Sabrina J. **Mielke**

SENIOR ML SCIENTIST @ GENENTECH | NEW YORK, NY, USA

✉ sjm@sjmielke.com | 🏠 sjmielke.com | sjmielke | sjmielke | Google Scholar | Semantic Scholar

Education

Ph. D. Computer Science THE JOHNS HOPKINS UNIVERSITY

Baltimore, MD, USA

09/2017 — 10/2023

- Affiliated with the Center for Language and Speech Processing (CLSP), advised by Prof. Jason Eisner
- Thesis: “Building and Evaluating Open-Vocabulary Language Models”
- Graduate courses: “Probability Theory I”, “Bayesian Statistics”, “Natural Language Processing”, “ML: Optimization”, “ML: Linguistic & Sequence Modeling”, “Mathematics of Deep Learning”, “Parallel Programming”, “SW Testing & Debugging”, “Preparation for University Teaching”, “Causal Inference”
- JHU CER “Teaching Academy” certification (pass/fail class “Preparation for University Teaching,” numerous workshops and brown bags on teaching, developing and teaching a course)

M. Sc. Computer Science TU DRESDEN

Dresden, Germany

10/2015 — 07/2017

- Graduate courses: “Machine Learning I”, “Computer Vision I”, “Seminar Natural Language Processing”, “Computer Graphics I”, “Scientific Visualization”, “Foundational Research: lectures and project”
- Thesis: “Soft matching of terminals for syntactic parsing”, supervisor: Prof. Heiko Vogler

B. Sc. Computer Science TU DRESDEN

Dresden, Germany

10/2012 — 08/2015

- Graduate level coursework: “Machine Translation of Natural Languages”, “Compiler Construction”, “Lab/Project: Haskell for NLP”, “C++ Programming for Computer Graphics”
- Thesis: “Extracting and Binarizing probabilistic linear context-free rewriting systems”, supervisor: Prof. Heiko Vogler

Work Experience

Applied Research on LLMs for drug discovery

New York, NY, USA

SENIOR ML SCIENTIST AT GENENTECH *2025-10 — ongoing*

- Biotech applications of LLMs in drug discovery (Genentech/Roche): leading key machine learning and scientific decisions for high-impact initiatives while owning the team’s generalized evaluation framework for Foundation Models.

MACHINE LEARNING ENGINEER, LARGE LANGUAGE MODELS AT GENENTECH *2024-06 — 2025-10*

- work as part of the Prescient Design Accelerator/MLDD (machine learning drug discovery) unit in gCS (Genentech Computational Sciences), part of gRED (Genentech Research and Early Development) in Genentech, which is a subsidiary of Roche’s Pharma division)
- implementing retrieval-augmented generation systems to survey tens of millions of scientific publications for the purpose of drug discovery, maintaining databases of hundreds of millions of snippets of knowledge
- collaborating with biomedical teams in deciding how to bring large language models into applications from foundations to clinical outcomes
- driving evaluation efforts, combining these two strands to produce results for the entire team to be presented to leadership
- part of a ~10-person team solely responsible for one of Roche’s company-wide goals in 2024, achieving the goal fully with our presented results, resulting in company-wide extra bonuses; recognized for this impact with a Roche Impact Award (individuialized to team members)

Applied Research on GenAI

New York, NY, USA

SENIOR AI RESEARCH ENGINEER AT ALPHASENSE

2023-11 — 2024-06

- work on AlphaSense Assistant, an AI-powered chatbot designed to answer business questions
- primarily working around the task of "citation," relating chatbot generations to underlying data in a model-agnostic way to aid verifiability of results by platform and user
- investigating various signals to aid in this task: metadata-based, neural, and non-neural statistical signals
- extensive visualization efforts for citations and signals, communicating results to leadership to drive key decisions in the rollout of the AI-powered Assistant model

