Amey Hengle

Predoctoral Researcher | IIT Delhi

ameyhengle.github.io @ ameyhengle@gmail.com github.com/AmeyHengle Google Scholar
linkedin.com/in/amey-hengle x.com/AmeyHengle

Research Interests: Parameter-efficient training of LLMs | LLM Evaluation | Low-resource NLP

Education

(Incoming) August 2025 — **University of Maryland, College Park** — **Maryland, USA**
Master of Science (M.S.) in Applied Machine Learning

June 2020 July 2016 — **Savtribai Phule Pune University | PVG's College of Engineering and Technology** — **Pune, India**
Bachelor of Engineering (B.E.) in Computer Science

Experience

Present Aug 2023 — **IIT Delhi | Laboratory for Computational Social Systems (LCS2)** [🌐] — **New Delhi, India**
Predoctoral Researcher | Research Associate | Advisors: Prof. Tanmoy Chakraborty & Prof. Md Shad Akhtar
- Working on mechanistic interpretability of LLMs in three key areas: prompt sensitivity, multilingual in-context learning, and linguocultural bias.
- Led a study to investigate the long-context behaviour of multilingual LLMs. (*accepted at NAACL 2025*)
- Worked on auto calibrating LLMs for reference-free NLG evaluation. (*accepted at NAACL 2025*)
- Devised a multi-task instruction-tuned model for non-toxic counterspeech generation. (*NAACL 2024*)

June 2024 Dec 2022 — **IIT Bombay | Cognitive & Behavioural Neuroscience Laboratory** [🌐] — **(Remote) Powai, India**
Research Collaborator | Advisor: Prof. Rashmi Gupta
- Assessed the performance of LLMs in diagnosing comorbid mental health conditions, identified their key limitations, and developed a novel dataset to address these gaps. (*EMNLP 2024*)

June 2023 March 2023 — **Carnegie Mellon University | Language Technologies Institute** [🌐] — **(Remote) Pittsburgh, PA**
Research Collaborator | Advisor: Prof. David Mortensen
- Worked on evaluating the morphological generalization capabilities of LLMs. (*EMNLP 2023*)

July 2023 Aug 2021 — **SKIT.ai | Forbes Asia 30-Under-30 AI Start-up** [🌐] — **Bengaluru, India**
Machine Learning Engineer
- Worked on end-to-end design, training, deployment, and evaluation of speech-based virtual assistants.
- Built light-weight speech-to-text and speech-to-intent models for 14 low-resource Indian languages.
- Developed LLM-agents using LangGraph to fully-automate code and functions for API calls.

Selected Publications (∗ denotes equal contribution)

[7] **Multilingual Needle in a Haystack: Investigating Long-Context Behavior of Multilingual Large Language Models** [Preprint | code]
Amey Hengle, Prasoon Bajpai, Soham Dan, Tanmoy Chakraborty
Conference of the North American Chapter of the Association for Computational Linguistics (Main) [NAACL '25]

[6] **CSEval: Towards Multi-Dimensional and Reference-Free Counterspeech Evaluation using Auto-Calibrated LLMs**
Amey Hengle, Aswini Kumar, Anil Bandhkavi, Tanmoy Chakraborty
Conference of the North American Chapter of the Association for Computational Linguistics (Main) [NAACL '25]

[5] **Still Not Quite There! Evaluating Large Language Models for Comorbid Mental Health Diagnosis** [Paper | code]
Amey Hengle*, Atharva Kulkarni*, Shantanu Patankar, Rashmi Gupta
Conference on Empirical Methods in Natural Language Processing (Main) [EMNLP '24]

[4] **Intent-conditioned and Non-toxic CS Generation using Multi-Task Instruction Tuning with RLAIF** [Paper | code]
Amey Hengle*, Aswini Kumar*, Sahajpreet Singh, Md Shad Akhtar, Tanmoy Chakraborty
Conference of the North American Chapter of the Association for Computational Linguistics (Main) [NAACL '24]

