

Yair Schiff

yairschiff@cs.cornell.edu

yair-schiff.github.io

RESEARCH INTERESTS

AI for Discovery, AI for Social Good, Generative modeling, Optimal transport



EDUCATION

CORNELL UNIVERSITY – DEPARTMENT OF COMPUTER SCIENCE, New York, NY

Present

PhD candidate in Computer Science

- *Teaching*: Applied Machine Learning (TA Fall 2022)

NEW YORK UNIVERSITY – COURANT INSTITUTE OF MATHEMATICAL SCIENCES, New York, NY

May 2019

MS in Computer Science, GPA: 3.97/4.00

- *Relevant Coursework*: Advanced Machine Learning, Artificial Intelligence, Computer Vision, Deep Learning, Graphical Processing Units, Machine Learning, Mathematics of Deep Learning, Predictive Analytics

UNIVERSITY OF PENNSYLVANIA – COLLEGE OF ARTS AND SCIENCES, Philadelphia, PA

May 2014

BA Summa Cum Laude with Distinction in Economics, GPA: 3.93/4.00

- *Academic Honors*: Phi Beta Kappa, Dean's list 2010-2014



RESEARCH AND PUBLICATIONS

PUBLICATIONS

Cloud-Based Real-Time Molecular Screening Platform
with MolFormer

[ECML PKDD 2022 Demo Track](#)

(*alphabetical order
equal contribution)

Brian Belgodere, Vijil Chenthamarakshan*, Payel Das*, Pierre Dognin*,
Toby Kurien*, Igor Melnyk*, Youssef Mroueh*, Inkit Padhi*, Mattia Rigotti*,
Jarret Ross*, Yair Schiff*, Richard A. Young**

Optimizing Functionals on the Space of Probabilities with
Input Convex Neural Networks

[Transactions of Machine Learning Research](#)

David Alvarez-Melis, Yair Schiff, Youssef Mroueh

Augmenting Molecular Deep Generative Models with
Topological Data Analysis Representations

[ICASSP 2022](#)

(*equal contribution)

Yair Schiff, Vijil Chenthamarakshan*, Samuel Hoffman*,
Karthikeyan Natesan Ramamurthy*, Payel Das**

Predicting Deep Neural Network Generalization with Perturbation Response Curves

[NeurIPS 2021](#)

Yair Schiff, Brian Quanz, Payel Das, Pin-Yu Chen

Tabular Transformers for Modeling Multivariate Time Series

[ICASSP 2021](#)

*Inkit Padhi, Yair Schiff, Igor Melnyk, Mattia Rigotti, Youssef Mroueh, Pierre Dognin,
Jarret Ross, Ravi Nair, Erik Altman*

Image Captioning as an Assistive Technology: Lessons Learned from
VizWiz 2020 Challenge

[Journal of AI Research](#)

(*alphabetical order
equal contribution)

Pierre Dognin, Igor Melnyk*, Youssef Mroueh*, Inkit Padhi*, Mattia Rigotti*,
Jarret Ross*, Yair Schiff*, Richard Young, Brian Belgodere*

WORKSHOPS

Optimizing Functionals on the Space of Probabilities with
Input Convex Neural Networks
*David Alvarez-Melis, **Yair Schiff**, Youssef Mroueh*

[NeurIPS Workshop 2021](#)
Spotlight presentation

Gi and Pal Scores: Deep Neural Network Generalization Statistics
***Yair Schiff**, Brian Quanz, Payel Das, Pin-Yu Chen*

[ICLR Workshop 2021](#)

Characterizing the Latent Space of Molecular Deep Generative Models
with Persistent Homology Metrics
***Yair Schiff**, Vijil Chenthamarakshan, Karthikeyan Natesan Ramamurthy, Payel Das*

[NeurIPS Workshop 2020](#)
Spotlight presentation

Alleviating Noisy Data in Image Captioning with Cooperative Distillation
Pierre Dognin, Igor Melnyk*, Youssef Mroueh*, Inkit Padhi*, Mattia Rigotti*,
Jarret Ross*, **Yair Schiff****

[CVPR Workshop 2020](#)
*(*alphabetical order,
equal contribution)*

TALKS AND PRESENTATIONS

Topological Data Analysis and Beyond Workshop

[NeurIPS 2020](#)

