schedule

| | |
|---|---|
| 08:45 AM | **Welcome** |
| 09:00 AM | **The How2 Database and Challenge** *Specia, Sanabria* |
| 10:15 AM | **Coffee Break** |
| 10:30 AM | **Forcing Vision + Language Models To Actually See, Not Just Talk** *Parikh* |
| 11:00 AM | **Topics in Vision and Language: Grounding, Segmentation and Author Anonymity** |
| 11:30 AM | **Learning to Reason: Modular and Relational Representations for Visual Questions and Referring Expressions** *Saenko* |
| 01:30 PM | **Multi-agent communication from raw perceptual input: what works, what doesn't and what's next** *Lazaridou* |
| 02:00 PM | **Overcoming Bias in Captioning Models** *Hendricks* |
| 02:30 PM | **Embodied language grounding** *Fragkiadaki* |
| 03:00 PM | **Poster Session and Coffee** *Sanabria, Srinivasan, Raunak, Zhou, Kundu, Patel, Specia, Choe, Belova* |
| 04:30 PM | **Unsupervised Bilingual Lexicon Induction from mono-lingual multimodal data** *Jin* |
| 05:00 PM | **New Directions for Vision & Language** *Metze, Palaskar* |

Abstracts (2):

Abstract 7: **Multi-agent communication from raw perceptual input: what works, what doesn't and what's next in The How2 Challenge: New Tasks for Vision & Language**, *Lazaridou* 01:30 PM

Multi-agent communication has been traditionally used as a computational tool to study language evolution. Recently, it has attracted attention also as a means to achieve better coordination among multiple interacting agents in complex environments. However, is it easy to scale previous research in the new deep learning era? In this talk, I will first give a brief overview of some of the previous approaches that study emergent communication in cases where agents are given as input symbolic data. I will then move on to presenting some of the challenges that agents face when are placed in grounded environments where they receive raw perceptual information and how environmental or pre-linguistic conditions affect the nature of the communication protocols that they learn. Finally, I will discuss some potential remedies that are inspired from human language and communication.

Abstract 8: **Overcoming Bias in Captioning Models in The How2 Challenge: New Tasks for Vision & Language**, *Hendricks* 02:00 PM

Most machine learning models are known to capture and exploit bias. While this can be

beneficial for many classification tasks (e.g., it might be easier to recognize a computer mouse given the context of a computer and a desk), exploiting bias can also lead to incorrect predictions. In this talk, I will first consider how over-reliance on bias might lead to incorrect predictions in a scenario where is inappropriate to rely on bias: gender prediction in image captioning. I will present the Equalizer model which more accurately describes people and their gender by considering appropriate gender evidence. Next, I will consider how bias is related to hallucination, an interesting error mode in image captioning. I will present a metric designed to measure hallucination and consider questions like what causes hallucination, which models are prone to hallucination, and do current metrics accurately capture hallucination?

## Machine Learning for Music Discovery

***Erik Schmidt, Oriol Nieto, Fabien Gouyon, Katherine Kinnaird, Gert Lanckriet***

**204, Sat Jun 15, 08:30 AM**

The ever-increasing size and accessibility of vast music libraries has created a demand more than ever for artificial systems that are capable of understanding, organizing, or even generating such complex data. While this topic has received relatively marginal attention within the machine learning community, it has been an area of intense focus within the community of Music Information Retrieval (MIR). While significant progress has been made, these problems remain far from solved.

Furthermore, the recommender systems community has made great advances in terms of collaborative feedback recommenders, but these approaches suffer strongly from the cold-start problem. As such, recommendation techniques often fall back on content-based machine learning systems, but defining musical similarity is extremely challenging as myriad features all play some role (e.g., cultural, emotional, timbral, rhythmic). Thus, for machines must actually understand music to achieve an expert level of music recommendation.

On the other side of this problem sits the recent explosion of work in the area of machine creativity. Relevant examples are both Google Magenta and the startup Jukedeck, who seek to develop algorithms capable of composing and performing completely original (and compelling) works of music. These algorithms require a similar deep understanding of music and present challenging new problems for the machine learning and AI community at large.

This workshop proposal is timely in that it will bridge these separate pockets of otherwise very related research. And in addition to making progress on the challenges above, we hope to engage the wide AI and machine learning community with our nebulous problem space, and connect them with the many available datasets the MIR community has to offer (e.g., Audio Set, AcousticBrainz, Million Song Dataset), which offer near commercial scale to the academic research community.

