

NAOMI SAPHRA

Curriculum Vitae

Center for Data Science
New York University
✉ nsaphra@nsaphra.net
📄 nsaphra.net
🌐 [nsaphra](#) 🐦 [nsaphra](#)



Education

- 2021 **Ph.D, Informatics**, *University of Edinburgh*, Edinburgh, UK.
Advisors: Adam Lopez and Tim Hospedales
Thesis: Training Dynamics of Neural Language Models.
- 2015 **MSE, Computer Science**, *Johns Hopkins University*, Baltimore, MD.
Mentors: Adam Lopez, Sanjeev Khudanpur, Raman Arora
- 2013 **B.Sc, Computer Science**, *Carnegie Mellon University*, Pittsburgh, PA.
Mentors: Noah Smith and Chris Dyer
Minor: Language Technologies
- Spring 2017 **Sabbatical: Recurser**, *The Recurse Center*, New York, NY.
Experimental unstructured educational programming retreat.

Work

- 2021–Present **NYU**, *Postdoctoral Researcher*, New York, NY.
Mentor: Kyunghyun Cho.
Connecting loss landscape geometry to generalization behavior.
Evolutionary analysis of language model training.
Compositionality in multimodal models.
- 2022–Present **MosaicML**, *Consultant*, New York, NY.
Improving the efficiency of the pretraining large language models.
- Winter 2020 **Google**, *Research Intern*, New York, NY.
Mentors: Dipanjan Das, Ian Tenney, Jasmijn Bastings, Thibault Sellam.
Better understanding of how BERT learns.
- Summer 2017 **Koko**, *Contractor*, New York, NY.
Neural classifiers for informal text at abuse-detection startup.
- Summer 2015 **Google**, *Software Engineering Intern*, Mountain View, CA.
Mentor: Marius Pasca (Common Sense Team, Machine Intelligence).
Unsupervised discovery of paraphrases of complex compositional noun phrases.
- Summer 2013 **Google**, *Software Engineering Intern*, Mountain View, CA.
Mentor: Eric Altendorf (Common Sense Team, Machine Intelligence).
Integrating models of real-world knowledge into entity linking systems.
- Summer 2012 **CLSP Summer Workshop at Johns Hopkins University**, *Undergraduate Research Fellow*, Baltimore, MD.
Mentor: Matthew Blaschko.
Unsupervised learning methods to verify and explore image annotations.
Selected from a competitive undergraduate pool.
- Summer 2011 **Facebook, Inc.**, *Engineering Intern*, Palo Alto, CA.
Mentor: Jen Burge (Engagement Team).
Engineering and testing new features for the People You May Know model.
- 2010–2013 **Carnegie Mellon University**, *Undergraduate Research Assistant*, Pittsburgh, PA.
Mentors: Chris Dyer and Noah Smith.
Projects include Twitter dependency parsing; multithreaded language models; political corpus analysis.

Publications

In Conference Proceedings

- 2022 Josef Valvoda, **Naomi Saphra**, Jonathan Rawski, Ryan Cotterell, and Adina Williams. Learning Transductions to Test Systematic Compositionality. In *COLING*, 2022.
- 2022 Thibault Sellam, Steve Yadlowsky, Ian Tenney, Jason Wei, **Naomi Saphra**, Alexander D'Amour, Tal Linzen, Jasmijn Bastings, Iulia Raluca Turc, Jacob Eisenstein, Dipanjan Das, and Ellie Pavlick. The MultiBERTs: BERT Reproductions for Robustness Analysis. In *ICLR*, 2022.
- 2021 Jennifer C. White, Tiago Pimentel, **Naomi Saphra**, Adina Williams, and Ryan Cotterell. A Non-Linear structural probe. In *NAACL*, 2021.
- 2020 **Naomi Saphra** and Adam Lopez. LSTMs Compose—and Learn—Bottom-Up. In *Findings of EMNLP*, 2020.
- 2020 Mohammad Tahaei, Kami Vaniea, and **Naomi Saphra**. Understanding privacy-related questions on stack overflow. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.
- 2020 Tiago Pimentel*, **Naomi Saphra*** (equal contribution), Adina Williams, and Ryan Cotterell. Pareto Probing: Trading Off Accuracy for Complexity. In *EMNLP*, 2020.
- 2019 **Naomi Saphra** and Adam Lopez. Understanding Learning Dynamics Of Language Models with SVCCA. In *NAACL*, 2019.
- 2015 **Saphra, Naomi** and Adam Lopez. AMRICA: an AMR Inspector for Cross-language Alignments. In *NAACL*, 2015.
- 2014 Andrea Vedaldi, Siddharth Mahendran, Stavros Tsogkas, Subhransu Maji, Ross Girshick, Juho Kannala, Esa Rahtu, Iasonas Kokkinos, Matthew B. Blaschko, David Weiss, Ben Taskar, Karen Simonyan, **Naomi Saphra**, and Sammy Mohamed. Understanding Objects in Detail with Fine-grained Attributes. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

