

Yair Schiff

yairschiff@cs.cornell.edu

yair-schiff.github.io

RESEARCH INTERESTS

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



EDUCATION

CORNELL UNIVERSITY – DEPARTMENT OF COMPUTER SCIENCE, New York, NY 2022 – Present

PhD student in Computer Science

- *Awards:* Hal & Inge Marcus PhD Fellowship (2022-2023)
- *Teaching:* Applied Machine Learning (TA Fall 2022), Deep Learning (TA Spring 2023)

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

Learning with Stochastic Orders

[ICLR 2023](#)

Carles Domingo-Enrich, **Yair Schiff**, Youssef Mroueh

(Notable Top 25% acceptance)

Semi Parametric Inducing Point Networks

[ICLR 2023](#)

Richa Rastogi, **Yair Schiff**, Alon Hachohen, Zhaozhi Li, Ian Lee, Yuntian Deng, Mert R. Sabuncu, Volodymyr Kuleshov

Cloud-Based Real-Time Molecular Screening Platform
with MolFormer

[ECML PKDD 2022 Demo Track](#)

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*

(*alphabetical order
equal contribution)

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](#)

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

(*equal contribution)

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

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

[Journal of AI Research](#)

(*alphabetical order
equal contribution)

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

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

[NeurIPS 2020](#)

VizWiz Grand Challenge Workshop

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

[CVPR 2020](#)

OPEN-SOURCE CONTRIBUTIONS

TabFormer: Tabular Transformers for Modeling Multivariate Time Series

- 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

[Github.com](#)

pytorch-PPUU: Prediction and Policy-learning Under Uncertainty

- 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

[Github.com](#)

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

- Contributed to continuous development and testing of Watson Machine Learning products
- Facilitated 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

- Volunteered 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-authored 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



GROUPS AND AFFILIATIONS

CORNELL TECH K-12 INITIATIVE, New York, NY

Present

Volunteer

- Facilitate coding and learning events for local students to make Computer Science education more accessible

GRADS FOR GENDER INCLUSION IN COMPUTING, New York, NY

Present

Member

- Member of a group dedicated to combating harassment, pushing for policy change, and creating supportive spaces

STUDENT-APPLICANT SUPPORT PROGRAM, New York, NY

Present

Volunteer

- Provide feedback and support to students applying to graduate programs in Computer Science and related fields

PHD PEER MENTOR PROGRAM, New York, NY

Present

Mentor

- Meet with fellow PhD student to provide guidance about graduate school life at Cornell Tech



SKILLS

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