## Schedule

| | |
|---|---|
| 09:00 AM | **From Listening to Watching, A Recommender Systems Perspective** *Raimond* |
| 10:00 AM | **Poster Presentations (Part 1)** *Agrawal, Choi, Gururani, Hamidi, Jhamtani, Kopparti, Krause, Lee, Pati, Zhdanov* |
| 11:00 AM | **Making Efficient use of Musical Annotations** *McFee* |
| 11:20 AM | **Two-level Explanations in Music Emotion Recognition** *Haunschmid* |
| 11:40 AM | **Characterizing Musical Correlates of Large-Scale Discovery Behavior** *Kaneshiro* |
| 12:00 PM | **NPR: Neural Personalised Ranking for Song Selection** *Levy* |
| 02:00 PM | **Personalization at Amazon Music** *Ellis* |
| 02:20 PM | **A Model-Driven Exploration of Accent Within the Amateur Singing Voice** *Noufi* |
| 02:40 PM | **What's Broken in Music Informatics Research? Three Uncomfortable Statements** *Salamon* |

03:00 PM **Poster Presentations (Part 2)** *Agrawal, Choi, Gururani, Hamidi, Jhamtani, Kopparti, Krause, Lee, Pati, Zhdanov*

04:30 PM **User-curated shaping of expressive performances** *Shi*

04:50 PM **Interactive Neural Audio Synthesis** *Hantrakul*

05:10 PM **Visualizing and Understanding Self-attention based Music Tagging** *Won*

05:30 PM **A CycleGAN for style transfer between drum & bass subgenres** *Vande Veire*

Abstracts (6):

Abstract 1: **From Listening to Watching, A Recommender Systems Perspective in Machine Learning for Music Discovery**, *Raimond* 09:00 AM

In this talk, I'll be discussing a few key differences between recommending music and recommending movies or TV shows, and how these differences can lead to vastly different designs, approaches, and algorithms to find the best possible recommendation for a user. On the other hand, I'll also discuss some common challenges and some of our recent research on these topics, such as better understanding the impact of a recommendation, enable better offline metrics, or optimizing for longer-term outcomes. Most importantly, I'll try to leave a lot of time for questions and discussions.

Abstract 3: **Making Efficient use of Musical Annotations in Machine Learning for Music Discovery**, *McFee* 11:00 AM

Many tasks in audio-based music analysis require building mappings between complex representational spaces, such as the input audio signal (or spectral representation), and structured, time-varying output such as pitch, harmony, instrumentation, rhythm, or structure. These mappings encode musical domain knowledge, and involve processing and integrating knowledge at multiple scales simultaneously. It typically takes humans years of training and practice to master these concepts, and as a result, data collection for sophisticated musical analysis tasks is often costly and time-consuming. With limited available data with reliable annotations, it can be difficult to build robust models to automate music annotation by computational means. However, musical problems often exhibit a great deal of structure, either in the input or output representations, or even between related tasks, which can be effectively leveraged to reduce data requirements. In this talk, I will survey several recent manifestations of this phenomenon across different music and audio analysis problems, drawing on recent work from the NYU Music and Audio Research Lab.

Abstract 5: **Characterizing Musical Correlates of Large-Scale Discovery Behavior in Machine Learning for Music Discovery**, *Kaneshiro* 11:40 AM

We seek to identify musical correlates of real-world discovery behavior by analyzing users' audio identification queries from the Shazam service. Recent research has shown that such queries are not uniformly distributed over the course of a song, but rather form clusters that may implicate musically salient events. Using a publicly available dataset of Shazam queries, we extend this research and examine candidate musical features driving increases in query likelihood. Our findings suggest a relationship between musical novelty -- including but not limited to structural segmentation boundaries -- and ensuing peaks in discovery-based musical engagement.

Abstract 7: **Personalization at Amazon Music in Machine Learning for Music Discovery**, *Ellis* 02:00 PM

I'll give an overview of some of the projects that we are working on to make Amazon Music more personalized for our customers. Projects include personalized speech and language understanding for voice search, personalizing "Play Music" requests on Alexa, and work with traditional recommender models as building blocks for many customer experiences.