Teaching assistant

Baltimore, MD, USA

TEACHING ASSISTANT AT THE JOHNS HOPKINS UNIVERSITY

2018-09 — 2023-10

- teaching, administration, and grading of undergraduate and graduate level "Natural Language Processing" and "Machine Learning: Linguistic & Sequence Processing" (both under Prof. Jason Eisner), as well as "Artificial Intelligence" (under multiple lecturers)

Working at co:here: Frameworks for Large Language Models

New York City, NY, USA (remote)

INTERN AT COHERE AI

2022-06-13 — 2022-08-26

- research and engineering for the Frameworks team, advised by Joanna Yoo and Kuba Perlin

Teaching a full "Artificial Intelligence" class at JHU

Baltimore, MD, USA

TEACHING AS A PHD STUDENT AT THE JOHNS HOPKINS UNIVERSITY IN FALL 2020 AND FALL 2021

2020-08 — 2021-12

- developing a combined undergrad/grad-level class on "Artificial Intelligence" based on previous years (including TAing the previous iteration in Spring 2020) and the Berkeley CS188 class
- focus on adding lectures and sections on AI ethics that were missing before, all the way to current-day research
- organizing guest lectures on causal inference, knowledge bases, embedding methods, neuro-symbolic hybrids
- making an all-virtual class (as the university only allowed in-person for smaller classes) still engaging by experimenting with asynchronous lectures and a flipped classroom approach mixed with lectures
- Fall 2020: ~50 undergraduate students, all-virtual, half-asynchronous across many timezones
- Fall 2021: ~40 undergraduate + ~30 graduate students, all-virtual, synchronous

Working at Hugging Face: Tokenization in Language Modeling

New York City, NY, USA

PART-TIME INTERN AT HUGGING FACE INC.

2021-05-31 — 2021-08-27

- conducting thesis research advised by Alexander Rush and Yacine Jernite
- part of the Tokenization Working Group of the globally cross-institutional BigScience project

Internship at Facebook AI Research (FAIR): Metacognition and calibration of chatbots

New York City, NY

INTERN AT FACEBOOK INC. (FULL- AND PART-TIME)

2020-05-26 — 2020-08-28, part-time until 2020-11-13

- analyzing state-of-the-art chatbots with Emily Dinan, Y-Lan Boureau, and Arthur Szlam
- developing annotation schemes for certainty and correctness in chat and collecting data
- training a "metacognition" probe to anticipate incorrect outputs and facilitate proper calibration
- paper "Linguistic calibration through metacognition: aligning dialogue agent responses with expected correctness" under review at TACL, preprint at `https://arxiv.org/abs/2012.14983`

Research assistant

Baltimore, MD, USA

RESEARCH ASSISTANT AT THE JOHNS HOPKINS UNIVERSITY

2017-09 — 2018-08, 2019-08 — 2020-05

- research assistant funded by NSF grant #1718846 ("Linguistic Structure in Neural Sequence Models")

Internship at Google: Transliteration of South Asian languages

New York City, NY, USA

intern at Google LLC *2019-05-20 — 2019-08-23*

- working on transliteration from Latin orthography into Indic scripts with Brian Roark
- building a modern, extensible transliteration system using TensorFlow 2.0
- evaluating the feasibility of using pronunciation data and cross-language multi-task approaches
- giving research talks and a final presentation
- paper "Processing South Asian Languages Written in the Latin Script: the Dakshina Dataset" at LREC 2020

Student teaching assistant

Dresden, Germany

student assistant at the TU Dresden, various institutes *2013-10-15 — 2017-07-21*

- "Programming" (2014, 2017), "Operating systems" (2016/17), "Formal systems" (2016/17), "Algorithms and data structures" (2013/14, 2014/15, 2015/16), "Introductory lab: RoboLab" (2013/14), "Computer architecture I" (2014/15, 2015/16), Seminar group mentor (2014/15)

Student research assistant

Dresden, Germany

student assistant at the TU Dresden, Institute of Theoretical Computer Science *2015-05-01 — 2016-09-30*

