

Steven Y. Feng

Stanford University
PhD in Computer Science

✉ syfeng@stanford.edu
🌐 styfeng.github.io
🌐 [linkedin.com/in/steven-feng](https://www.linkedin.com/in/steven-feng)
🐙 <https://github.com/styfeng>

EDUCATION

- **Stanford University** Stanford, United States
PhD in Computer Science 2022-Present
- **Carnegie Mellon University (CMU)** Pittsburgh, United States
Master of Language Technologies (MLT) – Fully-funded research master's in NLP 2020-2022
- **University of Waterloo** Waterloo, Canada
Bachelor of Mathematics, Statistics Major, Computer Science Minor 2015-2020
- **Wilfrid Laurier University** Waterloo, Canada
Bachelor of Business Administration, Finance Concentration 2015-2020

PEER-REVIEWED PUBLICATIONS AND CONFERENCE PROCEEDINGS

- **Steven Y. Feng**, Vivek Khetan, Bogdan Sacaleanu, Anatole Gershman, and Eduard Hovy. “CHARD: Clinical Health-Aware Reasoning Across Dimensions for Text Generation Models.” Accepted to *European Chapter of the Association for Computational Linguistics (EACL) 2023* [Long Paper]. Arxiv: <https://arxiv.org/abs/2210.04191>
- Sedrick Scott Keh, **Steven Y. Feng***, Varun Gangal*, Malihe Alikhani, and Eduard Hovy. “PANCETTA: Phoneme Aware Neural Completion to Elicit Tongue Twisters Automatically.” Accepted to *European Chapter of the Association for Computational Linguistics (EACL) 2023* [Long Paper]. Arxiv: <https://arxiv.org/abs/2209.06275> (* Equal Contribution)
- Sedrick Scott Keh, Kevin Lu, Varun Gangal*, **Steven Y. Feng***, Harsh Jhamtani, Malihe Alikhani, and Eduard Hovy. “PINEAPPLE: Personifying INanimate Entities by Acquiring Parallel Personification data for Learning Enhanced generation.” Proceedings of *International Conference on Computational Linguistics (COLING) 2022* [Long Paper]. <https://aclanthology.org/2022.coling-1.547/> (* Equal Contribution)
- **Steven Y. Feng**, Kevin Lu, Zhuofu Tao, Malihe Alikhani, Teruko Mitamura, Eduard Hovy, and Varun Gangal. “Retrieve, Caption, Generate: Visual Grounding for Enhancing Commonsense in Text Generation Models.” Proceedings of *AAAI Conference on Artificial Intelligence 2022* (Acceptance rate: 15%) and *AKBC 2021 Commonsense Reasoning and Knowledge Bases (CSKB) Workshop*. <https://ojs.aaai.org/index.php/AAAI/article/view/21306>
- Varun Gangal*, **Steven Y. Feng***, Malihe Alikhani, Teruko Mitamura, and Eduard Hovy. “NAREOR: The Narrative Reordering Problem.” Proceedings of *AAAI Conference on Artificial Intelligence 2022* (Acceptance rate: 15%). <https://ojs.aaai.org/index.php/AAAI/article/view/21309> (* Equal Contribution)
- **Steven Y. Feng**, Jessica Huynh, Chaitanya Prasad Narisetty, Eduard Hovy, and Varun Gangal. “SAPPHIRE: Approaches for Enhanced Concept-to-Text Generation.” Proceedings of *International Conference on Natural Language Generation (INLG) 2021* [**Best Long Paper**]. <https://aclanthology.org/2021.inlg-1.21/>
- **Steven Y. Feng***, Varun Gangal*, Jason Wei, Sarath Chandar, Soroush Vosoughi, Teruko Mitamura, and Eduard Hovy. “A Survey of Data Augmentation Approaches for NLP.” Proceedings of *Association for Computational Linguistics (ACL) 2021 Findings* [Long Paper]. <https://aclanthology.org/2021.findings-acl.84/> (* Equal Contribution)
- **Steven Y. Feng***, Varun Gangal*, Dongyeop Kang, Teruko Mitamura, and Eduard Hovy. “GenAug: Data Augmentation for Finetuning Text Generators.” Proceedings of *EMNLP 2020 Deep Learning Inside Out (DeeLIO) Workshop* [Long Paper]. <https://aclanthology.org/2020.deelio-1.4/> (* Equal Contribution)