Abstract 9: **What's Broken in Music Informatics Research? Three Uncomfortable Statements in Machine Learning for Music Discovery**, *Salamon* 02:40 PM

Melody extraction has been an active topic of research in Music Information Retrieval for decades now. And yet - what *is* a melody? As a community we still (mostly) shy away from this question, resorting to definition-by-annotation. How well do past/present/future algorithms perform? Despite known limitations with existing datasets and metrics, we still (mostly) stick to the same ones. And last but not least, why do melody extraction at all? Despite great promise (e.g. query-by-humming, large-scale musicological analyses, etc.), melody extraction has seen limited application outside of MIR research. In this talk I will present three problems that are common to several of research area in music informatics: the challenge of trying to model ambiguous musical concepts by training models with somewhat arbitrary reference annotations, the lack of model generalization in the face of small, low-variance training sets, and the possible disconnect between parts of the music informatics research community and the potential users of the technologies it produces.

Abstract 11: **User-curated shaping of expressive performances in Machine Learning for Music Discovery**, *Shi* 04:30 PM

Musical statements can be interpreted in performance with a wide variety of stylistic and expressive inflections. We explore how different musical characters are performed based on an extension of the basis function models, a data-driven framework for expressive performance. In this framework, expressive dimensions such as tempo, dynamics and articulation are modeled as a function of score features, i.e. numerical encodings of specific aspects of a musical score, using neural networks. By allowing the user to weight the contribution of the input score features, we show that predictions of expressive dimensions can be used to express different musical characters

## Workshop on Self-Supervised Learning

***Aaron van den Oord, Yusuf Aytar, Carl Doersch, Carl Vondrick, Alec Radford, Pierre Sermanet, Amir Zamir, Pieter Abbeel***

**Grand Ballroom A, Sat Jun 15, 08:30 AM**

Self-supervised learning is a promising alternative where proxy tasks are developed that allow models and agents to learn without explicit supervision in a way that helps with downstream performance on tasks of interest. One of the major benefits of self-supervised learning is increasing data efficiency: achieving comparable or better performance with less labeled data or fewer environment steps (in Reinforcement learning / Robotics).
The field of self-supervised learning (SSL) is rapidly evolving, and the performance of these methods is creeping closer to the fully supervised approaches. However, many of these methods are still developed in domain-specific sub-communities, such as Vision, RL and NLP, even though many similarities exist between them.
While SSL is an emerging topic and there is great interest in these techniques, there are currently few workshops, tutorials or other scientific events dedicated to this topic.
This workshop aims to bring together experts with different backgrounds and applications areas to share inter-domain ideas and increase cross-pollination, tackle current shortcomings and explore new directions. The focus will be on the machine learning point of view rather than the domain side.

https://sites.google.com/corp/view/self-supervised-icml2019

## Schedule

08:50 AM **Opening remarks**
09:00 AM **Jacob Devlin**
09:30 AM **Alison Gopnik**
10:00 AM **Learning Latent Plans from Play**

| | |
|---|---|
| 10:15 AM | **Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty** |
| 10:30 AM | **Poster Session + Coffee Break** |
| 11:30 AM | **Chelsea Finn** |
| 12:00 PM | **Lunch** |
| 02:00 PM | **Yann Lecun** |
| 02:30 PM | **Revisiting Self-Supervised Visual Representation Learning** |
| 02:45 PM | **Data-Efficient Image Recognition with Contrastive Predictive Coding** |
| 03:00 PM | **Poster session + Coffee Break** |
| 04:00 PM | **Andrew Zisserman** |
| 04:30 PM | **Abhinav Gupta** |
| 05:00 PM | **Alexei Efros** |

## Learning and Reasoning with Graph-Structured Representations

***Ethan Fetaya, Zhiting Hu Hu, Thomas Kipf, Yujia Li, Xiaodan Liang, Renjie Liao, Raquel Urtasun, Hao Wang, Max Welling, Eric Xing, Richard Zemel***

**Grand Ballroom B, Sat Jun 15, 08:30 AM**

Graph-structured representations are widely used as a natural and powerful way to encode information such as relations between objects or entities, interactions between online users (e.g., in social networks), 3D meshes in computer graphics, multi-agent environments, as well as molecular structures, to name a few. Learning and reasoning with graph-structured representations is gaining increasing interest in both academia and industry, due to its fundamental advantages over more traditional unstructured methods in supporting interpretability, causality, transferability, etc. Recently, there is a surge of new techniques in the context of deep learning, such as graph neural networks, for learning graph representations and performing reasoning and prediction, which have achieved impressive progress. However, it can still be a long way to go to obtain satisfactory results in long-range multi-step reasoning, scalable learning with very large graphs, flexible modeling of graphs in combination with other dimensions such as temporal variation and other modalities such as language and vision. New advances in theoretical foundations, models and algorithms, as well as empirical discoveries and applications are therefore all highly desirable.