- analyzing a count-based state-merging algorithm (2016-08-22 — 2016-09-30): used an algorithm for generalizing regular tree grammars (RTGs) to recover grammars from a corpus sampled from handwritten RTGs and study its influence on parsing performance for natural language corpora
- working on machine translation software (2016-01-01 — 2016-04-30): implementation work (Haskell, Java) for the machine translation software Vanda Toolkit and the GUI Vanda Studio
- creating lecture materials (2015-05-01 — 2015-10-31): writing, figure planning and creation of both lecture notes and lecture slides for the lecture "Algorithms and data structures", held by Prof. Heiko Vogler

Internship: Low-resource machine translation at USC/ISI

Los Angeles, CA, USA

intern at the University of Southern California Information Sciences Institute (USC/ISI) *2016-05-01 — 2016-07-31*

- working for the DARPA LORELEI efforts under Prof. Kevin Knight and Prof. Daniel Marcu
- helping our team (USC/ISI, USC, UoND, RPI) clearly win the first two of three checkpoints
- making significant contributions to final result, handling dictionary preprocessing and out-of-vocabulary words
- holding an NL seminar talk at ISI: "Let's not be clever: simple pre- and post-processing tricks in machine translation"

Internship: Training a dependency parser on a noisy multi-paradigm Kannada treebank

Manipal, India

Intern at the Manipal Institute of Technology (constituent of Manipal University) *2015-08-10 — 2015-09-20*

- developing procedures to clean and convert a noisily annotated treebank of the Kannada language into pure dependency information, used to experiment with existing off-the-shelf parsers
- significantly shaped the ongoing treebank construction with feedback and error reports, summarized in a report and a talk

Work on a particle simulation library

Dresden, Germany

Student assistant at the TU Dresden, Instutite of Computer Engineering *2015-04-15 — 2015-07-14*

- working on a C++ reimplementation of the PPM library under Prof. Dr. Jerónimo Castrillón-Mazo and Prof. Dr. Ivo F. Sbalzarini
- implementing efficient multi-layer grids (AR-lists) yielding a 5x speedup for adaptive resolution simulations

Publications

double-blind preprints under review, most recent submission to ICML 2025

written with various co-authors.

"BLOOM: A 176B-Parameter Open-Access Multilingual Language Model" ArXiv *arXiv preprint*

BigScience Workshop (390 authors, detailed contributions in paper).

"Reducing conversational agents' overconfidence through linguistic calibration" ArXiv MIT Press *TACL (presented at NAACL 2022)*

Mielke, Szlam, Boureau, and Dinan.

"UniMorph 4.0: Universal Morphology" arXiv ACLweb *LREC 2022*

Batsuren, Goldman, Khalifa, Habash, Kieraś, Bella, Leonard, Nicolai, Gorman, Ate, Ryskina, **Mielke**, Budianskaya, El-Khaissi, Pimentel, Gasser, Lane, Raj, Coler, Samame, Camaiteri, Sagot, Rojas, Francis, Oncevay, Bautista, Villegas, Hennigen, Ek, Guriel, Dirix, Bernardy, Scherbakov, Bayyr-ool, Anastasopoulos, Zariquiey, Sheifer, Ganieva, Cruz, Karahóğa, Markantonatou, Pavlidis, Plugaryov, Klyachko, Salehi, Angulo, Baxi, Krizhanovsky, Krizhanovskaya, Salesky, Vania, Ivanova, White, Maudslay, Valvoda, Zmigrod, Czarnowska, Nikkarinen, Salchak, Bhatt, Straughn, Liu, Washington, Pinter, Ataman, Wolinski, Suhardijanto, Yablonskaya, Stoehr, Dolatian, Nuriah, Ratan, Tyers, Ponti, Aiton, Arora, Hatcher, Kumar, Young, Rodionova, Yemelina, Andrushko, Marchenko, Mashkovtseva, Serova, Prud'hommeaux, Nepomniashchaya, Giunchiglia, Chodroff, Hulden, Silfverberg, McCarthy, Yarowsky, Cotterell, Tsarfaty, and Vylomova.

