Chaitanya Ahuja

(412) 726 2373 • chahuja.com • chahuja • chahuja
ahujachaitanya@gmail.com

Education

Ph.D. in Language and Information Technologies — **Pittsburgh, PA**
School of Computer Science, Carnegie Mellon University, 4.02/4.0 — *Aug 2015 – May 2022*
Advisor: Prof. Louis-Philippe Morency

B.Tech. in Electrical Engineering — **Kanpur, India**
Indian Institute of Technology Kanpur, 9.5/10 — *Aug 2011 – May 2015*
Minor: Artificial Intelligence

Experience

- **AI Research Scientist, Meta Platforms, Inc.** — **Los Angeles, CA**
  Multimodal Recommendation Systems — *May 2022 – present*
  - **Drove the design and deployment of multimodal foundation models** jointly optimized with content and engagement objectives, improving cold-start performance and long-term engagement across **billion-user recommendation surfaces**.
  - **Architected multimodal representation learning pipelines** spanning pre-training and post-training over video–text data, directly influencing relevance, retrieval, and ranking quality in large-scale recommender systems.
  - **Led the development of generative, sequence-based recommendation models** that unify candidate generation and ranking by modeling user engagement trajectories and content semantics in a single multimodal model, simplifying system complexity while maintaining recommendation quality at scale.

  Multimodal Machine Learning
  - **Advanced multimodal generative modeling** by developing a novel RL method for attribute binding improvement (+15%) in image diffusion models, informing training strategies for stronger cross-modal alignment.
  - Demonstrated that **multimodal LLMs act as effective visual learners**, leveraging language supervision to acquire strong visual representations without task-specific vision architectures.
  - Pioneered **continual multimodal learning methods for personalized co-speech gesture generation**, enabling long-term adaptation while preventing catastrophic forgetting across audio–vision–language modalities.
- **Graduate Researcher, Carnegie Mellon University** — **Pittburgh, PA**
  Advisor: Prof. Louis-Philippe Morency — *Aug 2015 – May 2022*
  PhD Thesis - Communication beyond words: Grounding Visual Body Motion with Language
  - **Multimodal Grounding:** Developed grounding algorithms that builds a common embedding space for language, acoustics and human body pose for for the purposes of co-speech gesture generation.
  - **Gesture Style Transfer and Control:** Designed efficient algorithms to transfer idiosyncratic gesture styles of one speaker to another and a many-to-many gesture style transfer set-up.
  - **Low-Resource Generative Models:** Designed algorithms to generate co-speech gestures for new speakers with significantly lesser supervision.
- **Research Intern, Meta Reality Labs (previously Facebook Reality Labs)** — **Pittburgh, PA**
  Advisor: Shugao Ma — *May 2018 – Aug 2018*
  - Designed a neural network model to generate upper body animations in a dyadic conversational setting. These animations are conditioned on avatar's speech, pose history and interlocutor's speech and pose history.
  - Used an attention-based model to focus on interpersonal and intrapersonal dynamics as and when indicated by the stimuli to the model.
  - Demonstrated the model's effectiveness in generating accurate and natural looking pose sequences via various objective and subjective metrics of evaluation.
- **Research Intern, Cornell University** — **Ithaca, NY**
  Advisor: Prof. Tsuhan Chen — *May 2014 – Aug 2014*
  - Designed a system to predict adjectives for a given noun based on an existing set of tags, which increased the vocabulary of the tags while maintaining the sanctity of the noun-adjective pair

- Improved the compatibility of adjectives with respect to nouns based on a probability measure by incorporating a sentence corpus (e.g. British-National-Corpus).

- **Undergraduate Researcher, Indian Institute of Technology** **Kanpur, India**
  Advisor: Prof. Rajesh Hegde *May 2013 – May 2015*
  - Worked towards mimicking a ear with digital filters that can help synthesize **Spatial Audio**.
  - Developed methods to construct ear contours generated by spectral notches of Head Related Transfer Functions (or HRTFs), hence mapping HRTFs to the anthropometry of the ear.
  - Explored relationships between structure of a ear and HRTFs.

- **Undergraduate Researcher, Indian Institute of Technology** **Kanpur, India**
  Advisor: Prof. Vinay Namboodiri *Aug 2014 – May 2015*
  - Proposed and implemented an online system for creating human-centric image summaries of videos.
  - Designed a Kernel-based tracking algorithm for automated live synthesis of video synopsis for human-centric videos.

