Kausik Lakkaraju

Linkedin: https://www.linkedin.com/in/kausik-lakkaraju-8a29833a
Github: https://github.com/kausik-l
Webpage: https://kausik-l.github.io/

Email : kausik0123456789@gmail.com
Mobile : +1-803-238-0960

EDUCATION

Artificial Intelligence (AI) Institute, University of South Carolina — Columbia, SC (USA)
Doctor of Philosophy (Ph.D.) in Computer Science & Engineering; GPA: 3.81 — December 2025 (Expected)
Courses: *Trusted AI, Neural Networks, Bayesian Networks, and Decision Graphs,*
Theory of Computation, Foundation of Analysis 1 (MATH)

University of South Carolina — Columbia, SC (USA)
Master of Science in Computer Science & Engineering; GPA: 3.87 — May 2021
Courses: *Artificial Intelligence, Data Mining and Warehousing, Natural Language Processing,*
Deep Reinforcement Learning, Robotics, Big Data Analytics, Database System Design,
Analysis of Algorithms, Computer Architecture

Osmania University — Hyderabad, Telangana (India)
Bachelor of Engineering in Electronics and Communication & Engineering; Percentage: 76% — May 2019
Major Project: *Multiple color line follower robot with obstacle detection*

SKILLS SUMMARY

- **Languages**: Python, JavaScript, C++, C, SQL, R.
- **Frameworks**: Tensorflow, PyTorch, Pandas, Matplotlib, Numpy, Keras.
- **Areas of Research**: Causal Inference, Trusted AI, Foundation Models, Generative AI, Multi-modal Decision Support, Chatbot Development, Rating of AI Systems, AI Fairness, Explainable AI (XAI), Machine Learning

RELEVANT RESEARCH EXPERIENCE

Mayo Clinic — Rochester, MN, USA
Data Science Intern - Dr. Hamid R. Tizhoosh — Sep 2024 - Present
- **Segmentation of CT scans**: We are working on segmenting CT scans using segmentation models like U-Net.
- **Evaluating patching algorithms**: Patching algorithms are used for efficient search and retrieval of digital histopathology data. I am working on evaluating such algorithms under different settings.

University of South Carolina — Columbia, SC, USA
Graduate Research Assistant - Dr. Biplav Srivastava — Jan 2021 - Present
- **Rating AI Systems for Bias**:
  - *Rating of AI Systems through a Causal Lens (Ph.D. thesis topic.):* As most AI systems are known for giving biased or uncertain outputs, we proposed a causality-based method to rate AI systems for bias. We assessed text-based sentiment analysis systems so far and built a tool using the Django framework that aids users in rating their own AI systems. We are currently working on evaluating multi-modal systems like CLIP.
- **Chatbot Testing**:
  - *LLM Testing:* We assessed LLM-based chatbots, ChatGPT and Bard, for efficacy and fairness. Currently, we are working on building a principled approach to evaluate LLMs for bias.
  - *SafeChat Framework:* The SafeChat framework we built enables users to build safe and trustworthy chatbots. We performed experiments and showed that chatbots built using SafeChat outperform LLM-based ChatGPT and Bard in the personal finance domain. We are currently adding more trust services to the framework including the rating service.

SELECTED PUBLICATIONS

Research Papers

- K. Lakkaraju, B. Srivastava and M. Valtorta, "Rating Sentiment Analysis Systems for Bias Through a Causal Lens," in IEEE Transactions on Technology and Society, doi: 10.1109/TTS.2024.3375519.