The aims of this workshop are to bring together researchers to dive deeply into some of the most promising methods which are under active exploration today, discuss how we can design new and better benchmarks, identify impactful application domains, encourage discussion and foster collaboration. The workshop will feature speakers, panelists, and poster presenters from machine perception, natural language processing, multi-agent behavior and communication, meta-learning, planning, and reinforcement learning, covering approaches which include (but are not limited to):

-Deep learning methods on graphs/manifolds/ relational data (e.g., graph neural networks)
-Deep generative models of graphs (e.g., for drug design)
-Unsupervised graph/manifold/relational embedding methods (e.g., hyperbolic embeddings)
-Optimization methods for graphs/manifolds/ relational data
-Relational or object-level reasoning in machine perception
-Relational/structured inductive biases for reinforcement learning, modeling multi-agent behavior and communication
-Neural-symbolic integration
-Theoretical analysis of capacity/generalization of deep learning models for graphs/manifolds/ relational data
-Benchmark datasets and evaluation metrics

### Schedule

| | |
|---|---|
| 08:45 AM | **Opening remarks** |
| 09:00 AM | **William L. Hamilton, McGill University** *Hamilton* |

| | |
|---|---|
| 09:30 AM | **Evolutionary Representation Learning for Dynamic Graphs; Aynaz Taheri and Tanya Berger-Wolf** *taheri* |
| 09:45 AM | **Poster spotlights #1** *Chen, Hadziosmanovic, Ramírez Rivera* |
| 10:00 AM | **Morning poster session and coffee break** |
| 11:00 AM | **Marwin Segler, Benevolent AI** *Segler* |
| 11:30 AM | **Yaron Lipman, Weizmann Institute of Science** *Lipman* |
| 12:00 PM | **PAN: Path Integral Based Convolution for Deep Graph Neural Networks; Zheng Ma, Ming Li and Yu Guang Wang** *Ma* |
| 12:15 PM | **Poster spotlights #2** *Teixeira, Baldassarre* |
| 12:30 PM | **Lunch break** |
| 02:00 PM | **Alex Polozov, Microsoft Research** *Polozov* |
| 02:30 PM | **Sanja Fidler, University of Toronto** *Fidler* |
| 03:00 PM | **On Graph Classification Networks, Datasets and Baselines; Enxhell Luzhnica, Ben Day and Pietro Lió** |
| 03:15 PM | **Poster spotlights #3** *Kumar, De Cao, Chen* |
| 03:30 PM | **Afternoon poster session and coffee break** *Tu, Zhang, Chen* |
| 04:30 PM | **Caroline Uhler, MIT** *Uhler* |
| 05:00 PM | **Alexander Schwing, University of Illinois at Urbana-Champaign** *Schwing* |

## Exploration in Reinforcement Learning Workshop

***Surya Bhupatiraju, Benjamin Eysenbach, Shixiang Gu, Harrison Edwards, Martha White, Pierre-Yves Oudeyer, Ken Stanley, Emma Brunskill***

**Hall A, Sat Jun 15, 08:30 AM**

Exploration is a key component of reinforcement learning (RL). While RL has begun to solve relatively simple tasks, current algorithms cannot complete complex tasks. Our existing algorithms often endlessly dither, failing to meaningfully explore their environments in search of high-reward states. If we hope to have agents autonomously learn increasingly complex tasks, these machines must be equipped with machinery for efficient exploration.

The goal of this workshop is to present and discuss exploration in RL, including deep RL, evolutionary algorithms, real-world applications, and developmental robotics. Invited speakers will share their perspectives on efficient exploration, and researchers will share recent work in spotlight presentations and poster sessions.

