

VARSHA KISHORE

Email: vk352@cornell.edu • Address: 514 Linn St, Apt B, Ithaca, NY 14850 • Number: (978) 269 7582

EDUCATION

Cornell University, Ithaca, NY

Computer Science, Ph.D.

August 2018 - Present

Advisor: Kilian Weinberger

Harvey Mudd College, Claremont, CA

Joint Major in Computer Science and Mathematics, B.S.

August 2014 - May 2018

Dean's List - Every Semester: Fall 2014 - Spring 2018

RESEARCH

Graduate Research Assistant - Cornell University

January 2019 - Present

Latent diffusion for language generation

- Learned a continuous text diffusion model in the latent space of an encoder-decoder model.
- Used the pre-trained auto-regressive decoder in conjunction with the diffusion model to generate text.
- Obtained state-of-the-art text diffusion results for conditional and unconditional text generation.
- Published at NeurIPS 2023.

IncDSI: Incrementally Updatable Document Retrieval

- Developed a real-time information retrieval system called IncDSI that extends Differentiable Search Index (DSI)¹ which is an end-to-end document retrieval system.
- With IncDSI, new documents can be added to an already trained system in less than a second, and the retrieval accuracy on queries corresponding to both the old and new documents is close to the accuracy obtained when retraining a system from scratch on the old and new documents.
- Published at ICML 2023.

Learning Iterative Neural Optimizers for Image Steganography

- Investigated using an iterative neural network to solve the optimization problem present in steganography, in order to reduce inference time for iterative steganography.
- The learned optimization procedure is fast, reliable and achieves new state-of-the-art results for steganography. The method can also be used to perform JPEG compression resistant steganography.
- This work is currently under review at ICLR2023.

Fined Neural Network Steganography

- Developed a novel iterative steganography method called Fixed Neural Network Steganography (FNNS). At a high level, FNNS makes changes akin to adversarial changes with a fixed (random or pre-trained) neural network for steganography.
- FNNS achieves a state-of-the-art error rate of 0% for hiding up to 3 bits per pixel (bpp) of information in an image. FNNS can also be used for provably secure face anonymization in image sharing.
- A paper about this project was published at ICLR2022.

BERTScore: Evaluating Text Generation with BERT

- Worked with two collaborators under the supervision of Professor Kilian Weinberger and Professor Yoav Artzi to design an automatic evaluation metric for text generation.

¹ <https://arxiv.org/pdf/2202.06991.pdf>

- The metric leverages contextual embeddings and outputs a similarity score between sentences. It achieved state of the art results on correlations with human judgments on multiple common benchmarks.
- This work was published at ICLR 2020.

Google Research Intern (Search team):

May 2023 - August 2023

- Worked with Ni Lao and Linting Xue and implemented a codebase for text diffusion from scratch in T5X.
- Tested different methods for improving the quality of text diffusion models for unconditional and conditional generation.
- Implemented consistency distillation for reducing the required number of diffusion steps to lower than 10.

Research Intern at Microsoft Research (MSR) - Relation Extraction

June 2021-August 2021

- Worked with Hoifung Poon, Tristan Naumann and Sheng Zhang on the Hanover Team on exploring contrastive learning for distantly supervised relation extraction.
- Found that training a supervised cross-entropy model on weakly labeled data is faster and often outperforms contrastive learning models. However, in settings with increased noise, contrastive learning can be helpful.

Research Intern at ASAPP

May 2019-August 2019

- Worked with Professor David Sontag to build classifiers for predicting whether patients have certain medical conditions. The end result was a tool that can be queried to produce a list of patients who are at risk for a particular medical condition.
- On a separate project, I worked on semi-supervised learning for a classification task on confidential company data. I worked on extending mixup, a data augmentation technique, to work with text data. I also investigated using co-training. These techniques led to an improvement of 5% on the classification task.

Intuitive Surgical Clinic Project (Harvey Mudd Capstone project)

September 2017 - April 2018

- Worked in a 4 person team to build machine learning and deep learning models to classify suturing gestures like threading the needle or knot tying, during a robotic minimally invasive surgical procedure. I was the project manager for this project.
- Investigated using different machine learning models like support vector machines, hidden markov models and neural networks (LSTMs, TDNNs) on video data and kinematics data from surgical procedures for prostatectomy.
- Highest accuracy achieved for classifying gestures was about 70%. The accuracy could further be improved by creating an ensemble of classifiers.

Computational Biology Research - Harvey Mudd College

January 2017 - April 2018

- Worked with Professor Jessica Wu to better understand the evolutionary history of organizations.
- Proved equivalence between the 3-Tree data structure and the Labeled Coalescent Tree data structure. Both these data structures contain a gene tree, a species tree and have a way to represent coalescence events. My project culminated in writing a supplement extending my advisor's work².
- Worked on developing and implementing a non-exponential time method to reconcile non-binary gene trees with binary species trees.

SMR Drives Research - Harvard University

June 2015 - August 2015

- Shingled Magnetic Recording (SMR) drives are hard disk drives with overlapping tracks to increase storage density. Developed a translation layer using C to improve the read write performance of SMR drives and to deal with problems that arise due to shingling.

² Wu Y.-C. et al. (2014). Most parsimonious reconciliation in the presence of gene duplication, loss, and deep coalescence using labeled coalescent trees. *Genome Res.* 24, 475–486.

