

# NAOMI SAPHRA

## Gradient Descent Spectator

@ nsaphra@nsaphra.net    www.nsaphra.net    @nsaphra    github.com/nsaphra



## WORK EXPERIENCE

### Kempner Fellow

#### Harvard University

2023 – present    Boston, MA

### Postdoctoral Researcher

#### NYU

2021 – 2023    New York, New York

- Supervisor: Kyunghyun Cho.

### NLP Consultant

#### MosaicML

2022    New York, New York

- Advised R&D on efficient LM training.

### Research Intern

#### Google

Winter 2020    New York, New York

- Supervisors: Dipanjan Das (Language, Google AI), Ian Tenney, Jasmijn Bastings, Thibault Sellam.
- Better understanding of how BERT learns.

### Contractor

#### Koko

Summer 2017    New York, New York

- Developed neural classifiers for informal text at anti-abuse startup.

### Software Engineering Intern

#### Google

Summer 2013 & 2015    Mountain View, California

- Supervisor: Marius Pasca (Common Sense, Machine Intelligence).
- Supervisor: Eric Altendorf (Common Sense, Machine Intelligence).

### Undergraduate Research Fellow

#### CLSP Summer Workshop at Johns Hopkins University

Summer 2012    Baltimore, Maryland

- Supervisor: Matthew Blaschko.
- Unsupervised methods to verify image annotations.

### Engineering Intern

#### Facebook

Summer 2011    Palo Alto, California

## EDUCATION

### Ph.D. in Informatics

#### Edinburgh University

2021

Supervisor: Adam Lopez.  
Thesis: Training Dynamics of Neural Language Models.

### M.Sc. in Computer Science

#### Johns Hopkins University

2015

### B.Sc. in Computer Science

#### Carnegie Mellon University

2013

Minor: Language Technologies

### Sabbatical

#### The Recurse Center

2017

## AWARDS

- MIT EECS Rising Stars.
- ICML outstanding reviewer 2024.
- ICLR outstanding reviewer 2021, 2022.
- New York Academy of Sciences - Natural Language, Dialogue and Speech Symposium: Best Poster Runner Up 2019.
- Daniella Sciamma Award for Achievement through Adversity 2019.
- Google Europe Scholarship for Students with Disabilities 2017.
- Best paper at LREC Workshop on Free Arabic Corpora 2014 (coauthor).
- Carnegie Mellon School of Computer Science Dragon Award 2013.
- National Merit Semifinalist 2009.

## SKILLS & HOBBIES

- NLP    interpretability    generalization
- language models    training dynamics
- Roller Derby    Standup    Clown

## SELECTED PUBLICATIONS

- Chen, Angelica et al. (2024). “Sudden Drops in the Loss: Syntax Acquisition, Phase Transitions, and Simplicity Bias in MLMs”. In: *ICLR*. Last author. Spotlight (top 5%).
- **Saphra, Naomi**, Eve Fleisig, et al. (2024). “First Tragedy, then Parse: History Repeats Itself in the New Era of Large Language Models”. In: *North American Association for Computational Linguistics (NAACL)*.
- Sherborne, Tom et al. (2024). “TRAM: Bridging Trust Regions and Sharpness Aware Minimization”. In: *ICLR*. Spotlight (top 5%).
- Hu, Michael et al. (2023). “Delays, Detours, and Forks in the Road: Latent State Models of Training Dynamics”. In: *TMLR*.
- Juneja, Jeevesh et al. (2023). “Linear Connectivity Reveals Generalization Strategies”. In: *ICLR*. Last author.
- Sellam, Thibault et al. (2022). “The MultiBERTs: BERT Reproductions for Robustness Analysis”. In: *ICLR*.
- Valvoda, Josef et al. (2022). “Learning Transductions to Test Systematic Compositionality”. In: *COLING*.
- White, Jennifer C. et al. (2021). “A Non-Linear Structural Probe”. In: *NAACL*.
- Pimentel, Tiago et al. (2020). “Pareto Probing: Trading Off Accuracy for Complexity”. In: *EMNLP*. Joint first author.
- **Saphra, Naomi** and Adam Lopez (2020). “LSTMs Compose—and Learn—Bottom-Up”. In: *Findings of EMNLP*.
- **Saphra, Naomi** and Adam Lopez (2019). “Understanding Learning Dynamics Of Language Models with SVCCA”. in: *NAACL*.
- **Saphra, Naomi** and Adam Lopez (2015). “AMRICA: an AMR Inspector for Cross-language Alignments”. In: *NAACL*.
- Cotterell, Ryan et al. (2014). “An Algerian Arabic-French Code-Switched Corpus”. In: *LREC Workshop on Free/Open-Source Arabic Corpora*. **Best paper award**.

## SERVICE

### Organizing Committee

**ACL Workshop on Representation Learning for NLP**

📅 2021, 2023

### Organizing Committee

**EMNLP Workshop on analyzing and interpreting neural nets for NLP**

📅 2022

### Area Chair

**EMNLP and AACL**

📅 2021–Present

### ACL Accessibility Subcommittee

**Association for Computational Linguistics**

📅 2020 📍 Seattle, WA

### Student social co-chair

**Association for Computational Linguistics**

📅 2014 📍 Baltimore, MD

### Puzzle tester and local organizer

**North American Computational Linguistics Olympiad**

📅 2012–2014

## TEACHING

### Project Supervisor

**NYU**

📅 2021

Machine Learning MS Capstone

### Tutor

**University of Edinburgh**

📅 2016 – 2019

Informatics Research Review  
Probabilistic Modeling & Reasoning  
Machine Learning & Pattern Recognition

### Course Assistant

**Johns Hopkins University**

📅 2013

Natural Language Processing

### Teacher’s Assistant

**Irvington High School**

📅 2008 – 2009

Ancient Greek

## INVITED TALKS

### Sources of Variance in Pretraining and Finetuning (Keynote)

**3rd Neural Scaling Laws Workshop**

📅 June 2022 📍 Quebec

### Transparent, Hackable, Accessible

**QueerInAI EMNLP Social**

📅 November 2020 📍 Remote (EMNLP)

### Accessible Means Hackable (Keynote)

**Pydatafest Amsterdam**

📅 June 15, 2020 📍 Amsterdam, Netherlands

### Blackbox NLP Panel Discussion

**ACL Blackbox NLP Workshop**

📅 Aug 1, 2019 📍 Florence, Italy

### Understanding Learning Dynamics Of Language Models with SVCCA

**Understanding & Analyzing Neural Networks Workshop**

📅 Aug 1, 2019 📍 Amsterdam, Netherlands

### Paying the Panopticon (Standup)

**The Stand Edinburgh Comedy Club**

📅 Feb 26, 2019 📍 Edinburgh, UK