☎ +1 (213)994-4146
☎ +91 72248 67609
✉ achaubey@usc.edu
www.ashutoshchaubey.in
Ashutosh Chaubey
ashutosh-chaubey

Ashutosh Chaubey

Overview

Areas of Interest. Multimodal LLMs, Post-training for Large Models, Multimodal Emotion and Social AI, Speech and Audio, Video Generation, Vision-Language Models

Summary. My research at USC aims to develop *multimodal social understanding* and *behavior generation* algorithms using *multimodal large language models* and large *diffusion models* respectively. Prior to USC, I have three years of industry experience on *speech understanding* and *multimodal content retrieval*.

Education

2024 - 2027 **PhD in Computer Science**, *University of Southern California, Los Angeles*.
GPA 4.0/4.0, Advisor - Prof. Mohammad Soleymani, Expected Graduation - 06/2027

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

Research Experience

Aug 2024 - Present **Graduate Researcher**, Institute for Creative Technologies, University of Southern California, [Part-time].
Advisor - Prof. Mohammad Soleymani

- Developed a new omni LLM preference optimization technique (**MoD-DPO**) to mitigate cross-modal hallucinations in omni LLMs – **relative improvements of 6-30%** on different cross-modal hallucination benchmarks. [CVPR 2026]
- Proposed **AVEm-DPO** for post-training multimodal LLMs (MLLMs) using preference optimization to enhance their emotion reasoning capabilities – performance **improvement of 6-19%** over different emotion benchmarks relative to the reference models. [ICLR 2026]
- Created **LibreFace-2.0**, an enhanced facial analysis toolkit with **diffusion-based synthetic data generation pipeline** to improve out-of-domain facial analysis – performance **improvement of 4-7%** over out-of-domain facial action unit benchmarks with ∼**20% reduced size**. [FG 2026]
- Worked on **Face-LLaVA**, a general vision-language model for different face analysis and face reasoning tasks, **outperforming all open source SOTA VLMs** on nine face analysis benchmarks. [WACV 2026]
- Evaluated **vision-language alignment** in MLLMs testing factual information retrieval and showed that probing internal states of the language model can reveal mis-alignment with **near 100% accuracy**. [EMNLP 2025]
- Proposed diffusion-based photorealistic listener behaviour animation using audio-visual speaker signals - **improvements of upto 73% on photorealism and 6% on motion generation**. [ICCV 2025]

Apr 2023 - Jul 2024 **Founding Research Engineer**, Anoki Inc, [Full-time].
Advisor - Dr. Susmita Ghose

- Worked on multimodal video retrieval using text, image, and audio – proposed system achieves **near 100% retrieval performance** with a much lighter framework compared to baselines [WACV 2025] [US Patent 1 2 3]

Jul 2021 - Mar 2023 **Data Scientist**, LG Ad Solutions (formerly Alphonso Inc.), [Full-time].
Advisor - Dr. Susmita Ghose

- Proposed a novel relation network-based pipeline for end-to-end speaker recognition, improving the baseline performance by **up to 12% relatively**. [Interspeech 2022] [ASRU 2023]

May 2020 - Jul 2020 **Research Intern**, Big-data Experience Lab, Adobe Research, [Full-time].
Advisor - Dr. Sumit Shekhar

- Used reinforcement learning (deep Q-learning) to learn an optimal acquisition function for active learning by modeling the active learning cycle as a Markov Decision Process. [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 - Mar 2020 **Research Intern**, Indian Institute of Science (IISc.), Bengaluru | Indian Institute of Technology, Roorkee, [Part-time].
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 text-to-speech (TTS) systems and proposed a GAN-decoder-based scoring mechanism. [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]

Publications

2025 **MoD-DPO : Towards Mitigating Cross-modal Hallucinations in Omni LLMs using Modality Decoupled Preference Optimization**.
Ashutosh Chaubey, Jiacheng Pang, Mohammad Soleymani
Conference on Computer Vision and Pattern Recognition **(CVPR) 2026** [Preprint]

2025 **AVERE : Improving Audiovisual Emotion Reasoning using Preference Optimization**.
Ashutosh Chaubey, Jiacheng Pang, Maksim Siniukov, Mohammad Soleymani
International Conference on Learning Representations **(ICLR) 2026** [Openreview][Webpage]

2025 **LibreFace 2.0 : Leveraging Large-Scale Synthetic Data for Fair and Generalizable Facial Analysis**.
Xulang Guan*, **Ashutosh Chaubey***, Maksim Siniukov, Belle Hsieh, Zongjian Li, Mohammad Soleymani
International Conference on Automatic Face and Gesture Recognition **(FG) 2026**

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** - R1 (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** [Paper]

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** [Paper] [Webpage]

2024 **ContextIQ : A Multimodal Expert-Based Video Retrieval System for Contextual Advertising**.
Ashutosh Chaubey, Anoubhav Agarwaal, 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]

Academic Services

Conference Reviewer: ICML 2026, CVPR 2026, ICCV 2025

Teaching Assistant: CS561 (Foundations of AI) – Fall 2025, CSCI 535 (Multimodal Probabilistic Learning of Human Communication) – Spring 2025

Supervised Students

Jiacheng Pang – Research Internship, Grad Student at USC, Summer 2025 - Present

Xulang Guan – CURVE Fellowship, Undergraduate at USC, Fall 2024 - Present

Belle Hsieh – Research Internship, Undergraduate at UPenn, Summer 2025

Hongkun Gong – CURVE Fellowship, Undergraduate at USC (Now, Grad Student at Columbia), Fall 2024

Skills

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

Frameworks/Libraries - PyTorch, NumPy, Pandas, Transformers

Tools - VSCode, Git, Anaconda, Docker