"Between words and characters: A Brief History of Open-Vocabulary Modeling and Tokenization in NLP" arXiv *arXiv preprint*

Mielke, Alyafeai, Salesky, Raffel, Dey, Gallé, Raja, Si, Lee, Sagot, and Tan.

"SIGMORPHON 2021 Shared Task on Morphological Reinflection: Generalization Across Languages" ACLweb *SIGMORPHON 2021*

Pimentel*, Ryskina*, **Mielke**, Wu, Chodroff, Leonard, Nicolai, Ate, Khalifa, Habash, El-Khaissi, Goldman, Gasser, Lane, Coler, Oncevay, Samame, Villegas, Ek, Bernardy, Shcherbakov, Bayyr-ool, Sheifer, Ganieva, Plugaryov, Klyachko, Salehi, Krizhanovsky, Krizhanovsky, Vania, Ivanova, Salchak, Straughn, Liu, Washington, Ataman, Kieraś, Woliński, Suhardijanto, Stoehr, Nuriah, Ratan, Tyers, Ponti, Aiton, Hatcher, Prud'hommeaux, Kumar, Hulden, Barta, Lakatos, Szolnok, Ács, Raj, Yarowsky, Cotterell, Ambridge, and Vylomova.

"SIGTYP 2021 Shared Task: Robust Spoken Language Identification" arXiv ACLweb *SIGTYP 2021*

Salesky*, Abdullah*, **Mielke***, Klyachko, Serikov, Ponti, Kumar, Cotterell, and Vylomova.

"Tired of Topic Models? Clusters of Pretrained Word Embeddings Make for Fast and Good Topics too!" arXiv ACLweb *EMNLP 2020*

Sia, Dalmia, and **Mielke**.

"SIGTYP 2020 Shared Task 0: Prediction of Typological Features arXiv ACLweb *SIGTYP 2020*

Bjerva, Salesky, **Mielke**, Chaudhary, Celano, Ponti, Vylomova, Cotterell, and Augenstein.

"SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection arXiv ACLweb *SIGMORPHON 2020*

Vylomova, White, Salesky, **Mielke**, Wu, Ponti, Maudslay, Zmigrod, Valvoda, Toldova, Tyers, Klyachko, Yegorov, Krizhanovsky, Czarnowska, Nikkarinen, Krizhanovsky, Pimentel, Hennigen, Kirov, Nicolai, Williams, Anastasopoulos, Cruz, Chodroff, Cotterell, Silfverberg, and Hulden.

"It's Easier to Translate out of English than into it: Measuring Neural Translation Difficulty by Cross-Mutual Information" *ACL 2020*

Bugliarello, **Mielke**, Anastasopoulos, Cotterell, and Okazaki. arXiv ACLweb

"Processing South Asian Languages Written in the Latin Script: the Dakshina Dataset" arXiv ACLweb *LREC 2020*

Roark, Wolf-Sonkin, Kirov, **Mielke**, Johny, Demirsahin, and Hall.

"UniMorph 3.0: Universal Morphology" ACLweb *LREC 2020*

McCarthy, Kirov, Grella, Nidhi, Xia, Gorman, Vylomova, **Mielke**, Nicolai, Silfverberg, Arkhangelskiy, Krizhanovsky, Krizhanovsky, Klyachko, Sorokin, Mansfield, Ernštreits, Pinter, Jacobs, Cotterell, Hulden, and Yarowsky.

"The SIGMORPHON 2019 Shared Task: Morphological Analysis in Context & Cross-Lingual Transfer for Inflection" *SIGMORPHON 2019*

McCarthy, Vylomova, Wu, Malaviya, Wolf-Sonkin, Nicolai, Silfverberg, **Mielke**, Heinz, Cotterell, and Hulden. arXiv ACLweb

"What Kind of Language Is Hard to Language-Model?" arXiv ACLweb *ACL 2019*

Mielke, Cotterell, Gorman, Roark, and Eisner.

