Jasivan Alex SIVAKUMAR

Department of Computer Science,
University of Sheffield,
United Kingdom
E-mail: jasivakumar1@sheffield.ac.uk

Webpage: jasivan.github.io
Github: jasivan
Twitter: @jasivan_s
LinkedIn: Jasivan Sivakumar

QUALIFICATIONS

Sep 21 - Sep 25 **PhD Computer Science** (Supervised by Dr. Nafise Sadat Moosavi)
- ▹ Natural Language Processing Group, University of Sheffield, United Kingdom
- ▹ Research title: *Numerical Reasoning for General-Purpose Language Models*
- ▹ Focus: AI, ML, NLP, NLG, LLMs, Reasoning, Number Representation, Tokenisation

Sep 20 - Sep 21 **MA Computational Linguistics - Distinction**
- ▹ University of Wolverhampton, United Kingdom

Sep 16 - Sep 17 **PGCE Secondary Education (Mathematics)**
- ▹ University of Cambridge, United Kingdom

Oct 12 - Jul 16 **Bachelor (MMath) Mathematics - First Class**
- ▹ University of Warwick, United Kingdom

PUBLICATIONS

Jul 23 (ACL) **Sivakumar, J.**, & Moosavi, N. (2023). FERMAT: An Alternative to Accuracy for Numerical Reasoning. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). ACL.
- ▹ Designed a mathematically informative multi-view test sets for numerical reasoning.
- ▹ Published a method for automatic data-augmentation of worded arithmetic problem.
- ▹ Demonstrated that diversity in language significantly improved performance.
- ▹ Explored data-leakage to justify improvement of BART and FLAN type LLMs.

Dec 21 (IALP-IEEE) **Sivakumar, J.**, et. al. (2021, December). A GRU-based pipeline approach for word-sentence segmentation and punctuation restoration in English. In 2021 International Conference on Asian Language Processing (IALP). IEEE.
- ▹ Trained GRU models using PyTorch for punctuation retrieval of concatenated strings.
- ▹ Created a binary classification system to identify insertions of punctuation.
- ▹ Generated synthetic training and testing data from punctuated text.
- ▹ Investigated automatic evaluations metrics against human perception.

RESEARCH ACTIVITIES

Sep 21 - Present **NLP for Endangered Language Revitalisation in Colombia**
- ▹ Collected and digitised text using OCR for Palenquero, an endangered language.
- ▹ Deploying low-resource language NLP research to develop pedagogical resource.

Nov 21 - Jun 22 **Speech Technology and NLP to improve Oral History Search functionality**
- ▹ Trained wav2vec ASR models on Legasee's oral history video archive i.e elderly speech.
- ▹ Explored existing NER systems to identify domain specific terms.
- ▹ Automated production of annotated transcripts to improve search and retrieval of videos based on timestamps.

Sep 20 - Sep 21 **Semantic Sentence Embeddings for Natural Language Inference**
- ▹ Trained BERT embeddings using siamese network contrastive loss over NLI tasks.
- ▹ Inductively biased embeddings with semantic parse trees over Graph Neural Network.

Sep 15 - May 16 **A Mathematical Approach at Analysing the Influence of English on French**
- ▹ Generated a diachronic dataset of French news articles to observe language evolution.
- ▹ Modelled adoption of English lexica into French with stochastic/differential equations.

SKILLS

Programming

- ▹ Python (PyTorch, Huggingface, scikit-learn, NLTK)
- ▹ MATLAB
- ▹ Bash (HPC)
- ▹ Version control (git/github)
- ▹ Latex

Languages

- ▹ French (Native)
- ▹ Tamil (Native)
- ▹ English (Near Native)
- ▹ Spanish (Fluent)
- ▹ German (Conversational)

ACADEMIC ACTIVITIES

Nov 23 - Present **HuggingFace Workshop for Undergraduates**
- ▹ Teaching the latest deep learning techniques both theoretically and practically.

Jan 24 **Reviewer for LREC-Coling 2024**
- ▹ Reviewed for Knowledge Discovery/Representation track.

Jun 23 **Invited Talk** - 3rd Speech and Language Technology CDT Conference
- ▹ Presented FERMAT research in main talk event.

Mar 23 - May 23 **Supervisor for final year research project in number representation**
- ▹ "How much does the number representation matter in downstream applications?"

Nov 22 - Feb 23 **Examiner for Text Processing and Professional Issues**
- ▹ Assessed student code and marked reports, provided feedback on assignments.

Jul 22 **12th Lisbon Machine Learning Summer School**
- ▹ Attended lectures on linear classifiers, seq2seq models, neural networks and talks from industry and academia on novel research, networked with other attendees.

Jan 22 - Jun 22 **Teaching Assistant for Foundations of Computer Science**
- ▹ Taught fundamental mathematics: probability, linear algebra and number theory.

Jan 21 - May 21 **Machine Learning Tutorial for Postgraduate Computational Linguists**
- ▹ Explained mathematics using linguistics examples for kNN, decision trees, Naive Bayes, linear regressions, regularisation, beam search, EM algorithm, gradient descent, entropy.