### Schedule

| | |
|---|---|
| 09:00 AM | **Doina Precup** |
| 09:30 AM | **Spotlight Talks** |
| 10:00 AM | **Poster Session #1** *Ali Taiga, Deshmukh, Rashid, Binas, Yasui, Pong, Imagawa, Clifton, Mysore, Tsai, Chuck, Vezzani, Eriksson* |
| 11:00 AM | **Emo Todorov** |
| 11:30 AM | **Best Paper Talks** |
| 12:00 PM | **Pieter Abbeel** |
| 12:30 PM | **Lunch** |
| 02:00 PM | **Raia Hadsell** |
| 02:30 PM | **Lightning Talks** |
| 03:00 PM | **Poster Session #2** |
| 04:00 PM | **Martha White - Adapting Behaviour via Intrinsic Rewards to Learn Predictions** |
| 04:30 PM | **Panel Discussion** |

Abstracts (3):

Abstract 3: **Poster Session #1 in Exploration in Reinforcement Learning Workshop**, *Ali Taiga, Deshmukh, Rashid, Binas, Yasui, Pong, Imagawa, Clifton, Mysore, Tsai, Chuck, Vezzani, Eriksson* 10:00 AM

This is the first poster session and coffee break. All the papers will be presented at both poster sessions.

Abstract 10: **Poster Session #2 in Exploration in Reinforcement Learning Workshop**, 03:00 PM

This is the second poster session and coffee break. All the papers will be presented at both poster sessions.

Abstract 12: **Panel Discussion in Exploration in Reinforcement Learning Workshop**, 04:30 PM

We will have a panel on exploration with panelists Martha White, Jeff Clune, Pulkit Agrawal, and Pieter Abbeel, and moderated by Doina Precup.

## Identifying and Understanding Deep Learning Phenomena

***Hanie Sedghi, Samy Bengio, Kenji Hata, Aleksander Madry, Ari Morcos, Behnam Neyshabur, Maithra Raghu, Ali Rahimi, Ludwig Schmidt, Ying Xiao***

**Hall B, Sat Jun 15, 08:30 AM**

Our understanding of modern neural networks lags behind their practical successes. As this understanding gap grows, it poses a serious challenge to the future pace of progress because fewer pillars of knowledge will be available to designers of models and algorithms. This workshop aims to close this understanding gap in deep learning. It solicits contributions that view the behavior of deep nets as a natural phenomenon to investigate with methods inspired from the natural sciences, like physics, astronomy, and biology. We solicit empirical work that isolates phenomena in deep nets, describes them quantitatively, and then replicates or falsifies them.

As a starting point for this effort, we focus on the interplay between data, network architecture, and training algorithms. We are looking for contributions that identify precise, reproducible phenomena, as well as systematic studies and evaluations of current beliefs such as “sharp local minima do not generalize well” or “SGD navigates out of local minima”. Through the workshop, we hope to catalogue quantifiable versions of such statements, as well as demonstrate whether or not they occur reproducibly.

## Schedule

| | |
|---|---|
| 08:45 AM | **Opening Remarks** |
| 09:00 AM | **Nati Srebro: Optimization’s Untold Gift to Learning: Implicit Regularization** *Srebro* |
| 09:30 AM | **Bad Global Minima Exist and SGD Can Reach Them** |
| 09:45 AM | **Deconstructing Lottery Tickets: Zeros, Signs, and the Supermask** |
| 10:00 AM | **Chiyuan Zhang: Are all layers created equal? -- Studies on how neural networks represent functions** |
| 10:30 AM | **Break and Posters** |
| 11:00 AM | **Line attractor dynamics in recurrent networks for sentiment classification** |
| 11:15 AM | **Do deep neural networks learn shallow learnable examples first?** |
| 11:30 AM | **Crowdsourcing Deep Learning Phenomena** |
| 12:00 PM | **Lunch and Posters** |
| 01:30 PM | **Aude Oliva: Reverse engineering neuroscience and cognitive science principles** |
| 02:00 PM | **On Understanding the Hardness of Samples in Neural Networks** |
| 02:15 PM | **On the Convex Behavior of Deep Neural Networks in Relation to the Layers' Width** |
| 02:30 PM | **Andrew Saxe: Intriguing phenomena in training and generalization dynamics of deep networks** *Saxe* |
| 03:00 PM | **Break and Posters** |
| 04:00 PM | **Olga Russakovsky** |

04:30 PM **Panel Discussion: Kevin Murphy, Nati Srebro, Aude Oliva, Andrew Saxe, Olga Russakovsky Moderator: Ali Rahimi**