"Counterfactual Data Augmentation for Mitigating Gender Stereotypes in Languages with Rich Morphology" arXiv ACLweb *ACL 2019*

Zmigrod, **Mielke**, Cotterell, and Wallach.

"Spell Once, Summon Anywhere: A Two-Level Open-Vocabulary Language Model" arXiv AAAI page *arXiv 2018 / AAAI 2019*

Mielke and Eisner.

"The CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection" arXiv ACLweb *CoNLL-SIGMORPHON 2018*

Cotterell, Kirov, Sylak-Glassman, Walther, Vylomova, McCarthy, Kann, **Mielke**, Nicolai, Silfverberg, Yarowsky, Eisner, and Hulden.

"A Structured Variational Autoencoder for Contextual Morphological Inflection" arXiv ACLweb *ACL 2018*

Wolf-Sonkin*, Naradowsky*, **Mielke***, and Cotterell*.

"Unsupervised Disambiguation of Syncretism in Inflected Lexicons" arXiv ACLweb *NAACL 2018*

Cotterell, Kirov, **Mielke**, and Eisner.

"Are All Languages Equally Hard to Language-Model?" arXiv ACLweb *NAACL 2018*

Cotterell, **Mielke**, Eisner, and Roark.

"UniMorph 2.0: Universal Morphology" LREC *LREC 2018*

Kirov, Cotterell, Sylak-Glassman, Walther, Vylomova, Xia, Faruqui, **Mielke**, McCarthy, Kübler, Yarowsky, Eisner, and Hulden.

"Incident-driven machine translation and name tagging for low-resource languages" Springer *Machine Translation (Springer Journal)*

Hermjakob, Li, Marcu, May, **Mielke**, Pourdamghani, Pust, Shi, Knight, Levinboim, Murray, Chiang, Zhang, Pan, Lu, Lin, and Ji.

Honors

2021 Rising Star in EECS

Cambridge, MA, USA (virtual)

SELECTION AS A PARTICIPANT TO THE 2021 RISING STARS IN EECS WORKSHOP AT MIT *2021*

- selection as a "rising star," participating in this "intensive workshop for graduate students and postdocs with historically marginalized or underrepresented genders who are interested in pursuing academic careers in electrical engineering, computer science, and artificial intelligence and decision-making"

Dean Robert H. Roy Fellowship

Baltimore, MD, USA

FELLOWSHIP GIVEN BY THE CS DEPARTMENT OF THE JOHNS HOPKINS UNIVERSITY *2017/18*

- a fellowship "given to a PhD student whom the CS Department believes has exceptional intellectual promise"

"Deutschlandstipendium"

Dresden, Germany

PUBLIC-PRIVATE SCHOLARSHIP, GIVEN BY THE TU DRESDEN AND A PRIVATE PARTNER *2013/14, 2014/15, 2015/16, 2016/17*

- with Deutsche Telekom (2013/14, 2014/15, and 2015/16) and IBM (2016/17)
- awarded for 300 "promising" students (roughly 1% of all students) of the university

Talks

"FROM STATEFUL CODE TO PURIFIED JAX: HOW TO BUILD YOUR NEURAL NET FRAMEWORK" *2023-07-01*
invited talk at the Diffusers community week organized by HuggingFace, ~50 attendees + ~2000 post-hoc recording views shared with two other talks

"FAIR COMPARISONS FOR GENERATIVE LANGUAGE AND TRANSLATION MODELS—WITH A BIT OF INFORMATION THEORY" *2022-01-26*
invited talk at the AI Suisse meetup, hosted by Steffen Konrath, ~30 attendees

"DAY 4 LECTURE: THE JUMP TO NLP" *2022-01-12*
invited lecture at the NYU AI School 2022