- Aaron W. Li, Veronica Jiang*, **Steven Y. Feng***, Julia Sprague, Wei Zhou, and Jesse Hoey. “ALOHA: Artificial Learning of Human Attributes for Dialogue Agents.” Proceedings of *AAAI Conference on Artificial Intelligence 2020* (Acceptance rate: **20.6%**) [Oral]. <https://arxiv.org/abs/1910.08293> (* Equal Contribution)
- **Steven Y. Feng***, Aaron W. Li*, and Jesse Hoey. “Keep Calm and Switch On! Preserving Sentiment and Fluency in Semantic Text Exchange.” Proceedings of *Empirical Methods in Natural Language Processing (EMNLP) 2019* (Acceptance rate: **23.8%**) [Long Paper]. <https://www.aclweb.org/anthology/D19-1272/> (* Equal Contribution)

WORK EXPERIENCE

- **Amazon** Sunnyvale, CA, USA
Applied Scientist Intern June – Sept. 2023
 - Working with the Alexa AI team (particularly LLM group) on projects to enhance the reasoning capabilities of LLMs.
 - Investigating ideas to improve and generalize chain-of-thought (CoT) reasoning for LLMs.
- **CBT Associates / MindBeacon** Toronto, Canada
Data Scientist Jan. – Apr. 2018
 - Built an e-therapist chatbot using Java and NLP methods that increased daily e-therapist capacity by approximately 50%.
 - Designed and implemented an automated client assessment scoring system using Python and SQL for customer segmentation.

AWARDS AND SCHOLARSHIPS

- **Stability Fellow** 2023
Awardeed by StabilityAI for research accomplishments and potential, particularly in image generation.
- **NSERC Postgraduate Scholarships – Doctoral (PGS D)** 2022
Prestigious NSF equivalent fellowship for the top Canadian researchers going into a PhD program.
- **NSERC Canada Graduate Scholarships - Doctoral (CGS D) [Declined]** 2022
- **Best Long Paper Award - International Conference on Natural Language Generation (INLG) 2021** 2021
- **CMU Graduate Research Fellowship** 2020-2022
- **Jessie W.H. Zou Memorial Award for Excellence in Undergraduate Research (2020) - Honorable Mention** 2020
Among the top three undergraduate researchers in the Faculty of Mathematics at the University of Waterloo.
Article: <https://cs.uwaterloo.ca/news/steven-feng-honourable-mention-2020-jessie-wh-zou-memorial-award>
- **CRA Outstanding Undergraduate Researcher Award (2020) – Honorable Mention** 2020
Award for the top North American undergraduate researchers in computing.
Article: <https://cs.uwaterloo.ca/news/steven-feng-shannon-veitch-honorable-mention-cra-outstanding-undergraduate-researcher>
- **NSERC Alexander Graham Bell Canada Graduate Scholarship - Master's (CGS M) [Declined]** 2020
- **Vector Scholarship in Artificial Intelligence [Declined]** 2020
- **Ontario Graduate Scholarship and Queen Elizabeth II Graduate Scholarship in Science & Technology [Declined]** 2020
- **NSERC Undergraduate Student Research Award (USRA)** Jan. – Apr. 2019, May – Aug. 2019
- **President’s and Math Endowment Fund Research Awards** 2019 - 2020
- **EMNLP and AACL Student Scholarships** 2019 - 2020
- **University of Waterloo Alumni Scholarship** 2015 - 2020
- **University of Waterloo President's Scholarship of Distinction** 2015

TALKS AND INTERVIEWS

- “Language Models: Emergent Abilities, Trends, and Applications.” Invited talk for *University of Connecticut*. April 2023. Video [here](#).
- “A Survey of Data Augmentation Approaches for NLP”. Invited talk with **Varun Gangal** for *Google Research*. Aug. 2021. Video [here](#).
- Guest (with Dr. **Eduard Hovy**) on the [Data Exchange Podcast with Ben Lorica](#). July 2021. Video [here](#).
 - Eduard Hovy and I discuss data augmentation for NLP and challenges + future directions in NLP and machine learning research.
- “ALOHA: Artificial Learning of Human Attributes for Dialogue Agents.” Our AAAI 2020 work (accepted as oral).
 - *AAAI Conference on Artificial Intelligence 2020*. New York, USA. Feb. 2020. Oral and poster presentations.
 - *University of Waterloo Cheriton School of Computer Science AI Seminar*. Waterloo, Canada. Jan. 2020.
- Panelist: Research in Data Science. *University of Waterloo Data Science Club*. July 2021.
- “Undergrad Research & Grad School Preparation.” *University of Waterloo Data Science Club*. Nov. 2020.