- Presented spotlight poster “Characterizing the Latent Space of Molecular Deep Generative Models” ([video](#))

VizWiz Grand Challenge Workshop

[CVPR 2020](#)

- Presented winning submission to VizWiz Grand Challenge ([video](#))
- Presented “Alleviating Noisy Data in Image Captioning with Cooperative Distillation” ([video](#))

OPEN-SOURCE CONTRIBUTIONS

TabFormer: Tabular Transformers for Modeling Multivariate Time Series

[Github.com](#)

- Wrote code for training and evaluating GPT-like models on tabular data to generate new, synthetic data that matches the underlying distributions of the real table variables

pytorch-PPUU: Prediction and Policy-learning Under Uncertainty

[Github.com](#)

- Added a new dataset on which the self-driving policy could be trained
- Enhanced the self-driving vehicle’s policy to enable dynamic lane changes

ONLINE PUBLICATIONS FOR WATSON MACHINE LEARNING PRODUCT LAUNCHES

“Data, data everywhere...”: Leveraging IBM Watson Studio for
private data with Federated Learning
***Yair Schiff**, Jim Rhyness*

[Medium.com](#)

Unlocking your data’s potential with IBM Watson Studio’s AutoAI
feature engineering on relational data
Yair Schiff

[Medium.com](#)

Breaking the Magicians’ code with IBM Watson Studio’s AutoAI Notebooks
Yair Schiff

[Medium.com](#)

Peeking behind the curtain with IBM Watson AutoAI Python Client
*Lukasz Cmielowski, **Yair Schiff**, Przemyslaw Czuba*

[Medium.com](#)

Automating the AI Lifecycle with IBM Watson Studio Orchestration Flow
***Yair Schiff**, Rafal Bigaj*

[Medium.com](#)

Right on time(series): Introducing Watson Studio’s AutoAI Time Series
Yair Schiff

[Medium.com](#)

PAST PROJECTS

Prediction and Planning Under Uncertainty

Research Assistant in Professor Yann LeCun's lab (under supervision of Dr. Alfredo Canziani)

- Contributed to ongoing research exploring planning under uncertainty in the context of self-driving vehicles
- Incorporated large aerial traffic dataset model training and enhanced vehicle's policy to enable lane changes



WORK EXPERIENCE

IBM WATSON MACHINE LEARNING, New York, NY

Aug 2019 – Aug 2022

Cognitive Software Developer

- Contribute to continuous development and testing of Watson Machine Learning products
- Facilitate weekly Cloud releases, Cloud Pak for Data platform releases, and the launch of AutoAI feature engineering on relational data, AutoAI Time Series, AutoAI Notebooks, and Federated Learning products
- Published 6 medium.com articles about Watson Machine Learning product releases
- Received Outstanding Technical Achievement Award for work on the release of AutoAI feature engineering on relational data
- Received CrushIT Team Excellence Award as part of the Watson Machine Learning Training team

Research Contributor to IBM Research AI Challenges

Aug 2019 – Aug 2022

- Volunteer to contribute to IBM Research AI challenges, working with the Trusted AI Department
- Member of the first-place winning team in the 2020 VizWiz Grand Challenge: Image Captioning as an Assistive Technology for the Visually Impaired
- Co-author with the Trusted AI team on several publications in the fields of Generative Modeling, Molecular Discovery, Deep Learning Generalization, and AI for Social Good
- Received two 2021 IBM Research Accomplishments awards for contributions to (1) trustworthy AI generative modeling and (2) deployment of large-scale transformer models on OpenShift environments

SIMON-KUCHER AND PARTNERS, New York, NY

Sept 2014 – Aug 2017

Consultant

- Advised global companies spanning various industries – including internet, media, consumer electronic goods, and chemicals – on areas for better revenue capture
- Synthesized large data sets (e.g., 300 million+ client transactions), customer research (surveys & conjoint studies sent to thousands of respondents), and secondary research to create solutions to client needs



SKILLS

- *Programming Languages:* C++, Java, Python
- *Deep Learning Frameworks:* PyTorch, PyTorch Lightning
- *Data Tools:* Excel, Stata, Tableau
- *Foreign Languages:* Fluent in Hebrew