- K. Lakkaraju, A. Gupta, B. Srivastava, M. Valtorta and D. Wu, "The Effect of Human v/s Synthetic Test Data and Round-Tripping on Assessment of Sentiment Analysis Systems for Bias," in 2023 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), Atlanta, GA, USA, 2023 pp. 380-389. doi: 10.1109/TPS-ISA58951.2023.00053

- Kausik Lakkaraju, Sara E Jones, Sai Krishna Revanth Vuruma, Vishal Pallagani, Bharath C Muppasani, and Biplav Srivastava. 2023. LLMs for Financial Advisement: A Fairness and Efficacy Study in Personal Decision Making. In Proceedings of the Fourth ACM International Conference on AI in Finance (ICAIF '23). Association for Computing Machinery, New York, NY, USA, 100–107. https://doi.org/10.1145/3604237.3626867

- Lakkaraju, K., 2022. Why is my System Biased?: Rating of AI Systems through a Causal Lens. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22). Association for Computing Machinery, New York, NY, USA, 902. https://doi.org/10.1145/3514094.3539556

- Lakkaraju, K., Hassan, T., Khandelwal, V., Singh, P., Bradley, C., Shah, R., Agostinelli, F., Srivastava, B., & Wu, D. (2022). *ALLURE: A Multi-Modal Guided Environment for Helping Children Learn to Solve a Rubik's Cube with Automatic Solving and Interactive Explanations.* Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 13185-13187. https://ojs.aaai.org/index.php/AAAI/article/view/21722

- Lakkaraju, K., Palaiya, V., Paladi, S.T., Appajigowda, C., Srivastava, B., Johri, L. (2022). Data-Based Insights for the Masses: Scaling Natural Language Querying to Middleware Data. In: , et al. Database Systems for Advanced Applications. DASFAA 2022. Lecture Notes in Computer Science, vol 13247. Springer, Cham. $https://doi.org/10.1007/978-3-031-00129-1_49$

Patents

- Srivastava, Biplav, Kausik Lakkaraju, Revathy Venkataramanan, Vishal Pallagani, Vedant Khandelwal, and Hong Yung Yip. "Robust useful and general task-oriented virtual assistants." U.S. Patent Application 17/714,508, filed November 10, 2022.

- Srivastava, Biplav, Kausik Lakkaraju, and Marco Valtorta. "Assigning trust rating to ai services using causal impact analysis." U.S. Patent Application No. 18/448,369.

Academic & Personal Projects

- **YOLOR Vs. YOLOv6 Face-off: A Comparison of SOTA Object Recognition Models (Dec '22)**: We deep-dived into the working of YOLOR and YOLOv6 object recognition models, which were released in 2022, and evaluated them on specific tasks to compare their performance. We tested the original models and also tested them after fine-tuning them on a different blood cells dataset which classifies different blood cells. **Keywords**: Computer Vision, Object Recognition.
- **Plant Disease Explanation (Dec '21)**: Trained neural network models like CNN and InceptionV3 (using transfer learning) on two well-known datasets, PlantDoc and PlantVillage to predict the plant name and plant disease (if it is diseased). In addition to this, the LIME explainer was used to provide the explanation for why the system had to make a particular decision for a certain sample. **Keywords**: Trusted AI, XAI, Deep Learning.
- **Explainable Pet Class Predicting System (Dec '21)**: Used the 'Pawpularity' dataset from Kaggle to train neural network models to predict a 'Pawpularity score' based on how cute a particular image of the pet is. LIME was used to provide an explanation for the predictions made by the models. **Keywords**: Trusted AI, XAI, Deep Learning.
- **Deep Reinforcement Learning Based Chatbot (Dec '20)**: Trained a Seq2Seq model using the 'Self-Critical Sequence Training (SCST)' method on movie dialogs dataset to create a chatbot (for chitchat). **Keywords**: Deep Reinforcement Learning, Chatbot.

Certifications & Prizes

- Awarded $ 1,000 NSF student travel grant to attend the IEEE TPS 2023 conference.

- Recipient of the best graduate student poster presentation award at a university-wide event that was held at the University of South Carolina in April 2023 for a poster presented on the topic, 'Rating of AI Systems through a Causal Lens'.

- Presented a student track paper (on Ph.D. dissertation topic) at the AIES 2022 conference which was held at the University of Oxford (UK).

- Secured first prize in ITT (Innovation Think Tank) certification program for two consecutive years.