

Shivam Sharma

Darmstadt, Germany • sharma.shivam2705@gmail.com • ssharma2705.github.io • Google Scholar • LinkedIn

Education

New Jersey Institute of Technology
MSc in Data Science

New Jersey, USA
Aug 2019 – May 2021

The LNM Institute of Information Technology
BTech in Computer and Communication Engineering

Rajasthan, India
Aug 2015 – May 2019

Publications

GroundInc: Grounding Inconsistencies for Video Caption Verification Jan 2026
Shivam Sharma, Y. Kementchedjieva, I. Gurevych
(Under Review, EMNLP 2026)

Exploring the benefits of domain-pretraining of generative large language models for chemistry Jan 2024
A. Acharya, **Shivam Sharma**, R. Cosbey, M. Subramanian, S. Howland, M. Glenski
(arXiv)

Benchmarking LLMs for Environmental Review and Permitting Jan 2024
R. Meyur, H. Phan, K. Hayashi, I. Stewart, **Shivam Sharma**, S. Chaturvedi, M. Parker, D. Nally, S. Montgomery, K. Pazdernik, A. Jannesari, M. Halappanavar, S. Munikoti, S. Horawalavithana, A. Acharya
(arXiv)

RAG vs. Long Context: Examining Frontier Large Language Models for Environmental Review Document Comprehension Jan 2024
H. Phan, A. Acharya, S. Chaturvedi, **Shivam Sharma**, M. Parker, D. Nally, A. Jannesari, K. Pazdernik, M. Halappanavar, S. Munikoti, S. Horawalavithana
(arXiv)

Foundation models of scientific knowledge for chemistry: Opportunities, challenges and lessons learned Jan 2022
S. Horawalavithana, E. Ayton, **Shivam Sharma**, S. Howland, M. Subramanian, S. Vasquez, R. Cosbey, M. Glenski, S. Volkova
(BigScience Workshop)

Expert: Public benchmarks for dynamic heterogeneous academic graphs Jan 2022
S. Horawalavithana, E. Ayton, A. Usenko, **Shivam Sharma**, J. Eshun, R. Cosbey, M. Glenski, S. Volkova
(arXiv)

Bang for your Buck: Performance Impact Across Choices in Learning Architectures for Crisis Informatics Jan 2022
Shivam Sharma, C. Buntain
(ISCRAM)

Evaluating and explaining natural language generation with genx Jan 2022
K. Duskin, **Shivam Sharma**, J.Y. Yun, E. Saldanha, D. Arendt
(DaSH, EMNLP)

An Evaluation of Twitter Datasets from Non-Pandemic Crises Applied to Regional COVID-19 Contexts Jan 2021
Shivam Sharma, C. Buntain
(ISCRAM)

Improving Classification of Crisis-Related Social Media Content via Text Augmentation and Image Analysis Jan 2020
Shivam Sharma, C. Buntain
(TREC-IS)

Combining neural, statistical and external features for fake news stance identification

Jan 2018

G. Bhatt, A. Sharma, **Shivam Sharma**, A. Nagpal, B. Raman, A. Mittal
(MSM, The Web Conference (WWW))

Research Experience

Research Assistant

Darmstadt, Germany

Ubiquitous Knowledge Processing (UKP) Lab, TU Darmstadt

Mar 2025 – present

- Developing benchmarks for misinformation detection and multimodal evidence retrieval in long-form videos

Data Scientist

Washington, USA

Pacific Northwest National Laboratory

Jan 2023 – Jan 2025

- Collaborated on research and development of application of Large Language Models on domain-specific scientific and environmental data

Research Associate

Washington, USA

Pacific Northwest National Laboratory

Oct 2021 – Dec 2022

- Collaborated on application of Natural Language Processing techniques for data curation and post-processing

Research Intern

Washington, USA

Pacific Northwest National Laboratory

Sept 2020 – May 2021

- Contributed towards evaluation of text generated from Large Language Models aimed towards improvement in detection and categorization of phishing emails

Research Assistant

New Jersey, USA

New Jersey Institute of Technology

Oct 2019 – Sept 2020

- Lead the research on application of Natural Language Processing techniques and Large Language Models on social media posts published during times of crisis

Research Intern

Uttarakhand, India

Indian Institute of Technology

June 2017 – July 2017

- Contributed towards research and application of Natural Language Processing techniques for fake news stance detection

Conference & Journal Reviewing

- Program Committee, ER 2024
- Review Committee, PERTRA Socio-NLP 2024
- Review Committee, IEEE TAI 2023
- Review Committee, Humanities & Social Sciences Communications Journal 2023
- Review Committee, EMNLP 2023

Invited Talks

- Lowercarbon AI Hackathon for Clean Energy, 2024

Programming Skills

Languages: Python, SQL, R

Libraries: HuggingFace Transformers, HuggingFace Accelerate, HuggingFace Datasets, PyTorch

Software: Git, Docker