TEACHING EXPERIENCE

- **Co-Instructor – CS 25: Transformers United (Stanford University)** *Jan. – Mar. 2023, Sept. – Dec. 2023*
Co-instructing this Transformers seminar course at Stanford. We feature in-depth discussion each week from exciting speakers about cutting-edge research in Transformers for a variety of Machine Learning areas/applications. Some speakers have included [Andrej Karpathy](#), [Jan Leike](#) (OpenAI), and [Jason Wei](#) (OpenAI, prev. Google Brain). Course webpage: <https://web.stanford.edu/class/cs25/>

MENTORSHIP AND ADVISING

- [Sedrick Scott Keh](#) - CMU Master's of Machine Learning (MSML), Class of 2022
Mentored several research projects on creative text generation. Two papers: a publication at COLING 2022 and another at EACL 2023.
- [Kevin Lu](#) – University of Waterloo Undergrad, Computer Science, Class of 2026
Mentored several research projects on controllable, creative, and visually-grounded text generation. Got his first two papers: publications at AAAI and COLING 2022, and an abstract at TADA which he presented in October 2021.
- [Zhuofu \(Derek\) Tao](#) - UCLA Ph.D. in Electrical Engineering, Class of 2025
Mentored a research project on controllable and visually-grounded text generation. Got his first NLP publication at AAAI 2022.
- [Jerry Huang](#), [Hongru Xiang](#), [Xintao Zhu](#), [Saidi Tang](#) - University of Waterloo Undergrads, Software Engineering, Class of 2022
Advised their software engineering capstone project on text simplification for ESL students.

PROFESSIONAL SERVICE AND VOLUNTEERING

- **Reviewer – Empirical Methods in Natural Language Processing (EMNLP) 2023** *July – Sept. 2023*
- **Reviewer – Association for Computational Linguistics (ACL) 2023** *Jan. – Apr. 2023*
- **Reviewer - Language Resources and Evaluation Conference (LREC) 2022** *Jan. – Apr. 2022*
- **Organizer – GEM Benchmark and Workshop** *July 2021 – July 2022*
Involved in the GEM benchmark and workshop for evaluation in NLG. <https://gem-benchmark.com/>
- **Lead Organizer – CtrlGen Workshop (NeurIPS 2021)** *Dec. 2020 – Dec. 2021*
Led organization of CtrlGen: Controllable Generative Modeling in Language and Vision workshop at NeurIPS 2021. Initiated idea and organization effort. In charge of planning & logistics. Main workshop host (recording: <https://neurips.cc/virtual/2021/workshop/21886>). Primary writer/creator of the workshop proposal and website: <https://ctrlgenworkshop.github.io/>
- **Reviewer – IEEE Transactions on Affective Computing (Journal)** *Aug. – Sept. 2021*
- **Reviewer – NeurIPS 2021 Workshop Proposals** *June - July 2021*

- **Reviewer – AAAI Conference on Artificial Intelligence 2021** Sept. – Nov. 2020
- **AAAI 2020 Student Volunteer** Feb. 2020
- **EMNLP 2019 Student Volunteer** Nov. 2019

RESEARCH EXPERIENCE

Stanford University

- PhD Graduate Research Assistant — Machine Learning, NLP, Computer Vision, Psychology, Cognitive Science Stanford, United States
Dr. **Michael C. Frank**, **Noah Goodman**, **Leonidas Guibas**, **Douwe Kiela** *Sept. 2022 - Present*
 - **BabyLM**: Investigating why babies can learn language and reason with much less data than LLMs. Trying to match SOTA performance while reducing orders of magnitude of training data and model size, e.g. using multimodal grounding, curricularization, knowledge injection, and external guidance. Will allow for greatly improved efficiency, open-source availability, alignment, and interpretability of LLMs.
 - **Vision-Language Models and Image Generation**: Working on projects to enhance the efficiency and controllability of text-to-image generation.

Carnegie Mellon University's Language Technologies Institute (LTI)

- Master's Graduate Research Assistant — Machine Learning and Natural Language Processing (NLP) Pittsburgh, United States
Dr. **Eduard Hovy**, **Malihe Alikhani**, **Teruko Mitamura**, **Graham Neubig** *Aug. 2020 – Aug. 2022*
 - **Commonsense Reasoning for NLG**: Devised effective approaches for generative commonsense reasoning and concept-to-text generation called SAPPHIRE. First-author paper published at INLG 2021 [**best long paper**]. Investigated visual-grounding for improving the commonsense reasoning of Transformer models for concept-to-text generation (first-author paper published at AAAI 2022) and commonsense QA.
 - **Creative Text Generation**: Investigated the automatic generation of personifications, tongue twisters, and narrative reordering (NAREOR). Publications (long papers) at AAAI 2022, COLING 2022, and EACL 2023. Also explored approaches for visual storytelling (VIST).
 - **Clinical Generative Reasoning**: Explored the generative reasoning capabilities of language models for the clinical healthcare domain. Investigated data augmentation and automatic template construction and infilling. First-author long paper published at EACL 2023.
 - **Data Augmentation for NLP**: Comprehensively surveyed existing data augmentation for NLP work and explored augmentation methods for text generation models. Co-first author papers published at ACL 2021 Findings and EMNLP 2020 DeeLIO Workshop.

