SHUBHAM LOHIYA

shubhlohiya.github.io | shubhlohiya@gatech.edu | shubhlohiya | lohiya-shubham

Research Interests

Generative AI, Natural Language Processing, Reinforcement Learning, Knowledge Graphs, AI in Healthcare

Education

Georgia Institute of Technology — Atlanta, GA, USA
Master of Science in Computer Science – Machine Learning Specialization, ***GPA: 4.0/4.0*** — *Aug 2022 – May 2024*

- **Key Coursework:** Limited Supervision in ML, Natural Language Processing, DL for Robotics, ML in Comp. Bio.

Indian Institute of Technology Bombay — Mumbai, India
Bachelor of Technology in Mechanical Engg with Minors in CS and AI, ***GPA: 9.59/10.0*** — *July 2018 – May 2022*

- **Rank 1 – Institute Silver Medal**; **Institute Academic Prize** [2020, 2021, and 2022]
- **Key Coursework:** Intelligent Agents, Digital Image Processing, Web Search and Text Mining, Algorithms, OS

Professional Experience

Amazon | Applied Scientist Intern — *Fall 2023*
Building scalable ***Mulimodal*** *systems for* ***Automated Item Mapping*** *for Amazon's Product Catalog* — *Seattle, WA*

Adobe | Machine Learning Engineer Intern — *Summer 2023*
Developed a ***Generative AI*** *system for KPI-optimized* ***personalized marketing content*** — *San Jose, CA*

- Fine-tuned **Llama2** via **RLHF** using a reward model that learns from past consumer-content interaction data
- Designed an automatic qualitative evaluation protocol for generated marketing content using a GPT-4 expert

Anheuser-Busch InBev | Data Scientist Intern — *Summer 2021*
Engineered a ML Framework for Accounts Receivable ***Forecasting****, delivering* ***$40 million****+ in working capital benefits*

- Developed trend and seasonal features from historical data, and data from Sales, Collections and the Economy
- Trained ensembles of models like **ARIMAX**, **XGBoost**, **Random Forest**, and **MLP** achieving $<$ **5%** dispersion

Python Developer | Avrio Energy — *May 2020 – Sep 2020*
The firm is developing AI and IoT powered technology to improve the energy efficiency of businesses

- Designed Schema, Models, and APIs in **Django** for version 1 of Avrio Energy's Outlet Manager android app
- Worked with raw time series data in **InfluxDB** from **1300**+ appliances to perform feature-extraction for **ML**

Research Experience

Graduate Research Assistant | NLP | Pathology Dynamics Lab — *Aug 2022 - Present*
Guide: Prof. Cassie S. Mitchell, Department of Biomedical Engineering — *Georgia Tech*

Document Filtering for Drug Repurposing and Clinical Meta-Analysis

- Involved in development of **BioSift**, a new human-annotated document classification dataset for drug repurposing
- Implemented a suite of **NLI-based Zero-shot Multi-label Classifiers** for a document classification benchmark

Biomedical Entity Linking Survey

- **Introduction:** *Biomedical entity linking (BioEL) is the identification of biomedical concepts in text using* ***NLP***
- Conducted a comprehensive survey of BioEL models on a **new benchmark** of uniformly formatted datasets

MARBLER: Multi-Agent RL Benchmark & Learning Env for the Robotarium — *Spring 2023*
An Open Platform for Standarized Evaluation of Multi-Robot Reinforcement Learning Algorithms — *Georgia Tech*

- Created an **open-source MRRL framework** enabling training in simulation and evaluation on physical robots
- Benchmarked popular learning algorithms and included functionality for **rapid prototyping** of new experiments

Joint Completion and Alignment of Multilingual Knowledge Graphs — *Aug 2021 - July 2022*
Bachelor's Thesis, Guide: Prof. Soumen Chakrabarti, Department of Computer Science — *IIT Bombay*

- **Problem Statement:** *For multilingual KGs, AlignKGC attempts to study whether Knowledge Graph Completion or fact prediction adds valuable information for Relation Alignment and Entity Alignment tasks, and vice versa.*
- Developed AlignKGC, a new **SOTA framework** for joint completion and alignment of Multilingual KGs

Online Reinforcement Learning for Lane Following — *Mar 2021 - Apr 2021*
Guide: Prof. Shivaram Kalyanakrishnan, Department of Computer Science and Engineering — *IIT Bombay*

