

# Kausik Lakkaraju

Ph.D. Student, Graduate Research Assistant, University of South Carolina, Columbia, SC-29201, USA

LinkedIn: <https://www.linkedin.com/in/kausik-lakkaraju-8a29833a>

Google Scholar: <https://scholar.google.com/citations?user=Cn0qKmYAAAAJ&hl=en>

Webpage: <https://kausik-l.github.io/>

Email: [kausik0123456789@gmail.com](mailto:kausik0123456789@gmail.com)

Phone No.: +1-803-238-0960

## RESEARCH INTERESTS

---

Ethical AI, AI Fairness, eXplainable AI (XAI), AI Robustness, Causal Reasoning, Time-series Forecasting, Multi-modal AI, Chatbot Development and Testing

## EDUCATION

---

**Artificial Intelligence (AI) Institute, University of South Carolina, Columbia, SC (USA)** December 2025  
Ph.D. in Computer Science Cumulative GPA: 3.81/4.00  
Thesis Title: Rating of AI Models through a Causal Lens

**University of South Carolina, Columbia, SC (USA)** May 2021  
MS in Computer Science Cumulative GPA: 3.87/4.00

**Osmania University, Hyderabad, India** May 2019  
B.E. in Electronics and Communication Percentage: 76 %

## WORK EXPERIENCE

---

**University of South Carolina** Columbia, SC, USA  
*Graduate Research Assistant - Dr. Biplav Srivastava* January 2021 — Present

- *Rating of AI Systems through a Causal Lens (Ph.D. thesis topic):* As most AI systems are known for giving biased outputs, we proposed a causality-based method to rate AI systems for robustness under different settings. I worked on assessing text-based sentiment analysis systems, multi-modal and foundation model based time-series forecasting models so far, and built a tool called ARC that aids users in rating AI systems. I am currently exploring how existing XAI methods can complement and improve our rating approach.
- *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 robustness.
- *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 chatbots in the personal finance domain. I am currently adding more trust services to the framework including the rating service.
- *IoT-based Chatbot:* I worked with a company called Tantiv4 to build a RASA-based chatbot that allows users to control their network usage and bandwidth using smart routers in a household or office setting. I helped build another chatbot that helps users monitor power usage in a household, office, or university setting using smart sensors. We, previously, published these chatbots as separate skills on Amazon devices.

**University of South Carolina** Columbia, SC, USA  
*Graduate Teaching Assistant (Artificial Intelligence) - Dr. Biplav Srivastava* Aug 2025 — Dec 2025

- Evaluating students' quizzes, providing detailed feedback to support learning in general AI concepts with more emphasis on eXplainable AI (XAI).

**University of South Carolina** Columbia, SC, USA  
*Graduate Teaching Assistant (Computer Architecture) - Dr. Rasha Karakchi* Jan 2025 — May 2025

- Evaluated students' projects and classwork, providing detailed feedback to support learning in computer architecture concepts related to MIPS assembly language programming.

**Mayo Clinic** Rochester, MN, USA  
*Data Science Intern - Dr. Hamid Tizhoosh* Sept 2024 — Jan 2025

- *Segmentation of CT scans:* I worked on segmenting CT scans using foundation model based segmentation models like SAM2.
- *Evaluating patching algorithms:* Patching algorithms are designed for the efficient search and retrieval of digital histopathology data. I evaluated these algorithms using foundation models as the backbone, testing their performance across various settings.

**University of South Carolina**

*Graduate Teaching Assistant (Artificial Intelligence) - Dr. Biplav Srivastava*

Columbia, SC, USA

Jan 2024 — May 2024

- Graded students' assignments and exams for the AI course.
- Held weekly office hours to assist students with conceptual understanding, project guidance, and debugging support.

**University of South Carolina**

*Graduate Research Assistant - Dr. Dezhi Wu*

Columbia, SC, USA

May 2020 — May 2021

- *Full-stack Development:* Involved understanding a clustering algorithm created by a Professor from the Physics department at our university. The project involved developing a UI using the ReactJS framework and building the back-end using the Django REST framework.
- *Front-end Development:* Involved building a website for a HCI experiment conducted by my advisor. I used ReactJS to build the front-end.