Publications

Updated list on Google Scholar

Pre-prints/Under review

1. A Simple and Effective Reinforcement Learning Method for Text-to-Image Diffusion Fine-tuning
   Shashank Gupta, **Chaitanya Ahuja**, Tsung-Yu Lin, Sreya Dutta Roy, Harrie Oosterhuis, Maarten de Rijke, Satya Narayan Shukla
   2026 Under Review

Refereed conferences/journals

1. Generative Context Improves Multimodal Embedding
   Xuanming Cui, Jianpeng Cheng, Hong-you Chen, Satya Narayan Shukla, Abhijeet Awasthi, Xichen Pan, **Chaitanya Ahuja**, Shlok Kumar Mishra, Yonghuan Yang, Jun Xiao, Qi Guo, Ser-Nam Lim, Aashu Singh, Xiangjun Fan
   International Conference on Learning Representations (ICLR 2026)

2. Multi-Modal Large Language Models are Effective Vision Learners
   Li Sun, **Chaitanya Ahuja**, Peng Chen, Matt D'Zmura, Kayhan Batmanghelich, Philip Bontrager
   IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2025)

3. Continual Learning for Personalized Co-speech Gesture Generation
   Chaitanya Ahuja, Pratik Joshi, Ryo Ishii, Louis-Philippe Morency
   International Conference on Computer Vision (ICCV 2023)
   [webpage]

4. Lecture Presentations Multimodal Dataset: Towards Understanding Multimodality in Educational Videos
   Dong Won Lee, **Chaitanya Ahuja**, Paul Pu Liang, Sanika Natu, Louis-Philippe Morency
   International Conference on Computer Vision (ICCV 2023)
   [code]

5. A Comprehensive Review of Data-Driven Co-Speech Gesture Generation
   Simbarashe Nyatsanga, Taras Kucherenko, **Chaitanya Ahuja**, Gustav Eje Henter, Michael Neff
   Annual Conference of the European Association for Computer Graphics (EUROGRAPHICS 2023)

6. Communication Beyond Words: Grounding Visual Body Motion with Language
   Chaitanya Ahuja
   PhD dissertation, Carnegie Mellon University, 2022

7. Low-Resource Adaptation for Personalized Co-Speech Gesture Generation
   Chaitanya Ahuja, Dong Won Lee, Louis-Philippe Morency
   Conference on Computer Vision and Pattern Recognition (CVPR 2022)
   [webpage] [code] [supp]

8. No gestures left behind: Learning relationships between spoken language and freeform gestures
   Chaitanya Ahuja, Dong Won Lee, Ryo Ishii, Louis-Philippe Morency
   Findings of Empirical Methods in Natural Language Processing (Findings of EMNLP, 2020)
   Presented at Natural Language Beyond Text Workshop @EMNLP 2020
   [code][video]

9. Impact of personality on nonverbal behavior generation
   Ryo Ishii, **Chaitanya Ahuja**, Yukiko I. Nakano, Louis-Philippe Morency
   ACM International Conference on Intelligent Virtual Agents (IVA, 2020)
10. Style transfer for co-speech gesture animation: A multi-speaker conditional mixture approach
    Chaitanya Ahuja, Dong Won Lee, Yukiko I. Nakano, Louis-Philippe Morency
    European Conference on Computer Vision (ECCV, 2020)
    [code] [demo] [video] Media: TechXplore
11. To react or not to react: End-to-end visual pose forecasting for personalized avatar during dyadic conversations
    Chaitanya Ahuja, Shugao Ma, Louis-Philippe Morency, Yaser Sheikh
    ACM International Conference on Multimodal Interaction (ICMI, 2019)
12. Coalescing Narrative and Dialogue for Grounded Pose Forecasting
    Chaitanya Ahuja
    Doctoral Consortium, ACM International Conference on Multimodal Interaction (ICMI 2019)
13. Language2pose: Natural language grounded pose forecasting
    Chaitanya Ahuja, Louis-Philippe Morency
    International Conference on 3D Vision (3DV, 2019)
    [code] [webpage] Media: Scientific American, Synced, VentureBeat
14. A complex matrix factorization approach to joint modeling of magnitude and phase for source separation
    Chaitanya Ahuja, Karan Nathwani, Rajesh M. Hegde
    IEEE International Symposium on Signal Processing and Information Technology (ISSPIT, 2019)
15. Multimodal machine learning: A survey and taxonomy
    Tadas Baltrušaitis, **Chaitanya Ahuja**, Louis-Philippe Morency
    Transactions on Pattern Analysis and Machine Intelligence (TPAMI, 2018)
16. Lattice recurrent unit: Improving convergence and statistical efficiency for sequence modeling
    Chaitanya Ahuja, Louis-Philippe Morency
    AAAI Conference on Artificial Intelligence (AAAI, 2018)
    [code] [webpage]
17. Fast modelling of pinna spectral notches from HRTFs using linear prediction residual cepstrum
    Chaitanya Ahuja, Rajesh M. Hegde
    IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)

Book Chapters

1. Challenges and applications in multimodal machine learning
   Tadas Baltrušaitis, **Chaitanya Ahuja**, Louis-Philippe Morency
   The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2, 2018, pp. 17–48

Refereed Workshops

1. Crossmodal clustered contrastive learning: Grounding of spoken language to gestures
   Dong Won Lee, **Chaitanya Ahuja**, Louis-Philippe Morency
   GENEA Workshop 2021 @ACM International Conference on Multimodal Interaction (ICMI, 2021)
2. Extraction of pinna spectral notches in the median plane of a virtual spherical microphone array
   Ankit Sohni, **Chaitanya Ahuja**, Rajesh M. Hegde
   Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA, 2014)

Organized Workshops

1. Expressive Encounters Workshop
   European Conference on Computer Vision (ECCV) 2024 Workshop Proceedings
2. First Workshop on Crossmodal Social Animation
   International Conference on Computer Vision (ICCV, 2021) Workshop Proceedings
   [video]
3. First Workshop on Multimodal Fact Checking and Hate Speech Detection
   AAAI Conference on Artificial Intelligence (AAAI, 2022) Workshop Proceedings
   [dataset] [workshop proc.]