- Extracted left & right distance features from dashboard camera feed using Semantic Segmentation and masking
- Employed **Tile Coding** to encode continuous state-variables like velocity, steer and throttle in discrete form

Publications and Pre-prints

1. Harkanwar Singh, Soumen Chakrabarti, **Shubham Lohiya**, Prachi Jain, Mausam, **Joint Completion and Alignment of Multilingual Knowledge Graphs**, *accepted at The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022*
2. Reza J. Torbati, **Shubham Lohiya**, Shivika Singh, Meher Shashwat Nigam, Harish Ravichandar, **MARBLER: An Open Platform for Standarized Evaluation of Multi-Robot Reinforcement Learning Algorithms**, *under review at The International Symposium on Multi-Robot & Multi-Agent Systems (MRS), 2023*
3. David Kartchner, Irfan Al-Hussaini, Haydn Turner, Jennifer Deng, **Shubham Lohiya**, Prasanth Bathala, Cassie Mitchell, **BioSift: A Dataset for Filtering Biomedical Abstracts for Drug Repurposing and Clinical Meta-Analysis**, *Resource Paper, accepted at the 46th International ACM SIGIR Conference, 2023*
4. David Kartchner, Jennifer Deng, **Shubham Lohiya**, Tejasri Kopparthi, Prasanth Bathala, Daniel Domingo Fernández, Cassie S. Mitchell, **A Comprehensive Evaluation of Biomedical Entity Linking Models**, *under review at The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023*
5. David Kartchner, Irfan Al-Hussaini, Haydn P McCary Turner, Jennifer Deng, Zihan Wei, **Shubham Lohiya**, Cassie S. Mitchell, **TrialSieve: An Information Extraction Dataset for Automating Clinical Meta Analysis**, *under review at The 37th Conference on Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*

Scholarships

- Awarded the prestigious **OPJEMS Scholarship** for displaying academic and leadership excellence (2021)
- Recipient of the **NSF UG Engineering Scholarship** for demonstrating scholastic excellence (2022)
- Recipient of the Narotam Sekhsaria Foundation's **PG Scholarship** for pursuing graduate studies (2022)

Key Technical Projects

Web corpus indexing and compression | *CS635: Information Retrieval and Web Mining* *Fall 2021*
- Developed an **inverted index** for a corpus of 50,000+ web documents by encoding **D-gaps** for each token
- Conducted a comparative analysis of Index Compression methods by encoding the D-gap posting lists using various techniques like **Elias Gamma** Coding, **Golomb** Coding, and **Arithmetic** Coding

Profile Recommendation System for Online Dating | *ME781: Statistical ML and Data Mining* *Fall 2021*
- Devised heuristics to construct **compatibility feature vectors** for pairs of dating profiles using profile content
- Synthesized supervision signals using **Proxy Labeling** for learning models like Logistic Regression, SVM, and MLP
- Generated profile recommendations by ranking pair-wise scores, achieved a relevancy score of **79.94%** on test data

Mastering Atari Games using Deep Reinforcement Learning | *CS419: Introduction to ML* *Spring 2021*
- Trained a Deep Reinforcement Learning agent capable of surpassing human performance on classic Atari games like Pong, Breakout and Boxing using **high-level sensory information** in the form of game screen pixels
- Compared the performance of **off-policy** frameworks like **Deep Q-Network (DQN)** and **Double-DQN**

Automatic Raga Recognition in Indian Classical Music | *IE643: Deep Learning* *Fall 2020*
- Leveraged **tonic** normalized **pitch-tracked** frequencies of a music sample as features for raga classification
- Trained a model based on **LSTM**s with **attention** on random subsequences from the Carnatic Music Dataset
- Achieved a testset accuracy of **96.67%** with 60% majority voting, and **100%** with 50% majority voting

Shortest Path in a Maze | *CS747: Foundations of Intelligent and Learning Agents* *Fall 2020*
- Modelled given 2D mazes as **MDP**s with appropriate states, actions, rewards and transition probabilities
- Compared Howard's Policy Iteration, Value Iteration and Linear Programming algorithms to find shortest path

Image Quilting for Texture Synthesis and Transfer | *CS663: Digital Image Processing* *Fall 2020*
- Implemented a **patch-based algorithm** to synthesize a texture of any desired size from the given sample
- Used a modified quilting algorithm to transfer any given texture to any target image and obtained good results