**University of South Carolina**

*Graduate Teaching Assistant - Ms. Catherine Matthews*

Columbia, SC, USA

Jan 2020 — Apr 2020

- *Front-end Development:* Helped students solve some assignment problems involving HTML, CSS, and Javascript.
- *Python Programming:* Helped students solve assignment problems involving Python programming.

**ECIL**

*Intern*

Hyderabad, Telangana, India

May 2018 — June 2018

- *Home-based Automation System using GSM Module:* With the device we built, one could control (turn any device ON and OFF) by just sending a simple text message with their phone, even when they are away from their home. GSM module converts the text message that user sends to a format which is understandable by the AT89S52 Micro-controller we used. This micro-controller controls the devices based on the command.

---

## PUBLICATIONS

### Journal

#### 2024

- **Lakkaraju, Kausik**, Biplav Srivastava, and Marco Valtorta. "Rating sentiment analysis systems for bias through a causal lens." *IEEE Transactions on Technology and Society* (2024).
- Srivastava, B., **Lakkaraju, K.**, Bernagozzi, M., & Valtorta, M. (2024). Advances in automatically rating the trustworthiness of text processing services. *AI and Ethics*, 4(1), 5-13.

### Conference

#### 2025

- [Accepted, not yet published] Gupta, N., Koppisetti, P., **Lakkaraju, K.**, & Srivastava, B. (2026, February). GAIco: A Deployed and Extensible Framework for Evaluating Diverse and Multimodal Generative AI Outputs. In Proceedings of the AAAI Conference on Artificial Intelligence.
- [Accepted, not yet published] Widener, M., **Lakkaraju, K.**, Aydin, J., & Srivastava, B. (2025). On identifying why and when foundation models perform well on time-series forecasting using automated explanations and rating. In Proceedings of the AAAI Fall Symposium on AI Trustworthiness and Risk Assessment for Challenged Contexts (ATRACC 2025). AAAI Press.
- **Lakkaraju, K.**(2025, August). Rating AI models for robustness through a causal lens. In J. Kwok (Ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)*, Doctoral Consortium (pp. 10975–10976). International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2025/1242>

#### 2023

- **Lakkaraju, K.**, Jones, S. E., Vuruma, S. K. R., Pallagani, V., Muppasani, B. C., & Srivastava, B. (2023, November). 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 (pp. 100-107).
- **Lakkaraju, K.**, Gupta, A., Srivastava, B., Valtorta, M., & Wu, D. (2023, November). 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) (pp. 380-389). IEEE.

## 2022

- **Lakkaraju, K.** (2022, July). 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 (pp. 902-902).

## Published Articles and Book Chapters

### 2025

- **Lakkaraju, K.**, Muppasani, B., Jones, S. E., & Srivastava, B. (2025). A Dataset and Visualization of Generalizable Election-Related Questions Compiled from Leading Global Democracies for Building AI-Enabled Tools. In PROMISE-PROMoting AI's Safe usage for Elections (pp. 105-113). Cham: Springer Nature Switzerland.
- Ayisi, A., Deepak, P., Smith, M., Srivastava, B., Nikolich, A., Hickerson, A., ... & **Lakkaraju, K.** (2025). Towards Better Elections: A Discussion About the United Kingdom and Africa. In PROMISE-PROMoting AI's Safe usage for Elections (pp. 197-207). Cham: Springer Nature Switzerland.

### 2023

- Muppasani, B., Pallagani, V., **Lakkaraju, K.**, Lei, S., Srivastava, B., Robertson, B., ... & Narayanan, V. (2023). On safe and usable chatbots for promoting voter participation. *AI Magazine*, 44(3), 240-247.

## Demos

### 2025

- [Accepted, not yet published] Koppisetti, P., Gupta, N., **Lakkaraju, K.**, & Srivastava, B. (2026, February). GAICo: Demonstrating a Unified Framework for Multi-Modal GenAI Evaluation. In Proceedings of the AAAI Conference on Artificial Intelligence.