Workshop Publications

- 2022 Jeevesh Juneja, Rachit Bansal, Kyunghyun Cho, Joao Sedoc, and **Naomi Saphra**. Linear Connectivity Reveals Generalization Strategies. In *ICML Workshop on Spurious Correlations, Invariance, and Stability*, 2022.
- 2019 **Saphra, Naomi** and Adam Lopez. Sparsity emerges naturally in neural language models. In *ICML Workshop on Identifying and Understanding Deep Learning Phenomena*, 2019.
- 2016 **Naomi Saphra** and Adam Lopez. Evaluating Informal-Domain Word Representations With UrbanDictionary. In *ACL Workshop on Evaluating Vector Space Representations for NLP*, 2016.
- 2014 Nathan Schneider, Brendan O'Connor, **Naomi Saphra**, David Bamman, Manaal Faruqui, Noah A. Smith, Chris Dyer, and Jason Baldridge. A framework for (under) specifying dependency syntax without overloading annotators. In *ACL Linguistic Annotation Workshop*, 2014.
- 2014 Ryan Cotterell, Adithya Renduchintala, **Naomi Saphra**, and Chris Callison-Burch. An Algerian Arabic-French Code-Switched Corpus. In *LREC Workshop on Free/Open-Source Arabic Corpora*, 2014.

Preprints

- 2022 Jeevesh Juneja, Rachit Bansal, Kyunghyun Cho, João Sedoc, and **Naomi Saphra**. Linear Connectivity Reveals Generalization Strategies. *arXiv preprint arXiv:2205.12411*, 2022.
- 2022 Dieuwke Hupkes, Mario Giulianelli, Verna Dankers, Mikel Artetxe, Yanai Elazar, Tiago Pimentel, Christos Christodoulopoulos, Karim Lasri, **Naomi Saphra**, Arabella Sinclair, Dennis Ulmer, Florian Schottmann, Khuyagbaatar Batsuren, Kaiser Sun, Koustuv Sinha, Leila Khalatbari, Maria Ryskina, Rita Frieske, Ryan Cotterell, and Zhijing Jin. State-of-the-art generalisation research in NLP: a taxonomy and review. *arXiv preprint arXiv:2210.03050*, 2022.

- 2017 Graham Neubig, Chris Dyer, Yoav Goldberg, Austin Matthews, Waleed Ammar, Antonios Anastasopoulos, Miguel Ballesteros, David Chiang, Daniel Clothiaux, Trevor Cohn, Kevin Duh, Manaal Faruqui, Cynthia Gan, Dan Garrette, Yangfeng Ji, Lingpeng Kong, Adhiguna Kuncoro, Gaurav Kumar, Chaitanya Malaviya, Paul Michel, Yusuke Oda, Matthew Richardson, **Naomi Saphra**, Swabha Swayamdipta, and Pengcheng Yin. Dynet: The dynamic neural network toolkit. *arXiv preprint arXiv:1701.03980*, 2017.

Awards

- 2022 **Outstanding Reviewer**, *ICLR*.
2021 **Outstanding Reviewer**, *ICLR*.
2019 **Best Poster Runner Up**, *NY Academy of Sciences - Natural Language, Dialogue & Speech*.
2019 **Daniella Sciama Award for Achievement through Adversity**, *University of Edinburgh*.
2017 **Google Europe Scholarship for Students with Disabilities**.
2013 **Dragon Award**, *Carnegie Mellon School of Computer Science*.
2009 **National Merit Semifinalist**.
2009 **Nannina Rasulo Memorial Scholarship Award for Technology**, *Irvington High School*.

Supervision

- 2021 **University of Amsterdam MSc.**, Sylke Gosen.
Understanding Language Models through Perturbed Datasets.
2018 **University of Edinburgh MSc.**, Alp Ozkan.
Combined Application of Pruning and Growing Approaches for Neural Networks.
2018 **University of Edinburgh MSc.**, Yekun Chai.
Discovering Spelling Variants on Urban Dictionary.

Teaching

- 2021 **Center for Data Science Capstone**, *NYU*.
Project Mentor
2019 **Probabilistic Modeling & Reasoning**, *University of Edinburgh*.
Tutor (Teaching recitations)
2017–2019 **Machine Learning & Pattern Recognition**, *University of Edinburgh*.
Tutor (Teaching recitations)
2016 **Informatics Research Review**, *University of Edinburgh*.
Tutor (Teaching recitations)
2013 **Natural Language Processing**, *Johns Hopkins University*.
Course Assistant (Grading)
2008–2009 **Ancient Greek**, *Irvington High School*.
Teacher's Assistant (Grading)

Talks

Invited event talks

- June 2022 **3rd Neural Scaling Laws Workshop**, Manoir Saint-Sauveur, Quebec.
Sources of Variance in Pretraining and Finetuning (Keynote)
Jan 2022 **NYU AI School**, New York, NY (Remote).
Mathematical Fundamentals of AI
Nov 2020 **QueerInAI EMNLP Social**, Seattle, WA (Remote).
Transparent, Hackable, Accessible