- Researched file systems and cleaning algorithms that would be optimal to use with SMR drives and wrote a report summarizing pros and cons of different file systems.

WORK EXPERIENCE

Software Engineering Intern: IoT NetOps - Google Inc.

May 2017 - August 2017

- Worked on a research project to position 10-100 Android Things IoT devices using multiple 802.11mc Wi-Fi access points (APs). Distances to multiple APs were used to compute the location of the device.
- Implemented a method of using readings from a single mobile AP to locate devices.
- Devices could be located with an accuracy of 2.5 meters using fixed APs and 8 meters using a mobile AP.
- Made a JavaScript app to visualize sensor data from devices and the location of those devices.

Engineering Practicum Intern: Google Maps iOS - Google Inc.

May 2016 - August 2016

- Used Objective C to create a new user interface (UI) for selecting and downloading offline map data. The new UI clearly shows the area being downloaded and displays area around the selection region for context.
- Implemented a viewfinder that automatically resizes in response to the area in the selection rectangle. Thus, no error message is displayed when users zoom out beyond maximum downloadable area.
- Ensured that you cannot download in places where offline maps is not supported and optimized code for checking if offline maps is supported.

Math News Android Project - Harvey Mudd College

September 2016 - December 2016

- Developed this app as part of a four person team for Francis Su, the president of the Mathematical Association of America. The app displays math related news stories.
- A Firebase database is used to store information and links to relevant articles, magazines and books. The android app displays these as stories. Users can view stories, maintain favorites and share to social media.

PUBLICATIONS

[1] Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019, September). BERTScore: Evaluating Text Generation with BERT. In International Conference on Learning Representations.

[2] Kishore, V., Chen, X., Wang, Y., Li, B., & Weinberger, K. Q. (2021, September). Fixed Neural Network Steganography: Train the images, not the network. In International Conference on Learning Representations.

[3] Chen, X., Kishore, V., & Weinberger, K. Q. (2022, September). Learning Iterative Neural Optimizers for Image Steganography. In *The Eleventh International Conference on Learning Representations*.

[4] Lovelace, J., Kishore, V., Wan, C., Shekhtman, E., & Weinberger, K. (2022). Latent diffusion for language generation. *arXiv preprint arXiv:2212.09462*.

[5] Kishore, V., Wan, C., Lovelace, J., & Weinberger, K. Q. (2023, July). IncDSI: Incrementally Updatable Document Retrieval. In *International conference on machine learning*.

[6] Venderley, J., Matty, M., Krogstad, M., Ruff, J., Pleiss, G., Kishore, V., ... & Weinberger, K. (2020). Harnessing Interpretable and Unsupervised Machine Learning to Address Big Data from Modern X-ray Diffraction. *arXiv preprint arXiv:2008.03275*.

[7] Liu, Z., Wan, C., Kishore, V., Zhou, J. P., Chen, M., & Weinberger, K. Q. (2023). Correction with Backtracking Reduces Hallucination in Summarization. *arXiv preprint arXiv:2310.16176*.

TUTORING

Computer Science TA - Cornell University

August 2018 - May 2019

- Teaching assistant for Machine Learning for Data Science class from Jan-May 2019.
- Teaching assistant for Object Oriented Design and Data Structures (Honors) class from Aug-Dec 2018.
- Responsibilities include grading, teaching recitation sections and holding office hours.
- Was awarded the Cornell Computer Science Teaching Assistant Award for contributions as a TA.

Computer Science Grutor (Grader/ Tutor) - Harvey Mudd College

January 2015 - May 2018

- Tutored students in office hours for the following classes - Introduction to Computer Science, Algorithms, Data Structures and Machine Learning.
- Graded and provided feedback on assignments to students.

Science Bus

September 2014 - December 2015

- Taught science lessons to middle school students once a week. Each lesson involved a simple experiment that the students would do in order to understand a concept in Science.

LEADERSHIP ACTIVITIES:

- Secretary for Graduate Students for Gender Inclusion in Computing (2020-present): An organization to advocate for gender equity and inclusion in computing and related fields at Cornell. We organize events, workshops and interface with the department on gender and diversity related things.
- CS Colloquium Czar (2018-present): As a czar, I help organize the weekly CS colloquium at Cornell.
- Cornell URM Application support program (2022): Volunteered to help underrepresented students with their graduate school applications.
- On the board of the Computer Science Grad Student Organization (CSGO) at Cornell (2019-2021): We voice student concerns within the department to administration and help mitigate these concerns. We also plan and organize events for the CS department.
- Mentor (2019-Present): Mentor for incoming PhD students at Cornell.
- Officer for the Women in Computing (ACM-W) chapter at Harvey Mudd (2014-2018): Served as Chair of ACM-W during 2017-2018. We organized events with professors, tech talks with companies, led short tutorials, hosted mixers, funded students to attend conferences, etc.
- Peer Academic Liaison (2017-2018): My main role as a peer academic liaison was to help underclassmen transition to college by holding check-ins and other events.
- Dorm President (2016-2017): As dorm president, I was part of the student council, I took part in student budgeting and I helped organize dorm events.
- Honor Board Member (2016-2017): Helped deal with violations of Harvey Mudd's honor code. We also discussed and made changes to the honor code as required.