### 2022

- **Lakkaraju, K.**, Hassan, T., Khandelwal, V., Singh, P., Bradley, C., Shah, R., ... & Wu, D. (2022, June). Allure: A multi-modal guided environment for helping children learn to solve a rubik's cube with automatic solving and interactive explanations. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 11, pp. 13185-13187).
- **Lakkaraju, K.**, Palaiya, V., Paladi, S. T., Appajigowda, C., Srivastava, B., & Johri, L. (2022, April). Data-Based Insights for the Masses: Scaling Natural Language Querying to Middleware Data. In International Conference on Database Systems for Advanced Applications (pp. 527-531).

## Workshop

### 2025

- **[IJCAI User-Aligned Assessment of Adaptive AI Systems 2025]** **Lakkaraju, K.**, Valluru, S. L., Srivastava, B., & Valtorta, M. ARC: A Tool to Rate AI Models for Robustness Through a Causal Lens. In IJCAI 2025 Workshop on User-Aligned Assessment of Adaptive AI Systems.
- **[IEEE ICDM International Workshop on AI for Nudging and Personalization (WAIN)]** Nagpal, V., Valluru, S. L., **Lakkaraju, K.**, & Srivastava, B. (2024). BEACON: Balancing Convenience and Nutrition in Meals With Long-Term Group Recommendations and Reasoning on Multimodal Recipes. arXiv preprint arXiv:2406.13714.

### 2024

- **[ICML TEACH 2024]** **Lakkaraju, K.**, Khandelwal, V., Srivastava, B., Agostinelli, F., Tang, H., Singh, P., ... & Kundu, A. (2024). Trust and ethical considerations in a multi-modal, explainable AI-driven chatbot tutoring system: The case of collaboratively solving Rubik's Cube. CoRR.

### 2023

- **[ICAPS FinPlan 2023]** **Lakkaraju, K.**, Vuruma, S. K. R., Pallagani, V., Muppasani, B., & Srivastava, B. Can LLMs be Good Financial Advisors?: An Initial Study in Personal Decision Making for Optimized Outcomes. FinPlan 2023, 51.

## 2022

- [ICAPS KEPS 2022] Pallagani, V., Ramamurthy, P., Khandelwal, V., Venkataramanan, R., **Lakkaraju, K.**, Aakur, S. N., & Srivastava, B. (2022). A Rich Recipe Representation as Plan to Support Expressive Multi Modal Queries on Recipe Content and Preparation Process. arXiv preprint arXiv:2203.17109.
- [ICAPS XAI 2022] Agostinelli, F., Panta, R., Khandelwal, V., Srivastava, B., Muppasani, B. C., **Lakkaraju, K.**, & Wu, D. (2022). Explainable pathfinding for inscrutable planners with inductive logic programming. In ICAPS 2022 Workshop on Explainable AI Planning.

## Manuscripts

- [Under Review] **Lakkaraju, K.**, Valluru, S. L., & Srivastava, B. (2025). Holistic Explainable AI (H-XAI): Extending Transparency Beyond Developers in AI-Driven Decision Making. arXiv preprint arXiv:2508.05792.
- [Under Review] Srivastava, B., **Lakkaraju, K.**, Gupta, N., Nagpal, V., Muppasani, B. C., & Jones, S. E. (2025). SafeChat: A Framework for Building Trustworthy Collaborative Assistants and a Case Study of its Usefulness. arXiv preprint arXiv:2504.07995.
- [Under Review] **Lakkaraju, K.**, Kaur, R., Zehtabi, P., Patra, S., Valluru, S. L., Zeng, Z., ... & Valtorta, M. (2025). On Creating a Causally Grounded Usable Rating Method for Assessing the Robustness of Foundation Models Supporting Time Series. arXiv preprint arXiv:2502.12226.
- [Under Review] **Lakkaraju, K.**, Kaur, R., Zeng, Z., Zehtabi, P., Patra, S., Srivastava, B., & Valtorta, M. (2024). Rating Multi-Modal Time-Series Forecasting Models (MM-TSFM) for Robustness Through a Causal Lens. arXiv preprint arXiv:2406.12908.
- Muppasani, B., **Lakkaraju, K.**, Gupta, N., Nagpal, V., Jones, S., & Srivastava, B. (2025). ElectionBot-SC: A Tool to Understand and Compare Chatbot Behavior for Safe Election Information in South Carolina.
- Nagpal, V., Valluru, S. L., **Lakkaraju, K.**, Gupta, N., Abdulrahman, Z., Davison, A., & Srivastava, B. (2024). A Novel Approach to Balance Convenience and Nutrition in Meals With Long-Term Group Recommendations and Reasoning on Multimodal Recipes and its Implementation in BEACON. arXiv preprint arXiv:2412.17910.
- Srivastava, B., **Lakkaraju, K.**, Koppel, T., Narayanan, V., Kundu, A., & Joshi, S. (2023). Evaluating Chatbots to Promote Users' Trust-Practices and Open Problems. arXiv preprint arXiv:2309.05680.
- Khandelwal, V., **Lakkaraju, K.**, Pallagani, V., Ramamurthy, P., Venkataramanan, R., Aakur, S. N., & Srivastava, B. (2022). A Multi-Modal Decision Support System with Allergy-Aware Recipe Understanding Powered by a Plan Representation.