Other Projects

- Designed a smart **Modular Vertical Farming** unit for climate-independent agriculture in compact spaces (2021)
- Analyzed data of Indian macro-economic indicators to **forecast Consumer Price Index** (CPI) using ML (2021)
- Trained a deep learning framework to perform **Camouflaged Object Detection** in nature images (2020)
- Developed a web-app to do **facial sentiment recognition** on a live video feed using a **CNN** architecture (2020)
- Created a Star Wars themed shooting game using **PyGame** to render graphics, animation, and sound (2020)
- Built sequence model for **Trigger Word Detection**, trained using synthetically generated audio data (2020)
- Constructed an autonomous Line-Following bot using **Arduino UNO**, IR sensors, and a **PID** controller (2019)

Technical Skills

Programming Languages	Python, C++, Scala, SQL, HTML, CSS, JavaScript, Bash, R, MATLAB, Java
Technologies/Frameworks	PyTorch, TensorFlow, OpenCV, Docker, Git, Databricks, Spark, WandB, MLFlow

Teaching Experience

Graduate Teaching Assistant | **Georgia Institute of Technology**

- **CS 6601 - Artificial Intelligence**, Prof. Thomas Ploetz, School of Interactive Computing *Fall 2022*

Teaching Assistantships | **IIT Bombay**

Facilitating smooth course organization, grading papers, mentoring students, conducting tutorials and help sessions

- **CS 419 - Introduction to Machine Learning**, Prof. Abir De, CSE Department *Spring 2022*
- **CS 101 - Computer Programming and Utilization**, Prof. S Akshay, CSE Department *Spring 2022*
- **IE 643 - Deep Learning**, Prof. P. Balamurugan, IEOR Department *Fall 2021*
- **ME 119 - Engineering Drawing**, Prof. Deepak Marla, Department of Mechanical Engineering *Fall 2021*
- **MA 106 - Linear Algebra**, Prof. Sudhir Ghorpade, Department of Mathematics *Spring 2021*
- **MA 108 - Differential Equations**, Prof. Prachi Mahajan, Department of Mathematics *Spring 2021*
- **PH 107 - Quantum Physics**, Prof. Shankaranarayanan S, Department of Physics *Fall 2020*

Python is Cool, Kids | **Student-run Summer Course** *Summer 2021*

- Spearheaded a team of students to conduct a summer course for **Practical Python Programming**, consisting of interactive live lectures and guided projects, with **1000+ enrollments**

Key Mentoring and Leadership Roles

Institute Secretary Technical Affairs | **Institute Technical Council, IIT Bombay** *Apr 2020 – Mar 2021*

Head of the Electronics and Robotics Club and part of a 23-member core team catering to 5000+ students

- Elected to lead and manage a team of **15+ members** to organize **20+** events, competitions and hackathons and mentor 1200+ electronics and robotics enthusiasts with an annual budget of over **INR 300,000**
- Coordinated the Institute Technical Summer Project program with **70% y-o-y** increase in completed projects
- Initiated the development of **'ERC Wiki'** - a repository of easily accessible resources for enthusiastic learners

Student Mentor | **Student Mentorship Programme, IIT Bombay** *May 2021 – Aug 2022*

Selected based on overall performance in a rigorous process comprising of interviews, SOP and peer reviews

- **Institute Student Mentor:** Guiding a group of **12 freshmen** through their first year at IIT Bombay
- **Department Academic Mentor:** Mentoring a group of **7 sophomores** with their academics and research

Extra Curricular Activities and Other Achievements

Achievements	• **Winner** of the Prospect 100 Global Tech Challenge – Covid-19 Hackathon judged by **Steve Wozniak**. • Selected among the four **delegates from India** to the 5-day virtual "Humanizing Digital 2021" **AI and Data Science** conference at **Chulalongkorn University, Thailand** • Ranked **4** in India's Best Student Contest 2015 organized by RaoIIT amongst **0.3** million participants • Selected among **top 30** students in a Nationwide Aptitude Test conducted by VNIT, Nagpur
Mentorship	• Mentored **9 freshmen** on a project to create an AI agent for mastering the snake game using **RL** • Guided 6 students with reading projects on **Deep Learning** and **Reinforcement Learning**
Technical	• Led a team of **4** to build a radio-controlled trainer aircraft capable of dropping payloads • Constructed an all-terrain obstacle manoeuvring bot controlled using a mobile application
E-Cell, IIT Bombay	• Led a team of **5** organizers during the **Entrepreneurship Summit 2020**, to successfully execute **20+** talks, interviews, and lectures as a **Corporate Relations Coordinator**