

Ashutosh Chaubey

✉ achaubey@usc.edu
🏠 www.ashutoshchaubey.in
🎓 [Ashutosh Chaubey](#)
📄 [ashutosh-chaubey](#)

Education

- 2024 - 2029 **PhD in Computer Science**, *University of Southern California, Los Angeles*.
GPA 4.0/4.0, Advisor - [Prof. Mohammad Soleymani](#)
- 2017 - 2021 **BS in Computer Science**, *Indian Institute of Technology, Roorkee*.
GPA 9.718/10.0 - **Third-highest GPA** amongst all the graduating students of IIT Roorkee 2021

Publications

- 2025 **Face-LLaVA : Facial Expression and Attribute Understanding through Instruction Tuning**.
Ashutosh Chaubey, Xulang Guan, Mohammad Soleymani
Winter Conference on Applications of Computer Vision (WACV) 2026 - Round 1 (6.4% accept.) [[Preprint](#)] [[Webpage](#)]
- 2025 **Can VLMs Recall Factual Associations From Visual References ?**.
Dhananjay Ashok, **Ashutosh Chaubey**, Hirona Arai, Jonathan May, Jesse Thomason
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2025 [[Preprint](#)]
- 2025 **DiTaiListener : Controllable High Fidelity Listener Video Generation with Diffusion**.
Maksim Siniukov, Di Chang, Minh Tran, Hongkun Gong, **Ashutosh Chaubey**, Mohammad Soleymani
International Conference on Computer Vision (ICCV) 2025 [[Preprint](#)] [[Webpage](#)]
- 2024 **ContextIQ : A Multimodal Expert-Based Video Retrieval System for Contextual Advertising**.
Ashutosh Chaubey, Anoubhav Agarwal, Sartaki Roy, Aayush Agrawal, Susmita Ghose
Winter Conference on Applications of Computer Vision (WACV) 2025 [[Paper](#)] [[Poster](#)]
- 2023 **Meta-Learning Framework for End-to-End Imposter Identification in Unseen Speaker Recognition**.
Ashutosh Chaubey, Sparsh Sinha, Susmita Ghose
IEEE Workshop on Automatic Speech and Understanding (ASRU) 2023 [[Paper](#)] [[Poster](#)]
- 2022 **Improved Relation Networks for End-to-End Speaker Verification and Identification**.
Ashutosh Chaubey, Sparsh Sinha, Susmita Ghose
Interspeech 2022 [[Paper](#)][[Poster](#)]
- 2022 **OPAD : An Optimized Policy-based Active Learning Framework for Document Content Analysis**.
Sumit Shekhar, Bhanu Prakash Reddy Guda, **Ashutosh Chaubey**, Ishan Jindal, Avneet Jain
CVPR 2022 Workshop on Fair, Data Efficient and Trusted Computer Vision [[Paper](#)] [[Patent](#)]
- 2020 **Universal Adversarial Perturbations : A Survey**.
Ashutosh Chaubey*, Nikhil Agrawal*, Kavya Barnwal, Keerat K. Guliani, Pramod Mehta
arXiv Preprint (50+ citations) [[Paper](#)]
- 2019 **A GAN-based Ensemble Technique for Automatic Evaluation of Machine Synthesized Speech**.
Ashutosh Chaubey*, Jaynil Jaiswal*, Bhimavarapu Sasi Kiran Reddy, Shashank Kashyap, Puneet Kumar, Raman Balasubramanian, Partha Pratim Roy
Asian Conference on Pattern Recognition (ACPR) 2019 [[Paper](#)] [[Poster](#)]

Research Experience

- Aug 2024 - Present **Graduate Researcher**, Institute for Creative Technologies, University of Southern California.
Advisor - [Prof. Mohammad Soleymani](#)
- Working on post-training approaches such as preference optimization (eg., DPO) for multimodal LLMs.
 - Worked on Face-LLaVA, a general VLM for different face analysis and face reasoning tasks, outperforming all open source SOTA VLMs on nine face analysis benchmarks. [[Paper \(WACV 2026 - Round 1\)](#)]
 - Worked on diffusion-based photorealistic listener behaviour animation using audio-visual speaker signals - improvements of upto 73% on photorealism and 6% on motion generation. [[Paper \(ICCV 2025\)](#)]
- Apr 2023 - Jul 2024 **Founding Research Engineer**, Anoki Inc..
Advisor - [Dr. Susmita Ghose](#)
- Worked with multimodal models for video retrieval using text, image, and audio. [[Paper \(WACV 2025\)](#)]
 - Proposed system achieves near 100% performance and outperforms heavier baselines on audio and video retrieval baselines. [[Patent-1](#)] [[Patent-2](#)] [[Patent-3](#)]
- Jul 2021 - Mar 2023 **Data Scientist**, LG Ad Solutions (formerly Alphonso Inc.).
Advisor - [Dr. Susmita Ghose](#)
- Worked with state-of-the-art speaker verification models such as RawNet and ECAPA-TDNN to exhibit their generalizability on near-/far-field audios, different languages, and under noisy environments.
 - Proposed a novel relation network-based pipeline for end-to-end speaker recognition, improving the baseline by up to 12% relatively. [[Paper \(Interspeech 2022\)](#)] [[Paper \(ASRU 2023\)](#)]

May 2020 - **Research Intern**, Big-data Experience Lab, Adobe Research.

Jul 2020 Advisor - [Dr. Sumit Shekhar](#)

- Used deep Q-learning to learn an optimal acquisition function for active learning by modeling the active learning cycle as a Markov Decision Process. [\[Paper \(CVPR 2022 Workshops\)\]](#)
- Reduced the annotation effort by using a weak learning setting where the annotator just has to verify the current model predictions on acquired samples. [\[US Patent App. 17/170,307\]](#)

Jan 2019 - **Research Intern**, Indian Institute of Science (IISc.), Bengaluru | Indian Institute of Technology, Roorkee.

Mar 2020 Advisors - [Prof. R Venkatesh Babu](#) | [Prof. R. Balasubramanian](#)

- Worked on multi-person human pose prediction using synthetic dual person dataset to mitigate data limitations.
- Worked on automatic evaluation of TTS and proposed a GAN-decoder-based scoring mechanism. [\[Paper \(ACPR 2019\)\]](#)
- Experimented with state-of-the-art universal adversarial perturbation techniques (attacks and defenses) and wrote a survey paper with over 50 citations. [\[Survey Paper\]](#)

Skills

Coding Languages - Python [Advanced], C++ [Intermediate]

Frameworks/Libraries - PyTorch, NumPy, Pandas, Transformers

Tools - VSCode, Git, Anaconda, Docker