## 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 12,067,983, issued August 20, 2024.
- Srivastava, Biplav, Vishal Pallagani, Revathy Chandrasekaran Venka, Vedant Khandelwal, and Kausik Lakkaraju. "Multimodal retrieval and execution monitoring using rich recipe representation." U.S. Patent 12,332,873, issued June 17, 2025.
- Srivastava, Biplav, Kausik Lakkaraju, and Marco Valtorta. "Assigning trust rating to ai services using causal impact analysis." U.S. Patent Application 18/448,369, filed February 22, 2024.

## ACADEMIC AND PERSONAL PROJECTS

---

- **YOLOR Vs. YOLOv6 Face-off: A Comparison of SOTA Object Recognition Models** Dec 2022  
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.
- **Causal Robot** May 2022  
I examined the capabilities and limitations of CausalWorld framework which allows us perform various interventions on the environment variables. **Keywords:** Causal Reasoning, Reinforcement Learning, Robotics.
- **Plant Disease Explanation** Dec 2021  
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 2021  
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.

- **Indian Sign Language Recognition using OpenCV** Dec 2020  
The project allows one to create an Indian Sign Language (ISL) dataset on their own in a few minutes. A CNN model was trained on one such dataset. It was able to achieve a test accuracy of more than 98%. **Keywords:** Computer Vision, Deep Learning, Image Processing.
- **Deep Reinforcement Learning Based Chatbot** Dec 2020  
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.
- **Covid-19 Tracker** July 2022  
Built a UI using ReactJS for tracking worldwide and country-wise COVID cases and deployed it using Firebase. (click here). **Keywords:** Front-end, ReactJS.
- **Obstacle Avoidance using 77 GHz RADAR module** Mar 2020  
Constructed a Robotic Operating System (ROS) package for 77GHz Radar module. **Keywords:** Robotics, ROS, C++.
- **Multiple Color Line Following Robot with Obstacle Detection** Apr 2019  
We constructed a multiple color line following robot which could detect obstacles on its way. We used wireless technology module called HC-12 and ultrasonic sensors for obstacle detection. **Keywords:** Robotics, Obstacle detection.
- **Black Line Following Robot** Oct 2017  
A simple black line follower which follows a black line. **Keywords:** Robotics.
- **Mobile Controlled Robot** Oct 2017  
We created a mobile controlled robot using 8051 Microcontroller. **Keywords:** Robotics.

## SELECTED COURSES

---

- **Ph.D. in Computer Science:** Trusted AI, Neural Networks, Bayesian Networks, and Decision Graphs, Theory of Computation, Foundation of Analysis 1 (MATH).
- **MS in Computer Science:** Artificial Intelligence, Data Mining and Warehousing, Natural Language Processing, Deep Reinforcement Learning, Robotics, Big Data Analytics, Database System Design, Analysis of Algorithms, Computer Architecture.