"FAIR COMPARISONS AND FUNDAMENTAL IDEAS FOR OPEN-VOCABULARY GENERATIVE LANGUAGE AND TRANSLATION MODELS" *2021-08-12*
invited talk at the ISI seminar series, hosted by Jonathan May, ~10 attendees + ~60 post-hoc recording views

"THE MODERN NLP RESEARCHER'S TOOLBOX" (WITH SONAL JOSHI) *2021-08-01*
a brief tutorial given at the ACL 2021 Widening NLP satellite event

"LINGUISTIC CALIBRATION THROUGH METACOGNITION FOR CHATBOTS" *2021-07-21*
invited talk at the ICML 2021 Zeitgeist in NLP social, hosted by Katharina Beckh and Vanessa Faber, ~30 attendees

"FAIR COMPARISONS FOR GENERATIVE LANGUAGE MODELS—WITH A BIT OF INFORMATION THEORY" *2021-07-15*
invited talk at the virtual SIGTYP lecture series, hosted by Ekatarina Vylomova, ~10 attendees + ~120 post-hoc recording views

"FROM STATEFUL CODE TO PURIFIED JAX: HOW TO BUILD YOUR NEURAL NET FRAMEWORK" *2021-07-01*
invited talk at the Flax/JAX community week organized by Google and HuggingFace, ~100 attendees + ~4800 post-hoc recording views shared with three other talks

"WRITING EXTENDED ABSTRACTS FOR NLP CONFERENCES" (WITH VAGRANT GAUTAM) *2021-04-19*
a brief tutorial given at the EACL 2021 Widening NLP satellite event, ~500 views on the post-hoc released recording

"FAIR COMPARISONS FOR GENERATIVE LANGUAGE MODELS—WITH A BIT OF INFORMATION THEORY" *2020-09-02*
invited talk at the virtual seminar series "NLP with Friends," hosted by Liz Salesky, ~250 attendees + ~500 post-hoc recording views

"MEASURING PERFORMANCE OF PROBABILISTIC GENERATION MODELS WITH A BIT OF INFORMATION THEORY" *2020-05-01*
CLSP plenary talk at JHU, ~30 attendees

"OPEN-VOCABULARY LANGUAGE MODELING IN 69 LANGUAGES" *2019-08-29*
CLSP plenary talk at JHU, ~30 attendees

"LANGUAGE MODELING: FAIR COMPARISONS AND COUNTERFACTUAL DATA AUGMENTATION" *2019-07-31*
invited talk at the Graduate School and University Center of the City University of New York (CUNY), hosted by Kyle Gorman, ~10 attendees

Selected academic blogposts and software

Parallax: a sketch for a JAX/PyTorch-hybrid neural network framework

OPEN-SOURCE, DEVELOPED WITH SASHA RUSH, HTTPS://GITHUB.COM/SRUSH/PARALLAX, 152 STARS *2020-05*

- implementing a prototype of Prof. Sasha Rush's idea of a pure module system for JAX
- main ideas: make param modules immutable trees, replace all imperative style coding and init, and avoid tracking state for most applications by first distributing seeds / globals through tree

From PyTorch to JAX: towards neural net frameworks that purify stateful code

BLOGPOST, HTTPS://SJMIELKE.COM/JAX-PURIFY.HTM, 28K VIEWS *2020-03-09*

- "Moving from object-oriented PyTorch- or TF2-code with tape-based backprop to JAX isn't easy—and while running grad() on numpy-oneliners is cool and all, you do wonder... how do I build actual big neural nets? Maybe you decided to look at libraries like flax, trax, or haiku [...] but what is it that actually happens there? What's the route from these tiny numpy functions to training big hierarchical neural nets?"

Can you compare perplexity across different segmentations?

BLOGPOST, HTTPS://SJMIELKE.COM/COMPARING-PERPLEXITIES.HTM, 4K VIEWS *2019-04-23*

- "Can you compare perplexity across different segmentations? Short answer: Not immediately. Long answer: Yes, as long as you have equal denominators and the same support!"

