

# NIHARIKA SHRIVASTAVA

Email: [niharika.shrivastava@u.nus.edu](mailto:niharika.shrivastava@u.nus.edu) Mobile: (+65) 91744710/ (+91) 9869338372  
LinkedIn: [niharikashrivastava](https://www.linkedin.com/in/niharikashrivastava) Website: [orionstar25.github.io](https://orionstar25.github.io) GitHub: [OrionStar25](https://github.com/OrionStar25)

## EDUCATION

---

### National University of Singapore

Aug 2022 - Jan 2024

*Master of Computing (Artificial Intelligence Specialization)*

- **Key courses:** Trustworthy ML, Natural Language Processing, AI Planning and Decision Making, Uncertainty Modelling in AI, Neural Networks and Deep Learning, Knowledge Discovery and Data Mining.
- **Thesis:** “Evaluating Large Language Models for Generation of Unstructured Synthetic Clinical Reports”.
- **Teaching Assistant:** Software Engineering (CS3203), Foundations of AI (CS3263), Big Data Systems (CS4225).
- **GPA: 4.4/5. Honours (Distinction).**

### Indian Institute of Information Technology, Allahabad

Jul 2016 - Jun 2020

*Bachelor of Technology, Information Technology*

- Recipient of the Prof. Dr.-Ing. Matthias Kleiner **University Gold Medal** for academic and innovative excellence.
- **Thesis:** “Congestion-Aware Routing for Vehicles using Graph Search Optimization and Game Theory”.
- **GPA: 9.23/10. First Class Honours.**

## TECHNICAL SKILLS

---

- **Programming:** Python, C++, Golang, SQL, ROS, Java
- **AI, ML, Data Science:** PyTorch, TensorFlow, Keras, Scikit-learn, NumPy, Pandas, LangChain, PEFT
- **Big Data and Deployment:** Spark, Hadoop, Kafka, RDBMS, CI/CD, GCP, Docker, DevOps, MLOps
- **Other Skills:** Multimodal Deep Learning, Retrieval Augmentation, Generative AI, A/B Testing

## WORK EXPERIENCE

---

### Software Engineer II

Apr 2022 - Jul 2022

*Gojek, India*

- Led the **architectural and design decisions** for key customer-facing issues for the GoFood app.
- Built **real-time streaming data pipelines** using Kafka, **integrated data science models**, and **created visualization dashboards** with Metabase for business analytics.

### Software Engineer I

Jul 2020 - Apr 2022

*Gojek, India*

- Maintained the **Order Management System and related micro-services** of GoFood mobile app to ensure reliable ordering at a scale of 1M+ orders/day.
- Built **high-throughput systems** and enhanced **system reliability** for a distributed **lambda architecture** on GCP.
- **Collaborated with cross-functional** engineering, product and design teams to refine the product requirements.

## RESEARCH EXPERIENCE

---

### Graduate Research Assistant

Nov 2023 - Jan 2024

*NUS Sound and Music Computing Lab, Singapore*

- Explored **multi-modal LLMs** with MIDI data to extract hierarchical musical structures in pop songs.

### Data Science Intern

May 2023 - Aug 2023

*Synapse, Singapore*

- Modelled a framework with **large language models (LLMs)** for generating synthetic clinical reports.
- Evaluated **zero-shot, few-shot, and chaining prompt techniques** using LangChain to generate medical reports for 5 major domains given a public dataset (MIMIC-III).
- Performed **fine-tuning** of WizardVicuna 13B using QLoRA and increased semantic utility by 7%.

### Undergraduate Research Assistant

Jan 2020 - Jun 2020

*SUTD Robotics Lab, Singapore*

- Developed a socio-optimal **vehicle path-planning framework** leveraging **Constraint Satisfaction Problems** to provide congestion-free routes in a city; with an 84% increase in network utilization.
- Proposed a **novel algorithm** for optimal path selection in a **multi-class set-up** (combination of cars, public transport, walking) to dissipate traffic in a capacity-bound transportation network; 64% decreased travel time.
- **Benchmarked** on a diverse set of road networks from 3 metropolitan cities.

