Shrusti Ghela

shrustighela.com | shrustighela1@gmail.com | linkedin.com/in/shrustighela | github.com/shrusti-ghela

Research Interests

Natural Language Processing, Evaluation, Human-AI Interaction

Education

University of Washington — *Seattle, WA*
Master of Science in Data Science — *Sept 2021 – Mar 2023*
Recipient of Betty G. Bengtson Endowed Scholarship, 2023

Vellore Institute of Technology — *India*
Bachelor of Technology in Information Technology — *May 2016 – June 2020*
Recipient of Scholarship for Academic Excellence, 2016 - 2020

Publications

HALoGEN: Fantastic LLM Hallucinations and Where to Find Them — *2025*
Shrusti Ghela*, Abhilasha Ravichander*, David Wadden, Yejin Choi
Association of Computational Linguistics (ACL) 2025
Outstanding Paper Award, ACL 2025
Best Paper Award, TrustNLP @ NAACL 2025

Unsupervised Learning Approaches for Dimensionality Reduction and Data Visualization — *2021*
Shrusti Ghela*, Anveshrithaa Sundareswaran*, B. K. Tripathy*
CRC Press (Taylor & Francis Group)
Recognized Technical Book, 2022

Research Experience

Allen Institute for AI and University of Washington — *Seattle, WA*
Research Collaborator — *July 2023 – Feb 2025*

HALoGEN: Fantastic LLM Hallucinations and Where to Find Them

- Defined and categorized 9 distinct prompting strategies across a broad spectrum of open-domain, content-grounded, response-based, and refusal-based tasks, establishing the first unified large-scale framework for studying how different prompt and task formulations induce LLM hallucinations.
- Constructed a comprehensive hallucination-identification pipeline spanning a 10,923-prompt benchmark, task-specific decomposition engines, and evaluators to detect and score hallucinations with high resolution.
- Evaluated ∼150,000 generations from 14 models (GPT-4, LLaMA-3, Mixtral, OLMo), revealing even the best-performing models hallucinate, with error rates ranging from 3% to 86%, and that larger models consistently hallucinate less on response-based tasks.
- Conducted an in-depth attribution study of why models hallucinate by tracing errors back to their underlying pretraining data using WIMBD, identifying three core types: Type A – Incorrect Recollection; Type B – Incorrect Knowledge; Type C – Fabrication.

University of Utah *(Advisor: Ana Marasović)* — *Salt Lake City, UT*
Research Collaborator — *Sept 2025 – present*

- Developing a cross-cultural framework to study general advice-seeking, dilemmas, and uncertainty-driven human-LLM interactions using multilingual surveys, human–LLM conversation data, and domain-expert evaluations.

COCO Health — *Seattle, WA*
Research Engineer — *July 2023 – Mar 2024*

- Implemented steering vectors to reduce sycophancy in an RLHF-trained LLM, achieving a 30% reduction in sycophantic behavior and significantly improving accuracy and robustness.
- Engineered a HIPAA-compliant API layer integrating a fine-tuned clinical chatbot with a mobile app using Python, Flask, FastAPI, OpenSSL, and encrypted PHI-safe data flows.

UW Medicine *Seattle, WA*
Data Scientist *Sept 2022 – Mar 2023*

- Automated audio-to-text processing for 100+ hours of clinical conversations, reducing transcription time by 90%.
- Applied NLP tools (NLTK, Kaldi, VADER, LDA) to cut manual annotation time by 60%.
- Built Tableau dashboards analyzing communication patterns in patient-provider interactions, improving provider awareness of effective communication by 37%.

Industry Experience

WM *(formerly Waste Management)* *Houston, TX*
Sr. Software Engineer - ML/NLP *May 2024 – present*

- Leading the development of WM's first AI Evaluation Lab, establishing organization-wide standards for responsible AI by designing automated and human-centered evaluation pipelines to test hallucinations, harmful outputs, fairness, and guardrail reliability in generative AI systems.
- Led WM's first Generative AI project, HeyWM, a Retrieval-Augmented Generation system for 50K+ employees that cut knowledge-lookup time by 40% and improved decision efficiency across business units.
- Redesigned key components of the Next-Day Optimization (NDO) engine, strengthening traffic-aware route generation, distance calibration, and time-bucket logic to ensure reliable and scalable optimization across 20K+ daily routes, improving route efficiency by 10% and reducing travel variance by 15%.
- Led WM's first AI/ML intern program, mentoring 3 interns on routing optimization, resulting in production-ready prototypes and a scalable framework for future cohorts.

Constellation Software *New York, NY*
Data Science Intern *Dec 2022 – Mar 2023*

- Designed and implemented an efficient Extract, Transform, Load (ETL) pipeline using Python and Snowflake to streamline the ingestion and processing of diverse data from databases, APIs, and streaming platforms, resulting in a 60% reduction in processing time and a 45% decrease in resource consumption.
- Curated comprehensive datasets of healthcare practitioners, facilitating targeted advertising campaigns for pharmaceutical clients during new drug launches. This involved meticulous data collection, cleaning, and integration, ensuring data quality and relevance. Concurrently, revamped data dictionaries to ensure domain-wide consistency, achieving a 100% reduction in data discrepancies.
- Devised and optimized complex SQL queries to extract data from the integrated datasets. Through optimization efforts, query speed increased by 35%, enabling faster data retrieval and analysis for decision-making.

Teaching Experience

Vellore Institute of Technology *India*
Teaching Assistant *Sept 2020 – Sept 2021*

Unsupervised Learning Approaches for Dimensionality Reduction and Data Visualization

- Designed and taught a graduate-level course on unsupervised learning, culminating in a published textbook adopted for coursework by students and instructors.

University of Texas *(Advisor: Rashmi Varma)* *Dallas, TX*
Teaching Assistant *May 2024 – Aug 2024*

Quantum Summer'24 Workshop

- Taught foundational quantum computing concepts to 35 high school students and guided them in developing their first quantum computing projects, fostering interest in STEM education.

eScience Institute *Seattle, WA*
Text Mining Specialist *June 2022 – Mar 2023*

- Delivered interactive lectures/workshops to educate students, researchers, and faculty members on the practical implementation of tools like Gensim, CoreNLP, PyNLPI, and more in their research projects, contributing to a 21% increase in NLP adoption across research initiatives.

Volunteering & Leadership Experience

Virufy: Developed an ML algorithm that prescreens for COVID-19 by analyzing voice patterns.
SnowEx 2022: Organized a hack week with 50+ students and 15 NASA researchers on real-time snowfall prediction.
Aliste Technologies: Launched a tech startup that develops affordable and user-friendly smart home solutions.
Vriksh: Utilized ML algorithms to identify high-potential tree-planting sites and directly contributed to planting 1,000+ trees to advance carbon-emission reduction goals.