Nickil Maveli

Last updated on February 5, 2026

Edinburgh, United Kingdom
✉ nickil.maveli@ed.ac.uk
⌂ https://nickilmaveli.com

Research Interests

I am broadly interested in Natural Language Processing, Machine Learning, and Information Retrieval. My recent research has focused on code generation using large language models (LLMs). Specifically, I aim to build and evaluate CodeLLMs to assist software engineering.

Education

2022 – current **PhD, Natural Language Processing**, The University of Edinburgh, Scotland, UK.
Research Unit: Institute for Language, Cognition and Computation (ILCC)
Research Areas: Program Synthesis, Code Generation, Neuro-Symbolic Reasoning, Compositional Generalisation, Large Language Models
Advisor: Prof. Shay Cohen and Prof. Antonio Vergari

2020 – 2021 **M.Sc. by Research, Informatics**, The University of Edinburgh, Scotland, UK.
Research Unit: Institute for Language, Cognition and Computation (ILCC)
Research Areas: Unsupervised Parsing, Knowledge Graph Completion
Grade: Pass with Distinction
Advisor: Prof. Shay Cohen

2011 – 2015 **B.E. in Electronics and Communication**, CMR Institute of Technology, Bangalore, India.
GPA: 7.58/10

Work Experience

Oct 2021 – July 2022 **Applied Research Scientist - NLP**, Cookpad, Bristol, UK.
- Operationalized an active learning, intermediate-task fine-tuning based multilingual meal-classification algorithm.
- Implemented a label suggestion algorithm for efficient recipe tagging.

May 2018 – Sept 2020 **Senior Software Developer - NLP**, Niki.ai, Bangalore, India.
- Developed a real-time pause detection algorithm on speech signals to identify end-of-speech conditions.
- Designed a probabilistic model to classify if an utterance is a valid parse or not.
- Developed a bot that could respond to about 75 FAQs in regional languages and English.
- Built a character-level RNN model to classify utterances as 'Hinglish' words with minimum lexicon support.
- Created an algorithm to automatically generate contextual synonyms and augment concept-grammar data.

Sep 2016 – Feb 2018 **Data Scientist - Research and Development**, Dave.ai, Bangalore, India.
- Built a time series model to generate weekly/monthly sales forecasts of products with scarce data.
- Devised a robust machine learning model to predict and optimize e-commerce metrics for retailers.
- Created an image search engine to recommend similar images based on segmentation as a selective search.

Feb 2016 – May 2016 **Junior Data Scientist**, Hotify.ai, Bangalore, India.
- Aggregated tweets from pre-selected news sources to estimate trending hashtags.
- Created actionable insights by analyzing how select twitter users through A/B testing react to real-time events.
- Prepared training data by examining RSS feeds of news sources and rated each source using multiple factors.

Internships

Aug 2024 – Feb 2025 **Research Intern**, Huawei R&D, Edinburgh, UK.
Topic: Integrating knowledge graphs in LLMs for language-to-code generation to solve math reasoning problems
Mentor: Prof. Jeff Z. Pan

Aug 2021 – Oct 2021	**Research Assistant**, Usher Institute (University of Edinburgh) and Institute of Health Informatics (UCL), Edinburgh, UK.
	Topic: Clinical Natural Language Processing for Automated Medical Coding [Details]
	Mentor: Prof. Honghan Wu and Dr. Hang Dong
May 2021 – Aug 2021	**Student Developer**, Google Summer of Code (GSoC), Remote.
	Topic: Identify Joint Meaning Construal Patterns using Co-speech Gestures [Details] [Slides]
	Mentor: Prof. Tiago Torrent