## ACHIEVEMENTS

---

- Presented my research at IJCAI 2025 Doctoral Consortium in Montreal, Canada with travel support from NSF.
- Presented my research at ACM FAccT 2025 Doctoral Consortium held in Athens, Greece with full travel support.
- Conducted a tutorial on the topic 'Evaluating and Rating AI Systems for Trust and its Applications to Finance' at the ACM ICAIF 2024 conference.
- 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) conducted by *Siemens Healthineers* certification program for two consecutive years.
- **Certifications:** Building Basic GANs (Coursera), Python for Data Science (IBM), Deep Learning A-Z (Udemy), Machine Learning (A-Z)(Udemy), Django Full Stack Web Development (Udemy).

## PROFESSIONAL SERVICE

---

- **Program Committee Member:** AAAI / ACM Conference on Artificial Intelligence, Ethics, and Society (2024, 2025), Annual AAAI Conference on Artificial Intelligence (Main track 2026, special track on AI alignment 2026, IAAI 2026).
- **Journal Reviewer:** IEEE Transactions on Neural Networks and Learning Systems (2023, 2025) IEEE Transactions on Technology and Society (2024), IEEE Internet Computing (2024).
- **Conference Reviewer:** International Joint Conferences on Artificial Intelligence (2024)
- **Workshop Reviewer:** TEACH Workshop at International Conference on Machine Learning (2023)

## OTHER EXPERIENCES

---

- [Sep 2021 — May 2023] President of Indian Student Organization (ISO) at the University of South Carolina.
- [Aug 2021] Vice President of Indian Student Organization (ISO) at the University of South Carolina.
- [Aug 2021 — July 2021] Public Relations Coordinator of Indian Student Organization (ISO) at the University of South Carolina.
- [May 2018 — May 2019] Started a book club during my undergrad with 50+ people.

## SKILLS

---

- **Programming Languages:** Python, JavaScript, C++, C, MySQL, R.
- **Frameworks:** Tensorflow, PyTorch, Pandas, Matplotlib, Numpy, Keras, Tableau, Power BI, Hadoop.
- **Areas of Research:** Causal Inference, Trusted AI, Foundation Models, AI Fairness, Time-Series Forecasting, Analytics, Statistics, Generative AI, Multi-modal Decision Support, Chatbot Development, Rating of AI Systems, Explainable AI (XAI), AI in Finance, Machine Learning, Deep Learning, Image Segmentation.

## REFERENCES

---

### Prof. Biplov Srivastava

*Professor, AI Institute, Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA*  
E-mail: biplav.s@sc.edu

Personal Page: <https://sites.google.com/site/biplavsrivastava>  
Google Scholar: <https://scholar.google.com/citations?user=mPC6wp4AAAAJ&hl=en>  
LinkedIn: <https://www.linkedin.com/in/biplav-srivastava>

### Prof. Dezhi Wu

*Professor, Department of Integrated Information Technology, University of South Carolina, Columbia, SC, USA*  
E-mail: dezhiwu@cec.sc.edu  
Personal Page: <https://sites.google.com/view/wudezhi/>  
Google Scholar: <https://scholar.google.com/citations?user=3Xt6B50AAAAJ&hl=en>

### Prof. Marco Valtorta

*Professor, Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA*  
E-mail: mgv@cse.sc.edu  
Personal Page: <https://www.cse.sc.edu/~mgv/>  
Google Scholar: <https://scholar.google.com/citations?user=l53UPjoAAAAJ&hl=en>  
LinkedIn: <https://www.linkedin.com/in/marco-valtorta-b990534>

### Dr. Sunandita Patra

*AI Research Lead, JP Morgan AI Research*  
E-mail: sunandita.patra@jpmchase.com  
Google Scholar: <https://scholar.google.co.in/citations?user=VMTRIgoAAAAJ&hl=en>  
LinkedIn: <https://www.linkedin.com/in/sunandita-patra-84b16b35>

### Dr. Dave Freeman

*Neurologist, Mayo Clinic, Jacksonville, Florida*  
E-mail: freeman.william1@mayo.edu  
Personal Page: <https://www.mayoclinic.org/biographies/freeman-william-d-m-d/bio-20054822>  
LinkedIn: <https://www.linkedin.com/in/wdfreemanmd1>