Abstracts (4):

Abstract 1: **Opening Remarks in Identifying and Understanding Deep Learning Phenomena**, 08:45 AM

Hanie Sedghi

Abstract 14: **Andrew Saxe: Intriguing phenomena in training and generalization dynamics of deep networks in Identifying and Understanding Deep Learning Phenomena**, *Saxe* 02:30 PM

In this talk I will describe several phenomena related to learning dynamics in deep networks. Among these are (a) large transient training error spikes during full batch gradient descent, with implications for the training error surface; (b) surprisingly strong generalization performance of large networks with modest label noise even with infinite training time; (c) a training speed/test accuracy trade off in vanilla deep networks; (d) the inability of deep networks to learn known efficient representations of certain functions; and finally (e) a trade off between training speed and multitasking ability.

Abstract 16: **Olga Russakovsky in Identifying and Understanding Deep Learning Phenomena**, 04:00 PM

Olga Russakovsky

Abstract 17: **Panel Discussion: Kevin Murphy, Nati Srebro, Aude Oliva, Andrew Saxe, Olga Russakovsky Moderator: Ali Rahimi in Identifying and Understanding Deep Learning Phenomena**, 04:30 PM

Panelists:
Kevin Murphy, Nati Srebro, Aude Oliva, Andrew Saxe, Olga Russakovsky

Moderator:
Ali Rahimi

## Adaptive and Multitask Learning: Algorithms & Systems

***Maruan Al-Shedivat, Anthony Platanios, Otilia Stretcu, Jacob Andreas, Ameet Talwalkar, Rich Caruana, Tom Mitchell, Eric Xing***

**Seaside Ballroom, Sat Jun 15, 08:30 AM**

Driven by progress in deep learning, the machine learning community is now able to tackle increasingly more complex problems—ranging from multi-modal reasoning to dexterous robotic manipulation—all of which typically involve solving nontrivial combinations of tasks. Thus, designing adaptive models and algorithms that can efficiently learn, master, and combine multiple tasks is the next frontier. AMTL workshop aims to bring together machine learning researchers from areas ranging from theory to applications and systems, to explore and discuss:

* advantages, disadvantages, and applicability of different approaches to learning in multitask settings,
* formal or intuitive connections between methods developed for different problems that help better understand the landscape of multitask learning techniques and inspire technique transfer between research lines,
* fundamental challenges and open questions that the community needs to tackle for the field to move forward.

Webpage: [www.amtl-workshop.org](https://www.amtl-workshop.org/)

## Schedule

08:30 AM **Opening Remarks**

08:40 AM **Building and Structuring Training Sets for Multi-Task Learning (Alex Ratner)** *Ratner*

09:10 AM **Meta-Learning: Challenges and Frontiers (Chelsea Finn)** *Finn*

09:40 AM **Contributed Talk: Learning Exploration Policies for Model-Agnostic Meta-Reinforcement Learning**

09:55 AM **Contributed Talk: Lifelong Learning via Online Leverage Score Sampling**

10:10 AM **Tricks of the Trade 1 (Rich Caruana)** *Caruana*

10:25 AM **Coffee Break**

11:00 AM **Poster Session** *Balazevic, Kwon, Lengerich, Asiaee, Lambert, Chen, Ding, Florensa, Gaudio, Jaafra, Fang, Wang, Li, GURUMURTHY, Yan, Cilingir, Thangarasa, Li, Lowe*

12:00 PM **Lunch Break**

01:45 PM **ARUBA: Efficient and Adaptive Meta-Learning with Provable Guarantees (Ameet Talwalkar)** *Talwalkar*

02:15 PM **Efficient Lifelong Learning Algorithms: Regret Bounds and Statistical Guarantees (Massimiliano Pontil)**

02:45 PM **Tricks of Trade 2 (Rich Caruana)** *Caruana*

03:00 PM **Coffee Break**

03:30 PM **Multi-Task Learning in the Wilderness (Andrej Karpathy)** *Karpathy*

04:00 PM **Recent Trends in Personalization: A Netflix Perspective (Justin Basilico)** *Basilico*

04:30 PM **Contributed Talk: Improving Relevance Prediction with Transfer Learning in Large-scale Retrieval Systems** *Wang*