Language diversity in ACL 2004 - 2016

BLOGPOST, HTTPS://SJMIELKE.COM/ACL-LANGUAGE-DIVERSITY.HTM, 1K VIEWS *2016-12-22*

- "Natural Language Processing == English Language Processing? Let's look at the languages that ACL long papers evaluated on in the last few years. Is is getting better or worse? Or maybe just a little bit of both?"

Describing discontinuous constituents with LCFRS

BLOGPOST, HTTPS://SJMIELKE.COM/DESCRIBING-DISCONTINUOUS-CONSTITUENTS-WITH-LCFRS.HTM, 1K VIEWS *2016-10-21*

- "About finding structure in language. Even weird things. Especially weird things."

Service

Area Chair

AREA CHAIR AT NLP CONFERENCES

- EACL (2023)
- ACL Rolling Review (February 2025)
- COLM (Conference on Language Modeling; 2025)

Reviewing

Program Committee member for various NLP and AI conferences and workshops and the NLE journal

- ACL Rolling Review (September 2021, October 2021, November 2021, January 2022, March 2022)
- NAACL (2019, 2021)
- ACL (2019, 2020, 2021, outstanding reviewer)
- EMNLP (2019, 2021)
- CoNLL (2019, 2020, 2021)
- LREC (2020)
- ICML (2020, top 33% reviewer, 2021)
- NeurIPS (2020, 2021)
- AACL (2020)
- ICLR (2021, top 10% reviewer, 2022)
- COLM (2024)
- Natural Language Engineering (Journal, *Cambridge University Press*)
- Workshop on Structured Prediction for NLP (NAACL 2019, EMNLP 2020, ACL 2021)
- Workshop on Representation Learning for NLP (ACL 2019, ACL 2021)
- Workshop on Widening NLP (ACL 2019, ACL 2020)
- Workshop on Neural Generation and Translation (EMNLP 2019, ACL 2020)
- Workshop on Human And Machine in-the-Loop Evaluation and Learning Strategies (NeurIPS 2020)
- Student Resarch Workshop at NAACL 2021
- Secondary reviewer for EMNLP 2018 and NeurIPS 2018

Organizing committee

Chair/organizer for workshops, shared tasks, and WiNLP

- SIGMORPHON Shared Task co-organization 2018, 2019, 2020, 2021
- SIGTYP Shared Task co-organization 2020, 2021
- SIGTYP Workshop co-organization 2021
- Widening NLP (WiNLP) chair, organizing committee 2020–2022

Other service within the ACL community

Outreach, panels, mentoring

- panelist for the Zeitgeist in NLP social at ICML 2021
- talk and Q&A for the inaugural event of the Queer in AI Undergraduate Mentorship Series at NAACL 2021
- panelist for a session on PhD applications at ACL 2020
- co-organizing/hosting a mentoring session at ACL 2020
- mentoring at the WiNLP satellite event at AACL 2020

University service at JHU

Baltimore, MD

Service on various JHU/CLSP committees

- organizing visit weekend as member of the CLSP student recruitment commitee (2018, 2019)
- leading the rewriting of internal CLSP computing documentation (2020)
- admissions committee (2020, 2021)

Skills

Languages	German (native), English (fluent, CEFR level C1, TOEFL 119, GRE 161/170/5.0) Latin (7 years of study, “Latinum”/Latin proficiency certificate) Arabic, Japanese, Chinese (script / basic vocabulary, phrases, and grammar)
Programming experience	Many years of Python3, a few years in Haskell, and a few fewer in Java, Rust, C and C++11
The usual tooling	Distributed version control (git), shell scripting (bash, sed, …), VSCode, GIMP, Inkscape
Comfortable frameworks	numpy, PyTorch, JAX, STAN, Flask, React
Fancy type and plots	LaTeX/BibTeX, TikZ/PGF, beamer, matplotlib, Altair