Kanishk Verma

kanishk.tcet@gmail.com kanishk-verma kanishk-r-verma.github.io kanishk-verma

Profile

AI researcher with 4+ years of experience in **NLP and LLMs**, focusing on autonomous context-sensitive online safety solutions. Skilled in fine-tuning transformers (LoRA, prompt-tuning) low-resource corpora in real-world domains. **Recipient of the Google–Research Ireland Fellowship** for high-impact AI work in online safety. Currently leveraging self-supervised learning with human-in-the-loop strategies to advance LLM applications toward zero-touch network automation in complex, multi-agent environments.

Education

PhD in Computer Science (Expected Dec 2025) *Sept 2021 – Dec 2025*
Dublin City University, Ireland

Thesis: A Data-Driven Toolkit to Combat Cyberbullying Amongst Teens (*project website*)
Research areas include transformer fine-tuning, explainable AI, and data augmentation with LLMs.

MSc in Computer Science *Sept 2019 – Aug 2020*
Dublin City University, Ireland

Thesis: Implicit aspect-based opinion mining of airline reviews (*Springer*).

BEng in Information Technology *Aug 2012 – July 2016*
University of Mumbai, India

Relevant coursework: Data Structures & Algorithms, RDBMS, OOP, and Web Programming.

Technical Skills

- **AI/ML:** PyTorch, Scikit-learn, Transformer Architectures, Parameter-Efficient Fine-Tuning (PeFT) (e.g., LoRA, prompt-tuning), Transfer Learning.
- **Programming:** Python, R, Java, SQL, HTML, CSS.
- **Tools/Cloud:** Flask, Django, Dash, GCP, AWS, Firebase.
- **Specialization:** LLM reasoning and planning, question answering, interpretability, low-resource NLP.

Experience

PhD Researcher – NLP & LLMs *Sept 2021 – Present*
ADAPT Centre, Dublin City University *Dublin, Ireland*

- Fine-tuned multilingual transformer models using transfer learning and parameter efficient fine-tuning (PeFT) for social media text classification, resulting in 78-81% detection of cyberbullying across three languages.
- Analyzed interpretability of BERT and RoBERTa on cyberbullying data; found attention aligned with human rationales only 41.6% of the time, highlighting its limits for safety-critical NLP tasks.
- Benchmarking encoder- and decoder-only LLMs for sexism detection in SemEval-2023 Task 10 (ACL); ranked 16th out of 54 global teams.

Research Assistant – AI for Content Moderation *Sept 2020 – Aug 2021*
ADAPT Centre, Dublin City University *Dublin, Ireland*

- Conducted corpus analysis of 150+ studies to derive actionable AI strategies for content moderation (Meta Content Policy Grant).
- Designed transformer-based classification models (transfer learning, encoder-based) achieving 81% F1 on cyberbullying detection.
- Co-led participatory design with 30+ teens to build an explainable AI toolkit for online safety; aligned outcomes with stakeholder needs.

Data Scientist Intern – NLP in Healthcare *Mar 2020 – Aug 2020*
Radmol AI (now Durotimi AI) *Dublin, Ireland*

- Built NLP pipeline integrating OCR and transformer-based encoders for radiology report parsing, reducing analysis time by 15%.
- Conducted A/B testing and statistical analysis to evaluate AI-driven workflow tools for colorectal cancer diagnostics.
- Deployed model on 300+ annotated clinical reports, achieving 79% classification accuracy in identifying diagnostic patterns.

Software Analyst *Nov 2016 – May 2019*
Accenture *Mumbai, India*

- Automated monthly financial reporting with SQL, Java, and Selenium, increasing efficiency by 12%.
- Developed clear technical guides and led onboarding sessions, accelerating new hire productivity and reducing ramp-up time by 8%.
- Supported cross-team integration of automation tool into 3 enterprise workflows, reducing manual reporting errors by 18%.

Publications

Journal Articles

- Leveraging Machine Translation for Cross-lingual Cyberbullying Classification. (*Journal of Natural Language Engineering*)
- Effectiveness of AI-Based Cyberbullying Interventions. (*Social Media + Society*)
- Implicit Aspect-Based Opinion Mining and Analysis. (*Springer*)

Conference Papers

- Beyond Binary: Complexities in Cyberbullying Detection. (*COLING–LREC 2024*)
- DCU at SemEval-2023 Task 10: Encoder vs Decoder LMs. (*ACL 2023*)
- Can Transformers Explain Cyberbullying Detection? (*COLING 2022*)
- Benchmarking Language Models for Cyberbullying Identification. (*LREC 2022*)
- Fine-tuning LMs for English-Maltese Social Data (*RANLP 2021*)

Reports

- Contribution to the **EU** Delegated Regulation on Data Access under the **Digital Services Act** (*Link*)
- Contribution to the **Online Safety Code** (Ireland) – **Coimisiún na Meán** (*Link*)

Awards & Fellowships

- Google–IRC Online Safety Fellowship (2021–2025) (*Link*)
- Early Stage Researcher of the Year – ADAPT Centre (2024) (*Link*)
- Best Student Contribution – ADAPT Annual Conference (2024) (*Link*)