## Software Research Intern

May 2019 - Aug 2019

*Outreachy, The Fedora Project, Remote*

- Created an **open-source Python library** to solve the problem of Localization (L10N) by providing **dynamic translations** for all Fedora packages' metadata to its end users. <https://tinyurl.com/ypup2mzt>.

## KEY PROJECTS

### Information Retrieval and Search for Big Data [\[Link\]](#)

Nov 2023 - Dec 2023

*Natural Language Processing, Recommender Systems*

- Built a **custom web-scraoper** to collect all Cisco products' documents and their categories. Created an **interactive visualization** to analyse 5K+ category-product dependencies.
- Constructed a FAISS index of 700K+ documents to **implement search functionality** for retrieving the top-k documents.
- Implemented **retrieval augmentation (RAG)** using the index with Llama2 LLM for retrieving the top-k documents.

### Singing Voice Synthesis with Avatar Generation [\[Link\]](#)

Aug 2023 - Nov 2023

*CS5647: Sound and Music Computing, NUS*

- **Generated singing audio** by optimizing DiffSinger with BigVGAN vocoder and PNLM for faster inference.
- **Generated singing faces** by **fine-tuning** Wav2Lip and used DeepFace augmentation to capture facial features.

### Enforcing Privacy and Robustness in Machine Learning [\[Link\]](#)

Aug 2023 - Nov 2023

*CS5562: Trustworthy Machine Learning, NUS*

- **Constructed adversarial images** to make self-driving cars deviate from their original prediction using Jacobian Saliency Map Attack. **Explored defences** employing gradient obfuscation; **created adaptive attacks** to circumvent the defences.
- **Performed membership inference attacks** for a model in the MLaaS setting. Implemented the **differential privacy SGD algorithm** to get meaningful privacy guarantees for the dataset.

### Grammatical Error Correction [\[Link\]](#)

Aug 2022 - Nov 2022

*CS4248: Natural Language Processing, NUS*

- **Ensembled** top-6 pre-trained models' outputs using a **CNN architecture** to decide whether a specific sentence edit should be present. All final edits in a sentence are combined using a greedy approach to avoid character conflict.
- Created a **state-of-the-art GEC system** with a 2.5 points increase in the F0.5 score on the CoNLL-2014 test set.

## PUBLICATIONS AND PRESENTATIONS

- **N. Shrivastava** and M. Meghjani, "Congestion-Aware Routing for Multi-Class Mobility-on-Demand Service," *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, 2022, pp. 2093-2099, doi: 10.1109/CASE49997.2022.9926619
- G. Ganapathy, **N. Shrivastava**, O.P. Vyas, M. Singh, R. Arora, and S. Mishra. "Intrusion Detection and Attack Classification using an Ensemble Approach." *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 10, Oct. 2020, pp. 1616-1620
- Synthetic Clinical Reports Generation using LLMs, *PyTorch Conference, 2023*
- Software Engineering at Practice, *National University of Singapore, 2022*, <https://tinyurl.com/bdhw3phn>
- Opportunities and Diversity in Tech, *Google Developer Student Club, 2021*, <https://tinyurl.com/5zw6sp63>
- Predicting the Traffic Jam: Congestion-Aware Routing, *DevConf.US AND PyCon India, 2020*, [tinyurl.com/2zzt9f8z](https://tinyurl.com/2zzt9f8z)
- Language is a bridge, not a barrier, *Open Source Summit Europe AND Flock to Fedora, 2019*, [tinyurl.com/hy5ja3sf](https://tinyurl.com/hy5ja3sf)
- Students in developing nations and FOSS contribution limitation, *Flock to Fedora, 2019*, [tinyurl.com/3x5dvd4w](https://tinyurl.com/3x5dvd4w)

## AWARDS & SCHOLARSHIP

- **Best Paper Award**, Advanced Reinforcement Learning (CS6284), 2023
- **University of British Columbia Master of Data Science International Scholarship**, 2022
- **Runner-Up**, Red Hat Women in Open Source Academic Award, 2020
- **Best Undergraduate Thesis Award**, 2020
- **Winner**, Smart India Hackathon, 2019
- **The Linux Foundation Diversity Scholarship**, 2019
- **Mozilla Open Source Recognition**, 2018
- **Dean's Merit List**, 2020; 2017