Selected Publications

Preprint	Can LLMs Compress (and Decompress)? Evaluating Code Understanding and Execution via Invertibility [PDF]
	Nickil Maveli, Antonio Vergari, Shay Cohen
	Under review
NAACL 2025 (Findings)	What can Large Language Models Capture about Code Functional Equivalence? [PDF]
	Nickil Maveli, Antonio Vergari, Shay Cohen
	Findings of NAACL 2025
ACL 2022 (Findings)	Co-training an Unsupervised Constituency Parser with Weak Supervision [PDF]
	Nickil Maveli, Shay Cohen
	Findings of ACL 2022
WNUT@EMNLP 2020	EdinburghNLP at WNUT-2020 Task 2: Leveraging Transformers with Generalized Augmentation for Identifying Informativeness in COVID-19 Tweets [PDF]
	Nickil Maveli
	Proceedings of the 2020 EMNLP Workshop W-NUT
Master's Thesis	Co-training a Weakly Supervised Constituency Parser of Natural Language [PDF]
	Nickil Maveli

Awards

2025	Student Volunteer Award at NAACL 2025.
2022	Student Volunteer Award at ACL 2022.
2022	Fully-funded UKRI CDT in NLP scholarship to pursue PhD studies.

Invited Talks

July 2025	Student Panel for MSc Dissertation Q&A event
July 2022	International Society for Gesture Studies (ISGS), Symposium IV [Details] [Video]
Dec 2021	University of Oxford, IMCC seminar [Details] [Video]
Dec 2017	Techfest IIT Bombay, Artificial Intelligence summit [Details]

Technical Skills

Specialties	Natural Language Processing, Machine Learning, Deep Learning, Reinforcement Learning, Recommender Systems, Time-series Forecasting
Languages	Python, R, Matlab, Javascript, SQL, HTML/CSS, LaTeX
Tools and Frameworks	Pandas, NumPy, SciPy, Matplotlib, Scikit-Learn, NLTK, Keras, PyTorch, TensorFlow, PySpark, Flask, Django, Docker, Kubernetes, AWS, Jira, Git, Elastic Stack, MongoDB

Key Achievements

2025	Selected to participate in Advanced Language Processing School (ALPS 2026).
2022	Selected to audition as a DataCamp instructor for the "Introduction to mlflow" course.

2022	Secured Top 3% in the Evaluating Student Writing Challenge hosted on Kaggle; Won a silver medal. [Details]
2021	Selected to participate in the Lisbon Machine Learning School (LxMLS 2021). [Details]
2020	Secured Top 2% in the Jigsaw Multilingual Toxic Comment Classification Challenge hosted on Kaggle; Won a silver medal. [Details]
2019	Finished 2nd out of 479 competitors in the Quality Forecasting in Cement Manufacturing Challenge hosted on CrowdAnalytix; Won $2,500 as prize money; Winner representing India. [Details]
2019	Secured Top 3% in the Quora Insincere Questions Classification Challenge hosted on Kaggle; Won a silver medal. [Details]
2018	Finished 2nd out of 750 competitors in the Predicting How Points End in Tennis Challenge hosted on CrowdAnalytix; Won $2500 as prize money; Featured in Tennis Australia's official press release. [Details] [Media Coverage]
2017	Placed 3rd out of 2975 competitors in the Click Prediction Challenge hosted on AnalyticsVidhya; Awarded a cash prize of ₹25000. [Details]
2016 – present	Among Top Stack Overflow Users for tags—Python, Pandas, NumPy, Matplotlib, with an overall 25000+ reputation score and more than 5 million post views worldwide. [Details]

Academic Services

Teaching Assistant

2025-26	Advanced Topics in Natural Language Processing (INFR11287) - Demonstrator
2024-25	Natural Language Understanding, Generation, and Machine Translation (INFR11157) - Tutor
2023-24	Cohort '22 Representative, CDTalks Organizer
2023-24	Accelerated Natural Language Processing (INFR11125) - Demonstrator
2022-23	Accelerated Natural Language Processing (INFR11125) - Tutor, Demonstrator, TA (Piazza)

Reviewer

2025-current	COLM
2022-current	ACL, NAACL, EMNLP
2021	ACL, AmericasNLP
2020	W-NUT@EMNLP (Shared Task 2), Scipy (Tutorials, Talk & Poster Presentations)

Volunteer

2021	ACL, ICML, NAACL, ICLR, EACL
2020	EMNLP

Respondent

2020	Semantic Scholar Diary Study project conducted by the Allen Institute for AI

Open Source Contributions

2022 – present	ivy [Commits]