- June 2020 **Pydatafest Amsterdam**, Amsterdam, Netherlands (Remote).
Accessible Means Hackable (Keynote)
- Aug 2019 **ACL Blackbox NLP Workshop**, Florence, Italy.
Blackbox NLP Panel Discussion
- [Outreach](#)
- March 2020 **!!Con West**, Santa Cruz, CA.
Get Hooked on Pytorch Hooks!
- Sept 2019 **Challenging the Work Society**, London, UK.
Carbon AI and the Concentration of Computational Work (jointly presented with Kate McCurdy)
- Feb 2019 **The Stand Edinburgh Comedy Club**, Edinburgh, UK.
Paying the Panopticon (standup comedy)
- Oct 2012 **Carnegie Mellon TechNights**, Pittsburgh, PA.
ML and NLP Guest Lectures (program for middle school girls)
- [Other invited talks](#)
- Jan 2023 **University of Copenhagen**, *Copenhagen, Denmark (Remote)*.
- Oct 2022 **Neural Sequence Models Theory Discord**, *Remote*.
- July 2022 **Oracle**, *Boston, MA (Remote)*.
- June 2022 **Stanford University**, *Palo Alto, CA (Remote)*.
- June 2022 **UC Irvine**, *Irvine, CA*.
- June 2022 **University of Southern California - Information Sciences Institute**, *Irvine, CA*.
- March 2022 **Neural Sequence Models Theory Discord**, *Remote*.
- Feb 2022 **University College London**, *London, UK*.
- Nov 2020 **UC Berkeley**, *Berkeley, CA (Remote)*.
- Oct 2020 **NLP With Friends**, *Remote*.
- June 2020 **New York University**, *New York, NY (Remote)*.
- May 2020 **Brown University**, *Providence, RI (Remote)*.
- Sept 2019 **Element AI**, *London, UK*.
- Aug 2019 **Allen Institute for AI**, *Seattle, WA*.
- May 2019 **City University of New York**, *New York, NY*.
- Jan 2019 **University of Amsterdam**, *Amsterdam, Netherlands*.

Service

[International](#)

- 2022 **Organizer**, *Analyzing and interpreting neural networks for NLP (BlackboxNLP) at EMNLP*.
- 2022 **Session Chair (Interpretability)**, *ACL*.
- 2021 **Organizer**, *6th Workshop on Representation Learning for NLP (RepL4NLP) at ACL*.
- 2014 **Student social co-chair**, *ACL*.

[Diversity, Equity, & Inclusion](#)

- 2020 **Panelist - D&I Session: Inclusivity in Conferences**, *NAACL*.
- 2020 **Accessibility Subcommittee**, *Association for Computational Linguistics (ACL)*.
- 2019-2020 **Disability representative**, *University of Edinburgh Staff Pride Network*.

[Local service and outreach](#)

- 2019-2020 **Radical Utopian Sci-Fi Enclave (RUSE)**, *University of Edinburgh*.
Co-organized monthly department sci-fi book club.

- 2017–2019 **ILCC Cookie Czar**, *University of Edinburgh*.
Organized weekly department social event.
- 2018–2019 **Edinburgh ML4NLP Reading Group**, *University of Edinburgh*.
Organized and led weekly reading group.
- 2012–2014 **North American Computational Linguistics Olympiad**, *Johns Hopkins University and Carnegie Mellon University*.
Tested puzzles and organized local high school competition.

Refereeing

Area Chair

- 2021–Present **Empirical Methods in Natural Language Processing (EMNLP)**, 2021, 2022.
- 2022 **Asian Association for Computational Linguistics (AACL)**, 2022.

Conference Reviewing

- 2021–Present **International Conference on Learning Representations (ICLR)**, 2021, 2022.
- 2017–Present **Neural Information Processing Systems (NeurIPS)**, 2017 (secondary), 2021, 2022.
- 2021–Present **Association for Computational Linguistics (ACL) Rolling Review**.
- 2019–2021 **Association for Computational Linguistics (ACL)**, 2019, 2020, 2021.
- 2019–2021 **North American Association for Computational Linguistics (NAACL)**, 2019, 2021.
- 2021 **European Association for Computational Linguistics (EACL)**.
- 2020 **International Conference on Machine Learning (ICML)**.
- 2017–2020 **Empirical Methods in Natural Language Processing (EMNLP)**, 2017, 2018, 2019, 2020.
- Workshop and Competition Reviewing**
- 2022 **Inverse Scaling Prize**.
- 2021 **I Can't Believe It's Not Better (NeurIPS Negative Results Workshop)**, *NeurIPS*.
- 2019–2021 **Analyzing and interpreting neural networks for NLP (BlackboxNLP)**, *ACL*, 2019, 2021.
- 2020 **The SIGNLL Conference on Computational Natural Language Learning (CoNLL)**, *ACL*.
- 2018–2020 **Widening NLP (WiNLP)**, *ACL*, 2018, 2019, 2020.
- 2017–2019 **Learning With Limited Data Workshop (LLD)**, *ICLR*, 2017, 2019.
- 2017 **Women in Machine Learning**, *NeurIPS*.
- 2017 **Representation Evaluation for NLP (RepL4NLP)**, *ACL*.