

# MAX TOROP

✉ [torop.m@northeastern.edu](mailto:torop.m@northeastern.edu) 🌐 [maxtorop.github.io](https://maxtorop.github.io) ☎ +1 347-224-6458

## EDUCATION

---

### Northeastern University

PhD Candidate in Electrical Engineering, GPA: 4.00/4.00

- **Advisor:** Professor Jennifer Dy

Boston, MA

Sep 2020 — Current

### Washington University in St. Louis

MS in Computer Science, GPA: 3.97/4.00

- **Advisor:** Professor Ulugbek Kamilov

St. Louis, MO

Sep 2018 — Dec 2019

### The University of Rochester

BS in Data Science, GPA: 3.40/4.00

Rochester, NY

Sep 2014 — May 2018

## ACADEMIC EXPERIENCE

---

### Machine Learning Lab

Northeastern University

Boston, MA

Sep 2020 — Current

- Developed a method for controlling LLM behavior using steering vectors on query and value representations. Empirically and theoretically characterized its properties and evaluated on open-source LLMs (**NeurIPS 2025**).
- Created SmoothHess, a geometrically intuitive method for interpreting neural network predictions through feature interactions. Evaluated on models trained with image and spirometry data (**NeurIPS 2023**).
- Collaborated to develop a method to quantify model explanation uncertainty as a function of the local decision boundary (**AISTATS 2024**).
- Working with doctors at **Memorial Sloan Kettering Cancer Center** on deep learning methods for dermatology spanning multi-modal LLMs and self-supervised learning.

### Computational Imaging Group

Washington University in St. Louis

St. Louis, MO

May 2019 — Dec 2019

- Collaborated with radiologists to develop a self-supervised deep learning model that jointly transforms and denoises MRI data into brain iron maps (**Magnetic Resonance in Medicine, 2020**).
- Mentored an undergraduate student in the lab for her senior thesis project.

## WORK EXPERIENCE

---

### Apple

ML Research Intern

Seattle, WA

May 2024 — Aug 2024

- Worked in the Data and Machine Learning Innovation team.
- Created a dynamic LLM-in-the-loop knowledge graph generation application.
- Developed methods for selecting LLM training data.

### iD Tech Camps

Instructor

New York, NY

July 2018 — Aug 2018

- Taught teenagers to create neural networks and familiarized them with basic machine learning practices.

## PUBLICATIONS AND TALKS

---

### *Published*

- **Max Torop**, Aria Masoomi, Masih Eskandar, Jennifer Dy. “DISCO: Disentangled Communication Steering for Large Language Models.” **NeurIPS 2025**.
- Davin Hill\*, Joshua Bone\*, Aria Masoomi, **Max Torop**, Jennifer Dy. “Axiomatic Explainer Globalness via Optimal Transport.” **AISTATS 2025**.
- Davin Hill, Aria Masoomi, **Max Torop**, Sandesh Ghimire, Jennifer Dy. “Boundary-Aware Uncertainty for Feature Attribution Explainers.” **AISTATS 2024**.
- **Max Torop\***, Aria Masoomi\*, Davin Hill, Kivanc Kose, Stratis Ioannidis, Jennifer Dy. “SmoothHess: ReLU Network Feature Interactions via Stein’s Lemma.” **NeurIPS 2023**.

- Davin Hill, **Max Torop**, Aria Masoomi, Peter J. Castaldi, Jennifer Dy, Michael H. Cho, Brian D. Hobbs. “Deep Learning Utilizing Discarded Spirometry Data to Improve Lung Function and Mortality Prediction in the UK Biobank.” ATS 2022 (Oral).
- **Max Torop**, Sandesh Ghimire, Wenqian Liu, Dana H. Brooks, *et al.* “Unsupervised Approaches for Out-Of-Distribution Dermoscopic Lesion Detection.” MedNeurIPS Workshop, NeurIPS 2021.
- **Max Torop**, Satya V.V.N. Kothapalli, Yu Sun, Jiaming Liu, Sayan Kahali, Dmitriy A. Yablonskiy, Ulugbek S. Kamilov. “Deep learning using a biophysical model for robust and accelerated reconstruction of quantitative, artifact-free and denoised images.” **Magnetic Resonance in Medicine**, vol. 84, no. 6, pp. 2932–2942, 2020.

### *Preprints*

- **Max Torop**, Masih Eskandar, Nicholas Kurtansky, Jinyang Liu, *et al.* “Grounding Multimodal Large Language Models with Quantitative Skin Attributes: A Retrieval Study.” arXiv 2025. In submission.
- Davin Hill, **Max Torop**, Aria Masoomi, Peter J. Castaldi, *et al.* “Deep Learning of Suboptimal Spirometry to Predict Respiratory Outcomes and Mortality.” medRxiv 2023. In submission.

### *Talks*

- “SmoothHess: ReLU Network Feature Interactions via Stein’s Lemma,” Prof. Finale Doshi-Velez’s Data to Actionable Knowledge (DtAK) lab, Harvard, Boston, MA, 10/2024.
- “Unsupervised representation learning for detecting out of distribution samples in dermoscopy images of eight types of skin lesions,” SPIE BiOS, San Francisco, CA (online), 03/2022.

## TEACHING AND LEADERSHIP

---

- Teaching assistant for the Advanced Machine Learning course at Northeastern University (Spring 2022).
- Organized the purchase and setup of a server for our group. Developed usage guidelines documentation.
- Co-organized a bi-weekly seminar for the SPIRAL group at Northeastern University. Invited speakers included: Brian Kulis (BU), Amin Karbasi (Yale), Michael Hughes (Tufts) and David Rosen (NEU).

## SKILLS AND SERVICE

---

- **Research:** Representation engineering for LLMs, interpretable ML.
- **Languages/Tools:** Python, PyTorch, pandas, NumPy, scikit-learn, L<sup>A</sup>T<sub>E</sub>X, SLURM.
- **Familiar:** MATLAB, Java.
- **Reviewer:** NeurIPS, AISTATS, AAAI and SIVP.

## AWARDS

---

- |   |                            |
|---|----------------------------|
| • Dean’s Fellowship (Northeastern)          | <i>Sep 2020 — May 2024</i> |
| • Dean’s List, 5 Terms (Rochester)          | <i>Sep 2014 — May 2018</i> |
| • Research and Innovation Grant (Rochester) | <i>Sep 2014</i>            |