Yair Schiff [yairschiff@cs.cornell.edu](mailto:yairschiff@cs.cornell.edu)

[yair-schiff.github.io](https://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, NYPresent

*PhD candidate in Computer Science*

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

**New York University** – Courant Institute of Mathematical Sciences, New York, NYMay 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, PAMay 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 [ECML PKDD 2022 Demo Track](https://arxiv.org/abs/2208.06665)

with MolFormer *(\*alphabetical order*

*Brian Belgodere\*, Vijil Chenthamarakshan\*, Payel Das\*, Pierre Dognin\*, equal contribution)*

*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 [Transactions of Machine Learning Research](https://arxiv.org/abs/2106.00774)

Input Convex Neural Networks

*David Alvarez-Melis,* ***Yair Schiff****,* *Youssef Mroueh*

Augmenting Molecular Deep Generative Models with [ICASSP 2022](https://arxiv.org/abs/2106.04464)

Topological Data Analysis Representations *(\*equal contribution)*

***Yair Schiff\*****, Vijil Chenthamarakshan\*, Samuel Hoffman\*,*

*Karthikeyan Natesan Ramamurthy\*, Payel Das\**

Predicting Deep Neural Network Generalization with Perturbation Response Curves [NeurIPS 2021](https://arxiv.org/abs/2106.04765)

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

Tabular Transformers for Modeling Multivariate Time Series [ICASSP 2021](https://ieeexplore.ieee.org/document/9414142)

*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 [Journal of AI Research](https://arxiv.org/abs/2012.11696)

VizWiz 2020 Challenge *(\*alphabetical order*

*Pierre Dognin\*, Igor Melnyk\*, Youssef Mroueh\*, Inkit Padhi\*, Mattia Rigotti\*, equal contribution)*

*Jarret Ross\*,* ***Yair Schiff****\*, Richard Young, Brian Belgodere*

**Workshops**

Optimizing Functionals on the Space of Probabilities with [NeurIPS Workshop 2021](https://arxiv.org/abs/2106.00774)

Input Convex Neural Networks *Spotlight presentation*

*David Alvarez-Melis,* ***Yair Schiff****,* *Youssef Mroueh*

Gi and Pal Scores: Deep Neural Network Generalization Statistics [ICLR Workshop 2021](https://arxiv.org/abs/2104.03469)

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

Characterizing the Latent Space of Molecular Deep Generative Models [NeurIPS Workshop 2020](https://arxiv.org/abs/2010.08548)

with Persistent Homology Metrics *Spotlight presentation*

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

Alleviating Noisy Data in Image Captioning with Cooperative Distillation [CVPR Workshop 2020](https://arxiv.org/abs/2012.11691)

*Pierre Dognin\*, Igor Melnyk\*, Youssef Mroueh\*, Inkit Padhi\*, Mattia Rigotti\*, (\*alphabetical order,*

*Jarret Ross\*,* ***Yair Schiff****\* equal contribution)*

**Talks and Presentations**

Topological Data Analysis and Beyond Workshop [NeurIPS 2020](https://tda-in-ml.github.io/papers)

* Presented spotlight poster “Characterizing the Latent Space of Molecular Deep Generative Models” ([video](https://slideslive.com/38941579/characterizing-the-latent-space-of-molecular-deep-generative-models-with-persistent-homology-metrics?ref=account-folder-62106-folders))

VizWiz Grand Challenge Workshop [CVPR 2020](https://vizwiz.org/workshops/2020-workshop/)

* Presented winning submission to VizWiz Grand Challenge ([video](https://ivc.ischool.utexas.edu/~yz9244/VizWiz_workshop/videos/MMTeam-oral.mp4))
* Presented “Alleviating Noisy Data in Image Captioning with Cooperative Distillation” ([video](https://ivc.ischool.utexas.edu/~yz9244/VizWiz_workshop/videos/103-poster.mp4))

**Open-Source Contributions**

TabFormer: Tabular Transformers for Modeling Multivariate Time Series [Github.com](https://github.com/IBM/TabFormer)

* 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](https://github.com/Atcold/pytorch-PPUU)

* 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 [Medium.com](https://yairschiff.medium.com/data-data-everywhere-leveraging-ibm-watson-studio-for-private-data-with-federated-learning-80e4310f46f4)

private data with Federated Learning

***Yair Schiff,*** *Jim Rhyness*

Unlocking your data’s potential with IBM Watson Studio’s AutoAI [Medium.com](https://medium.com/ibm-data-ai/unlocking-your-datas-potential-with-ibm-watson-studio-s-autoai-feature-engineering-on-relational-71a5dcac7b0d)

feature engineering on relational data

***Yair Schiff***

Breaking the Magicians’ code with IBM Watson Studio’s AutoAI Notebooks [Medium.com](https://yairschiff.medium.com/breaking-the-magicians-code-with-ibm-watson-studio-s-autoai-notebooks-839059333487)

***Yair Schiff***

Peeking behind the curtain with IBM Watson AutoAI Python Client [Medium.com](https://lukasz-cmielowski.medium.com/peaking-behind-the-curtain-with-ibm-watson-autoai-python-client-3062836c048)

*Lukasz Cmielowski,* ***Yair Schiff****,**Przemyslaw Czuba*

Automating the AI Lifecycle with IBM Watson Studio Orchestration Flow [Medium.com](https://medium.com/ibm-data-ai/automating-the-ai-lifecycle-with-ibm-watson-studio-orchestration-flow-4450f1d725d6)

***Yair Schiff,*** *Rafał Bigaj*

Right on time(series): Introducing Watson Studio’s AutoAI Time Series [Medium.com](https://medium.com/ibm-data-ai/right-on-time-series-introducing-watson-studios-autoai-time-series-5175dbe66154)

***Yair Schiff***

**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 EXPERINCE

**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 Challenge*s 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