University of Waterloo

- Undergraduate (USRA) Research Intern — Machine Learning and Natural Language Processing (NLP) Waterloo, Canada
Dr. **Jesse Hoey** and the Computational Health Informatics Lab *Jan. 2019 – Apr. 2020*
 - **Semantic Text Exchange**: Proposed semantic text exchange (STE) to adjust the semantic content of text while preserving its sentiment and fluency. Developed a pipeline called SMERTI for STE. Co-first author of a long paper published at EMNLP 2019. [News article](#).
 - **Personalized Dialogue Agents**: Proposed Human Level Attributes (HLAs) to model human personality based on character tropes. Helped design and develop ALOHA, a personalized dialogue agent. Main writer and presenter (oral) of a paper published at AAAI 2020.
- Undergraduate (USRA) Research Intern — Machine Learning and Natural Language Processing (NLP) Waterloo, Canada
Dr. **Pascal Poupart** and the Artificial Intelligence Lab *May 2019 – Apr. 2020*
 - **Robust Embeddings**: Devised a novel method for making word embeddings robust to noise using Bayesian statistics, where out-of-vocabulary words are modeled as Gaussian mixture distributions over vocabulary words weighted by their Levenshtein distance.
- Undergraduate Research Assistant (URA) — Robotics and Human-Computer Interaction (HCI) Waterloo, Canada
Dr. **Edith Law** and the HCI Lab *Sept. - Dec. 2018*
 - **Autonomous Teaching Robots**: Developed autonomous NAOqi robots that teach and interact with children, particularly those with learning disabilities. Investigated the effectiveness of various activities and robot personalities on children's learning.

POSTER PRESENTATIONS AND RECORDED TALKS

- “CHARD: Clinical Health-Aware Reasoning Across Dimensions for Text Generation Models.”
 - *EACL 2023*. Hybrid. May 2023. Poster presentation and recorded talk.
- “Retrieve, Caption, Generate: Visual Grounding for Enhancing Commonsense in Text Generation Models”
 - *AAAI Conference on Artificial Intelligence 2022*. Online. Feb. 2022. Poster presentation.
- “SAPPHIRE: Approaches for Enhanced Concept-to-Text Generation”
 - *International Conference on Natural Language Generation (INLG) 2021*. Aberdeen, UK. Sept. 2021. Poster presentation.
- “A Survey of Data Augmentation Approaches for NLP”
 - *Association for Computational Linguistics (ACL) 2021 Findings*. Online. Aug. 2021. Recorded talk.
 - *ACL 2021 Representation Learning for NLP (RepLANLP) Workshop*. Online. Aug. 2021. Poster presentation.
- “GenAug: Data Augmentation for Finetuning Text Generators”
 - *EMNLP 2020 Deep Learning Inside Out (DeeLIO) Workshop*. Online. Nov. 2020. Recorded talk.
- “Keep Calm and Switch On! Preserving Sentiment and Fluency in Semantic Text Exchange.”
 - *Empirical Methods in Natural Language Processing (EMNLP) 2019*. Hong Kong, China. Nov. 2019. Poster presentation.
 - *Vector’s Evolution of Deep Learning Symposium*. Toronto, Canada. Oct. 2019. Poster presentation.

EXTRACURRICULARS

- **Co-Founder and President – Stanford Piano Society (SPS)** (<https://piano.stanford.edu/>) *Sept. 2022 - Present*
- **Piano Cover Artist – YouTube** (Over 2000 subscribers) *2013 – Present*
- **Vice President of Education - University of Waterloo Data Science Club** *Jan. – Apr. 2020*
Helped lead events and workshops to promote and teach aspects of data science research to fellow undergraduate students.
- **President - University of Waterloo Piano Society** *2015 – 2020*
- **Software Developer - WATonomous** (University of Waterloo Autonomous Vehicle Design Team) *May – Aug. 2018*

PROJECTS AND HACKATHONS

- **StoryGen**: Generate stories from images using NLP and deep learning. <https://devpost.com/software/storygen> *Hack the North 2019*
- **iDoc**: Medical conditions automatic classifier built using ML and NLP. <https://devpost.com/software/idoc2-0> *Hack the North 2018*

Last Updated: *Sept 6, 2023*