Nadiia Chepurko

nadiiachepurko@gmail.com
Linkedin
Personal Website

EDUCATION

Massachusetts Institute of Technology — Cambridge, MA
Ph.D. in Computer Science — 2020–Current
Advisor: Antonio Torralba

Massachusetts Institute of Technology — Cambridge, MA
M.S. in Computer Science — 2017–2020

- GPA: 5.00/5.00
- Thesis: "Automatic Relational Data Augmentation for Machine Learning"
- Hewlett Packard Fellowship

Rutgers University — New Brunswick, NJ
B.S. in Computer Science — 2013–2016

- GPA : 4.00/4.00
- Summa Cum Laude graduation honor
- Academic Excellence Scholarship (2015)
- Academic Excellence Award (2014)
- Winner of Database challenge project (2014)

EXPERIENCE

Microsoft Research — Redmond, WA
Software Engineering Intern in Research — Summer 2020

- Worked in the DMX group on AutoML research project FLAML.
- Improved and added various system components in the system to increase the speed and accuracy of prediction.
- Worked with regression, SVM, random forest, XGBoost, Catboost, neural models, and other predictive models, methods, and algorithms.
- Improved sampling techniques.
- Incorporated ensemble models in the pipeline.
- Tuned various hyper-parameters for improved performance.
- Tested and verified correctness of computational pipelines.
- Identified and fixed bugs.

Adobe Research — San Jose, CA
Software Engineering Intern in Research — Summer 2018

- Worked on Question-Answering (NLP) system for Creative Cloud.
- Designed and implemented neural models for Question-Answering on text.
- Used language modeling as unsupervised pre-training.
- Built a demo system with model integration.

Rutgers University — New Brunswick, NJ
Software Engineer and Research Fellow — Jan 2016-Jan 2017

- Integrated various APIs and services that are currently available in the form of different mobile applications.

- Combined the chain of APIs and Deeplinks to unify multiple services under one interactive chat interface.
- Developed a web crawler to mine app descriptions and analyze the content.
- Designed stratified sampling techniques for Apache Spark, including how to maximize precision for stratums under fixed sample size, and how to sample the data under user-provided error bound for estimating aggregation functions such as sum and mean.
- Benchmarked sampling in Apache Spark framework versus Hadoop framework.

CareCloud   Miami, FL
Software Engineering Intern   Summer 2015
- Optimized queries and procedures for medical reports.
- Migrated and extended pivot table for multidimensional database (OLAP cube).
- Developed server health-check service.
- Gathered information from third-party software providers.

Luxoft   Kyiv, Ukraine
QA Engineer   Sept 2012-Aug 2013
- Customer: UBS Bank
- Worked in FICC department with CDF derivatives.
- Automated routines to verify correctness of daily reports to clearing house.
- Tested various financial trading platforms.

SELECTED PROJECTS

Dissecting Backpropagation
Interpreting how the learning dynamics of a deep network evolves over the course of the training phase.
- Developed a new *network dissection* method to visualize and quantify the emergence of semantics in representation gradients during training.
- Studied overfitting and adversarial robustness through the lens of interpretability.
- Studied emergent backpropagation neuron features with VQGAN+CLIP.

Generator-based Adversarial Attacks
Creating generator-based adversarial attacks for vision models that are transferable between different architectures and datasets in black-box fashion.
- Mentoring (UROP): Sauhaarda (Raunak) Chowdhuri (MIT EECS).
- Developed a generator-based adversarial attacks with novel learning techniques that incorporate deep model ensembles and produce cross-domain, black-box, and universal attacks in milliseconds.
- Introduced diversity induced objective approach and improved generator-based attacks in cross-domain and black-box fashion by 10-15%.

Automatic Relational Data Augmentation
Finding new features relevant to the user's predictive task with minimal "human-in-the-loop" involvement.
- Designed and implemented an end-to-end system that takes as input a data repository, and outputs an augmented data set such that training a predictive model on this augmented dataset results in improved prediction accuracy.
- Designed novel feature selection randomization based algorithm that outperformed other feature selection algorithms in highly noisy setup.
- Outperformed multiple AutoML systems performance in both accuracy and time.
- Experimented with targeted structural similarity based attacks to decrease attack's visual perceptibility.

Teaching

- **Teaching Assistant** at MIT   Fall 2019
  Advances in Computer Vision (6.819/6.869)
- **Teaching Assistant** at Get Smarter   Spring 2019
  Human Computer Interaction

Programming

- Languages
  Python, Java, C, JavaScript, SQL, HTML/CSS
- Libraries and Tools
  PyTorch, Scikit-learn, Pandas, SciPy, JUnit

Talks

1. **"Weighted Maximum Independent Set of Geometric Objects in Turnstile Streams"**. APPROX 2020
2. **"Robust and Sample Optimal Algorithms for PSD Low-Rank Approximation"**. FOCS 2020
3. **"ARDA: Automatic Relational Data Augmentation for Machine Learning"**. VLDB 2020

Publications in systems and machine learning

1. **Learning Program Representations for Food Images and Cooking Recipes.**
   Dim P. Papadopoulos, Enrique Mora, Nadiia Chepurko, Kuan Wei Huang, Ferda Ofli, Antonio Torralba
   CVPR 2022.
2. **ARDA: Automatic Relational Data Augmentation for Machine Learning**.
   Nadiia Chepurko, Ryan Marcus, Emmanuel Zggragen, Raul Castro Fernandez, Tim Kraska, David Karger
   VLDB 2020.

Publications in algorithms

1. **Near-Optimal Algorithms for Linear Algebra in the Current Matrix Multiplication Time.**
   Nadiia Chepurko* , Kenneth L. Clarkson*, Praneeth Kacham *, Lior Horesh*, David Woodruff*
   SODA 2022.
2. **Quantum-Inspired Algorithms from Randomized Numerical Linear Algebra.**
   Nadiia Chepurko* , Kenneth L. Clarkson*, Lior Horesh*, Honghao Lin*, David Woodruff*
   ICML 2022.
3. **Robust and Sample Optimal Algorithms for PSD Low-Rank Approximation.**
   Nadiia Chepurko*, Ainesh Bakshi*, David Woodruff *
   FOCS 2020.
4. **Testing Positive Semi-Definiteness via Random Submatrices.**
   Nadiia Chepurko*, Ainesh Bakshi*, Rajesh Jayaram*
   FOCS 2020.
5. **Weighted Maximum Independent Set of Geometric Objects in Turnstile Streams.**
   Nadiia Chepurko*, Ainesh Bakshi*, David Woodruff*
   APPROX 2020.

---

*Equal contribution