Karan Taneja

karantaneja@gatech.edu • linkedin.com/in/karantaneja1/

Education and Research Experience

- **Georgia Institute of Technology (Georgia Tech)** — Atlanta, US
  PhD+MS in Computer Science, *Advisor*: Prof. Ashok Goel, GPA: 4.00 /4, *Expected Graduation*: May'26 — *Jan'21 – Present*
  • LLM-BASED VIRTUAL TEACHING ASSISTANTS. Developed and evaluated an AI Teaching Assistant with a focus on document-grounding, safety, and use of publicly-available resources. Deployed in 20+ classes across multiple institutions. Analysis showed lower hallucinations, lower toxicity, and improved student perception of courses. See AIED'24, ITS'24, L@S'24.
  • MACHINE TEACHING. Developed a continual learning pipeline for LLM-based AI agents using active label correction and a framework for designing human-AI collaborative annotation tools. Results show oracle performance can be achieved with a tiny fraction of data annotations and a 16% improvement over previous methods. See EMNLP'24, CHAI-IJCAI'22, arXiv Preprint.
- **Indian Institute of Technology Bombay (IITB)** — Mumbai, India
  BS+MS in Electrical Engineering, *Minor* in Computer Science, GPA: 9.63 /10 — *Jul'15 – Jun'20*
  • DEEP LEARNING FOR F-MRI RECONSTRUCTION. Developed CNN-based pipeline for Bayesian reconstruction of 3D under-sampled fMRI scans with uncertainty prediction. Results showed quantitative and qualitative improvement over SOTA in terms of connectivity maps for many functional networks. Advised by Prof. Suyash Awate & Shabbir Merchant. See ICPR'20.
  • AWARDS. Institute Academic Prize (2019) and Undergraduate Research Award (2020)

Selected Publications (See Google Scholar for the complete list.)

1. Karan Taneja, Ashok Goel, **Can Active Label Correction Improve LLM-based Modular AI Systems?**, *Empirical Methods in Natural Language Processing (EMNLP-Main) 2024* [pdf]
2. Karan Taneja, Pratyusha Maiti, Sandeep Kakar, Pranav Guruprasad, Sanjeev Rao, Ashok Goel, **Jill Watson: A Virtual Teaching Assistant powered by ChatGPT**, *Artificial Intelligence in Education (AIED) 2024* [pdf]
3. Karan Taneja, Richard Segal, Richard Goodwin, **Monte Carlo Tree Search for Recipe Generation using GPT-2**, *International Conference on Computational Creativity (ICCC) 2023* [pdf]

Industry Experience

- **IBM Thomas J. Watson Research Center, *Research Intern*** — Yorktown Heights (NY), US
  Recipe Generation with Text Constraints, work with Dr. Richard Goodwin & Dr. Richard Segal — *May'22 – Aug'22*
  Developed RecipeMC, an LLM for recipe generation and Monte Carlo Tree Search-based inference method, to implement constraints on text generation, and evaluated the method to show improved credibility of generated recipes. Results showed that human evaluators preferred RecipeMC most often over real recipes compared to other strong baselines. See ICCC'23.
- **Microsoft India Development Center, *Research Intern & Collaborator*** — Hyderabad, India
  English-Hindi Code-mixed Speech Recognition, collaboration with Prof. Preethi Jyothi at IIT Bombay. — *Oct'18 – Feb'20*
  Developed and compared algorithms for synthesizing English-Hindi code-mixed datasets using monolingual datasets and text-to-speech systems to train code-mixed speech recognition and language models. Results showed significant improvements in word error rate and code-switching points using our synthetic English-Hindi speech utterances. See InterSpeech'19, InterSpeech'20.
- **Sony Semiconductor Solutions, *Research Intern*** — Atsugi, Japan
  Object Detection for 360-degree Images with Spherical CNNs — *May'18 – Jul'18*
  Enabled the use of previously infeasible deep architectures for 360-degree images in a software product by developing computationally light spherical CNNs for S^2 and $SO(3)$ group by factorization into depth-wise and point-wise spherical convolutions.
- **Philips Innovation Campus, *Software Intern*** — Bengaluru, India
  Patient-motion Detection in MRI Scans with Signal Processing & Deep Learning — *May'17 – Jul'17*
  Implemented and evaluated 3D CNN, cascaded 2D CNN-RNN, and several feature-based classifiers that used wavelet and spectral analysis for the task of patient motion detection for Philips MRI scans.

Skills and Courses

Expertise	AI Product Design | Natural Language Processing | Educational Technology | Research & Experimental Design | UI/UX Design | Probability & Statistics | Technical Writing & Communication | Leadership & Management
Coding	Python | Java | C++ | MATLAB/Octave | Web Development (HTML5, CSS3, JavaScript)
Packages	HuggingFace, SpaCy, LangChain, NLTK | Django, MongoDB, Google Cloud | NumPy, Pandas, Matplotlib | PyTorch, TensorFlow, Keras, Scikit-Learn | Streamlit, ReactJS, OpenCV | Git, Docker, Linux, Bash
Courses	ML, Advanced ML, RL, Knowledge-based AI | NLP, ASR | CV, Graphics | Optimization, Matrix Computations | Data Analysis, Markov Chains, Graphical Models | Qualitative Methods (HCI) , EdTech, Game Design

Leadership and Volunteering

- **Co-president, Asha for Education, Atlanta Chapter** (May'23 – Apr'24) Managed large-scale events with 1000+ participants with a team of 20+ volunteers, and fundraised over USD 100K for 6 NGOs focused on the education of underprivileged.
- **Reviewer** ACL Rolling Reviews 2023 | AAAI Special Award Judge at International Science and Engineering Fair 2022 | International Conference on Pattern Recognition 2020
- **Teaching Assistant** GEORGIA TECH: Knowledge-based Artificial Intelligence | IITB: Machine Learning, Data Analysis and Interpretation (×2), Electronic Design Lab