Gaurav Negi

Galway, Ireland · g.negi1@universityofgalway.ie · He/him
LinkedIn · Huggingface · Google Scholar · Github

Research Profile

My research focuses on representing, extracting, and reasoning about opinions in text, specifically on unifying the existing representations with ontology-driven methods. Alongside this, we continue experimenting with scalable, LLM-based annotation alternatives. My broader interests include subjective language understanding, *"learning from disagreement"* and knowledge-grounded NLP.

Research Experience

Doctoral Researcher, University of Galway Sept 2022 – Present

Thesis: Unified Opinion Representation and Extraction (expected defence: Oct 2026); supervisor: Prof. Paul Buitelaar

- Formalised the Unified Opinion Concepts (UOC) ontology, unifying heterogeneous opinion-mining schemas; proposed the UOCE extraction task with an evaluation dataset and tailored metrics.
- Built LLM-based annotation and adjudication pipelines for fine-grained opinion analysis; released the resulting models and datasets on HuggingFace.
- Developed linguistically enhanced extraction models, and integrated opinion semantics into GNN classifiers for downstream emotion analysis.

Visiting Researcher, University of Oslo Jan 2025 – Feb 2025

Language Technology Group (LTG)
Collaborated on a system for the extraction of narrative roles from text, under the supervision of Prof. Lilja Øvrelid.

Research Assistant, Data Science Institute (DSI), Galway Aug 2021 – Aug 2022

Machine Learning, AI, Natural Language Processing, Semantic Web

- PANDEM-2 (EU H2020): suggestion mining, sentiment/emotion analysis, and aspect mining of social media posts, enabling two-way communication between the public and pandemic-management authorities.
- Designed, trained, and deployed neural models end-to-end (TensorFlow, TensorFlow Serving, Docker).
- Term and taxonomy extraction; corpus collection.

Work Experience

Software Engineer, Centre for Development of Advanced Computing, Pune Sept 2018 – June 2021

Education, Training and Research, Online Examination & Registration Portals

- Data processing, consistency checks, and BI dashboards (Grafana, Plotly) for national-scale IAF examination platforms (**STAR**, **AFCAT**).
- Built an automated help-desk chatbot (RASA) and content-based image-validation services (classification/detection models served as RESTful APIs).
- Containerised deployment of applications and models.

Teaching

Teaching Assistant, University of Galway 2022 – 2024

- Introduction to NLP (3 semesters)
- Advanced Topics in NLP (1 semester)

Academic Service

Shared-Task Co-Organizer, WILDRE-7 Shared Task on Code-Mixed Sentiment Analysis 2024
LREC-COLING 2024 — Configured and managed the CodaLab evaluation platform for task participants.

Education

PhD in Computer Science & Engineering 2022 – present
University of Galway, Ireland (expected defence: Oct 2026)

Post Graduate Diploma in Big Data Analytics 2018
Centre for Development of Advanced Computing (C-DAC), Pune, India
Bachelor of Technology in Electronics and Communication 2013 – 2017
THDC Institute of Hydropower Engineering and Technology, Tehri, India

Technical Skills

Programming Languages: Python, Scala, Java
Data Visualisation: Plotly, Grafana, Matplotlib, Streamlit
Data Engineering: Pandas, Elastic Search (ELK) stack, MongoDB, PostgreSQL
Web Frameworks: Scalatra, Play Framework
Machine Learning/AI Tools: PyTorch, Transformers, HuggingFace, vLLM, DSPy, Weights & Biases, scikit-learn, spaCy, RASA, TensorFlow ecosystem (Keras, TensorBoard, Serving)
Misc.: Docker, GitHub, OpenCV

Publications

[1] **Gaurav Negi**, Dhairya Dalal, Omnia Zayed, and Paul Buitelaar. 2025. Towards Semantic Integration of Opinions: Unified Opinion Concepts Ontology and Extraction Task. In Proceedings of the 5th Conference on Language, Data and Knowledge, pages 174–185, Naples, Italy. Unior Press.

[2] Dhairya Dalal, **Gaurav Negi**, and Davide Picca. 2025. LLMs and Emotional Intelligence: Evaluating Emotional Understanding Through Psychometric Tools. In Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '25). ACM, New York, NY, USA, 323–328.

[3] **Gaurav Negi**, Atul Kr. Ojha, Omnia Zayed, and Paul Buitelaar. 2025. Towards Temporal Knowledge-Base Creation for Fine-Grained Opinion Analysis with Language Models. In Proceedings of the 3rd Workshop on Knowledge Base Construction from Pre-Trained Language Models (KBC-LM), co-located with ISWC 2025. CEUR-WS Vol. 4041.

[4] Omnia Zayed, **Gaurav Negi**, Sampritha Hassan Manjunath, Devishree Pillai, and Paul Buitelaar. 2025. LUCE: A Dynamic Framework and Interactive Dashboard for Opinionated Text Analysis. In Proceedings of the 31st International Conference on Computational Linguistics: System Demonstrations, pages 104–116, Abu Dhabi, UAE. ACL.

[5] Egil Rønningstad and **Gaurav Negi**. 2025. LTG at SemEval-2025 Task 10: Optimizing Context for Classification of Narrative Roles. In Proceedings of SemEval-2025, pages 440–447, Vienna, Austria. ACL.

[6] **Gaurav Negi**, Rajdeep Sarkar, Omnia Zayed, and Paul Buitelaar. 2024. A Hybrid Approach to Aspect Based Sentiment Analysis Using Transfer Learning. In Proceedings of LREC-COLING 2024, pages 647–658, Torino, Italia. ELRA and ICCL.

[7] Priya Rani, Theodorus Fransen, John P. McCrae, and **Gaurav Negi**. 2024. MaCmS: Magahi Code-mixed Dataset for Sentiment Analysis. In Proceedings of LREC-COLING 2024, pages 10880–10890, Torino, Italia. ELRA and ICCL.

[8] Priya Rani, **Gaurav Negi**, Saroj Jha, Shardul Suryawanshi, Atul Kr. Ojha, Paul Buitelaar, and John P. McCrae. 2024. Findings of the WILDRE Shared Task on Code-mixed Less-resourced Sentiment Analysis for Indo-Aryan Languages. In Proceedings of the 7th Workshop on Indian Language Data: Resources and Evaluation, pages 17–23, Torino, Italia. ELRA and ICCL.

[9] **Gaurav Negi**, Omnia Zayed, Martin Serrano, and Paul Buitelaar. 2026. Semantically Enriching Investor Microblogs for Opinion-Aware Emotion Analysis: A Practical Approach. In Proceedings of the SMART AIDA Workshop, co-located with the IEEE International Conference on Smart Computing (SMARTCOMP 2026), Messina, Italy. To appear.

[10] (Pre-print) **Gaurav Negi**, M. A. Waskow, John P. McCrae, Omnia Zayed, and Paul Buitelaar. 2026. Large Language Models as Automatic Annotators and Annotation Adjudicators for Fine-Grained Opinion Analysis. arXiv preprint arXiv:2601.16800.

Languages

Hindi (Native), English (Full professional proficiency)