04:45 PM **Contributed Talk: Continual Adaptation for Efficient Machine Communication** *Kwon*

05:00 PM **Toward Robust AI Systems for Understanding and Reasoning Over Multimodal Data (Hannaneh Hajishirzi)**

05:30 PM **Closing Remarks**

Abstracts (5):

Abstract 4: **Contributed Talk: Learning Exploration Policies for Model-Agnostic Meta-Reinforcement Learning in Adaptive and Multitask Learning: Algorithms & Systems**, 09:40 AM

Meta-Reinforcement learning approaches aim to develop learning procedures that can adapt quickly to a distribution of tasks with the help of a few examples. Developing efficient exploration strategies capable of finding the most useful samples becomes critical in such settings. Existing approaches to finding efficient exploration strategies add auxiliary objectives to promote exploration by the pre-update policy, however, this makes the adaptation using a few gradient steps difficult as the pre-update (exploration) and post-update (exploitation) policies are quite different. Instead, we propose to explicitly model a separate exploration policy for the task distribution. Having two different policies gives more flexibility in training the exploration policy and also makes adaptation to any specific task easier. We show that using self-supervised or supervised learning objectives for adaptation stabilizes the training process and also demonstrate the superior performance of our model compared to prior works in this domain.

Abstract 5: **Contributed Talk: Lifelong Learning via Online Leverage Score Sampling in Adaptive and Multitask Learning: Algorithms & Systems**, 09:55 AM

In order to mimic the human ability of continual acquisition and transfer of knowledge across various tasks, a learning system needs the capability for life-long learning, effectively utilizing the previously acquired skills. As such, the key challenge is to transfer and generalize the knowledge learned from one task to other tasks, avoiding interference from previous knowledge and improving the overall performance. In this paper, within the continual learning paradigm, we introduce a method that effectively forgets the less useful data samples continuously across different tasks. The method uses statistical leverage score information to measure the importance of the data samples in every task and adopts frequent directions approach to enable a life-long learning property. This effectively maintains a constant training size

across all tasks. We first provide some mathematical intuition for the method and then demonstrate its effectiveness with experiments on variants of MNIST and CIFAR100 datasets.

Abstract 8: **Poster Session in Adaptive and Multitask Learning: Algorithms & Systems**, *Balazevic, Kwon, Lengerich, Asiaee, Lambert, Chen, Ding, Florensa, Gaudio, Jaafra, Fang, Wang, Li, GURUMURTHY, Yan, Cilingir, Thangarasa, Li, Lowe* 11:00 AM

Accepted papers: https://www.amtl-workshop.org/accepted-papers

---

TuckER: Tensor Factorization for Knowledge Graph Completion
Authors: Ivana Balazevic, Carl Allen, Timothy Hospedales

Learning Cancer Outcomes from Heterogeneous Genomic Data Sources: An Adversarial Multi-task Learning Approach
Authors: Safoora Yousefi, Amirreza Shaban, Mohamed Amgad, Lee Cooper

Continual adaptation for efficient machine communication
Authors: Robert Hawkins, Minae Kwon, Dorsa Sadigh, Noah Goodman

Every Sample a Task: Pushing the Limits of Heterogeneous Models with Personalized Regression
Authors: Ben Lengerich, Bryon Aragam, Eric Xing

Data Enrichment: Multi-task Learning in High Dimension with Theoretical Guarantees
Authors: Amir Asiaee, Samet Oymak, Kevin R. Coombes, Arindam Banerjee

A Functional Extension of Multi-Output Learning
Authors: Alex Lambert, Romain Brault, Zoltan Szabo, Florence d'Alche-Buc

Interpretable Robust Recommender Systems with Side Information
Authors: Wenyu Chen, Zhechao Huang, Jason Cheuk Nam Liang, Zihao Xu

Personalized Student Stress Prediction with Deep Multi-Task Network
Authors: Abhinav Shaw, Natcha Simsiri, Iman Dezbani, Madelina Fiterau, Tauhidur Rahaman

SuperTML: Domain Transfer from Computer Vision to Structured Tabular Data through Two-Dimensional Word Embedding
Authors: Baohua Sun, Lin Yang, Wenhan Zhang, Michael Lin, Patrick Dong, Charles Young, Jason Dong

Goal-conditioned Imitation Learning
Authors: Yiming Ding, Carlos Florensa, Mariano Phielipp, Pieter Abbeel