[3] **Counting the Bugs in ChatGPT's Wugs: A Multilingual Investigation into the Morphological Capabilities of a Large Language Model** [Paper | code]
Leonie Weissweiler, Valentin Hofmann, ..., Amey Hengle, ..., Hinrich Schuetze, Kemal Oflazer, David Mortensen
Conference on Empirical Methods in Natural Language Processing (Main) [EMNLP '23]

Academic Service

Conference Reviewer ICLR'25, ARR'24 (EMNLP'24, ACL'24, NAACL'24), NAACL'23, EMNLP'21, EACL'21
Conference Volunteer EMNLP'22, EACL'22

Research Projects

Quantifying Prompt Sensitivity in Large Language Models Nov 2024 - present
Advisor: Prof. Tanmoy Chakraborty | IIT Delhi

› Exploring **prompt sensitivity index** – a metric to measure sensitivity of LLMs towards *intent-preserving* variations in prompts such as prompt templates, spelling errors, or alterations in the position and language of instruction (user query).

Understanding Knowledge-Conflicts in Multilingual Long-Context LLMs Nov 2024 - present
Advisor: Dr. Soham Dan | Microsoft Research

› Conducted large-scale experiments on the **Dunning-Kruger effect** in long-context LLMs (> 64k tokens), finding a correlation between longer contexts and the models' failure to identify knowledge conflicts and locate relevant information.
› Developed fact-aware instruction tuning & attention calibration technique to improve LLMs' detection of conflicting information in long inputs.

Understanding Cross-Task and Cross-Lingual Knowledge Trasfer during ICL in LLMs July 2024 - Nov 2024
Advisor: Prof. Tanmoy Chakraborty | IIT Delhi

› Conducted a large-scale analysis of LLMs to explore if 1) In-context learning creates task-specific vectors 2) demonstrations from high-resource tasks improve performance on low-resource ones.
› Developed a dynamic learning setup using multi-armed bandits to optimize high-resource tasks for few-shot ICL.

Probing Information Loss during Long-Context Retrieval in Multilingual LLMs April 2024 - December 2024
Advisor: Dr. Soham Dan | Microsoft Research

› Found that LLMs struggle to understand, retrieve, and reason explicitly-stated answers in multilingual long-input sequences.
› Analyzed attention patterns across the feed-forward layers of an LLM to identify layers responsible for information loss.

Industry Projects

Inference Latency Optimizing for Dialog-Act Classification Models *Skit.ai*

› Reduced inference latency for dialog-act classification models at Skit.ai by 80% using structured pruning, TensorFlow Lite quantization, and a dynamic TorchServe framework, optimizing performance under hardware constraints.

Open-Source Contributions (Primary Developer) *Skit.ai*

› Developed a Python library for spoken language understanding (SLU) and dialog agents (250k downloads). [🔗]
› Developed a Python library to compute geo-spatial concepts like bearing angle and rhumb line distance (16k downloads). [🔗]

Honours & Awards

1. Shared Task Winner | WANLP @ EACL'21 [🔗] Best system for sarcasm identification in Arabic posts.

2. Shared Task Winner | TechDoFication @ ICON'20 [🔗] – Coarse-grained technical domain identification of documents in indigenous Indian languages.

3. Bachelor's Thesis Award | ZS ASPIRE 2020 [🔗] – 2^{nd} place amongst 33,000 teams & cash prize of 2,00,000 INR for Bachelor's Thesis project.

References

1) Dr. Tanmoy Chakraborty *Associate Professor, IIT Delhi, India* [🌐]

2) Dr. Md Shad Akhtar *Assistant Professor, IIIT Delhi, India* [🌐]

3) Dr. Soham Dan *Senior Research Scientist, Microsoft Research, USA* [🌐]

4) Ajay Nair *Associate Director, Skit.ai, India* [🌐]