Tasks Without Borders: A New Approach to Online Multi-Task Learning
Authors: Alexander Zimin, Christoph H. Lampert

The Role of Embedding-complexity in Domain-invariant Representations
Authors: Ching-Yao Chuang, Antonio Torralba, Stefanie Jegelka

Lifelong Learning via Online Leverage Score Sampling
Authors: Dan Teng, Sakyasingha Dasgupta

Connections Between Optimization in Machine Learning and Adaptive Control
Authors: Joseph E. Gaudio, Travis E. Gibson, Anuradha M. Annaswamy, Michael A. Bolender, Eugene Lavretsky

Meta-Reinforcement Learning for Adaptive Autonomous Driving
Authors: Yesmina Jaafra, Jean Luc Laurent, Aline Deruyver, Mohamed Saber Naceur

PAGANDA: An Adaptive Task-Independent Automatic Data Augmentation
Authors: Boli Fang, Miao Jiang, Jerry Shen

Improving Relevance Prediction with Transfer Learning in Large-scale Retrieval Systems
Authors: Ruoxi Wang, Zhe Zhao, Xinyang Yi, Ji Yang, Derek Zhiyuan Cheng, Lichan Hong, Steve Tjoa, Jieqi Kang, Evan Ettinger, Ed Chi

Federated Optimization for Heterogeneous Networks
Authors: Tian Li*, Anit Kumar Sahu*, Manzil

Zaheer, Maziar Sanjabi, Ameet Talwalkar, Virginia Smith

Learning Exploration Policies for Model-Agnostic Meta-Reinforcement Learning
Authors: Swaminathan Gurumurthy, Sumit Kumar, Katia Sycara

A Meta Understanding of Meta-Learning
Authors: Wei-Lun Chao, Han-Jia Ye, De-Chuan Zhan, Mark Campbell, Kilian Q. Weinberger

Multi-Task Learning via Task Multi-Clustering
Authors: Andy Yan, Xin Wang, Ion Stoica, Joseph Gonzalez, Roy Fox

Prototypical Bregman Networks
Authors: Kubra Cilingir, Brian Kulis

Differentiable Hebbian Plasticity for Continual Learning
Authors: Vithursan Thangarasa, Thomas Miconi, Graham W. Taylor

Active Multitask Learning with Committees
Authors: Jingxi Xu, Da Tang, Tony Jebara

Progressive Memory Banks for Incremental Domain Adaptation
Authors: Nabiha Asghar, Lili Mou, Kira A. Selby, Kevin D. Pantasdo, Pascal Poupart, Xin Jiang

Sub-policy Adaptation for Hierarchical Reinforcement Learning
Authors: Alexander Li, Carlos Florensa, Pieter Abbeel

Learning to learn to communicate
Authors: Ryan Lowe*, Abhinav Gupta*, Jakob Foerster, Douwe Kiela, Joelle Pineau

Abstract 16: **Contributed Talk: Improving Relevance Prediction with Transfer Learning in Large-scale Retrieval Systems in Adaptive and Multitask Learning: Algorithms & Systems**, *Wang* 04:30 PM

Machine learned large-scale retrieval systems require a large amount of training data representing query-item relevance. However, collecting users' explicit feedback is costly. In this paper, we propose to leverage user logs and implicit feedback as auxiliary objectives to improve relevance modeling in retrieval systems. Specifically, we adopt a two-tower neural net architecture to model query-item relevance given both collaborative and content information. By introducing auxiliary tasks trained with much richer implicit user feedback data, we improve the quality and resolution for the learned representations of queries and items. Applying these learned representations to an industrial retrieval system has delivered significant improvements.

Abstract 17: **Contributed Talk: Continual Adaptation for Efficient Machine Communication in Adaptive and Multitask Learning: Algorithms & Systems**, *Kwon* 04:45 PM

To communicate with new partners in new contexts, humans rapidly form new linguistic conventions. Recent language models trained with deep neural networks are able to comprehend and produce the existing conventions present in their training data, but are not able to flexibly and interactively adapt those conventions on the fly as humans do. We introduce a repeated reference task as a benchmark for models of adaptation in communication and propose a regularized continual learning framework that allows an artificial agent initialized with a generic language model to more accurately and efficiently understand their partner over time. We evaluate this framework through simulations on COCO and in real-time reference game experiments with human partners.