Technical Reports

1. Training Segmentation Models for Extractive and Generative NLP Tasks with Reinforcement Learning
   Akash Bharadwaj*, **Chaitanya Ahuja***
   Course Project, Deep RL and control at CMU
2. Topological Data Analysis
   Bhuwan Dhingra*, **Chaitanya Ahuja***
   Course Project, Statistical Machine Learning at CMU [slides]
3. Video Captioning
   Salvador Medina*, **Chaitanya Ahuja***
   Course Project, Advanced Multimodal Machine Learning at CMU
4. Visual Summarization of foreground object motion using boundary initialization of object tracking
   Chaitanya Ahuja*, Pratik Somani*
   B.Tech Final Year Project at IIT Kanpur

Honors and Awards

- Highlighted Reviewer at ICLR 2022
- CMU Graduate Research Fellowship, 2015 – 2020
- Honorable Mention at the LTI Student Research Symposium, 2019
- Cornell-India summer program at Cornell University, 2014
- Viterbi-India summer program at the University of Southern California (declined), 2014
- Summer Undergraduate Research Grant for Excellence (SURGE) 2013, IIT Kanpur
- One of the top 7 projects (out of 70) in SURGE 2013
- Academic Excellence Awards for distinctive performance, 2011 – 2013, IIT Kanpur
- All India Rank 231 - Top 0.05% (amongst 4,75,000 students) in IIT-JEE 2011.
- All India Rank 124 - Top 0.05% (amongst 10,00,000 students) in AIEEE 2011.

Student Mentorship

- Dong Won Lee (CMU BS → CMU MS in Machine Learning → MIT Media Lab): Self-supervised generative models.
- Pratik Joshi (CMU MS): Continual learning for generative models.
- Sanika Natu (CMU MS): Understanding multimodality in educational slides
- Shradha Sehgal (IIIT Hyderabad B.Tech. → UIUC MS in Computer Science): Evaluation of generative models.
- Arvin Wu (CMU BS): Social intelligence benchmarking.
- Nikitha Murikinati (CMU BS): Study of relationships between co-speech gestures and prosody.
- Sharath Rao (CMU MS → PlayStation): Back-channel prediction in dyadic conversations.
- Qingtao Hu (CMU MS → Amazon): Unsupervised disentanglement of style and content in images.
- Anirudha Rayasam (CMU MS → Google): Language grounded pose forecasting.

Teaching

- Head TA: 11763 Structured Prediction for language and discrete data by Taylor Berg-Kirkpatrick and Bhiksha Raj, CMU
  3 recitations on Viterbi Decoding [slides], ILP and Dependency Parsing [slides] and Neural CRFs [slides] *Spring 2018*
- Head TA: 11-777 Multimodal Machine Learning by Louis-Philippe Morency, CMU *Spring 2017*

Talks

- **Communication Beyond Words: Grounding Visual Body Motion with Spoken Language**
  KTH Stockholm, Online *April 2021*
- **Learning Relationships between Spoken Language and Freeform Gestures**
  EMNLP 2020 Workshop on NLP Beyond Text, Online *November 2020*
- **Natural Language Grounded Pose Forecasting**
  LTI Student Research Symposium, Pittsburgh PA *August 2019*

- **End-to-End Visual Pose Forecasting for Personalized Avatar during Dyadic Conversations**
  ACM International Conference on Multimodal Interaction, Suzhou, China *October 2019*

Resources

- **PATS Dataset**: Designed and constructed a large benchmark to study the complex multimodal relationships between Body **P**oses, **A**udio, **T**ranscripts, and individual gesture **S**tyles
- Multimodal Lecture Presentations Dataset: Designed benchmark to study AI models capable of understanding multimodal information present in lecture slides
- chahuja/aisle: Learning relationships between spoken language and freeform gestures
- chahuja/mix-stage: Style transfer for co-speech gesture generation
- chahuja/language2pose: Natural language grounded pose forecasting
- chahuja/lru: Lattice recurrent units

Professional Activities and Service

- Co-organizer: ECCV 2024 Expressive Encounters Workshop
- Co-organizer: ICCV 2021 First Workshop on Crossmodal Social Animation
- Co-organizer: Multimodal Machine Learning Reading Group, CMU, Spring 2020
- Conference Program Committee: NeurIPS, ICLR, CVPR, ECCV, SIGGRAPH, ACL, EMNLP, ICMI
- Workshop Program Committee: NeurIPS workshop on Multimodal Machine Learning, ACL Workshop on Multimodal Language, NAACL-HLT Student Research Workshop, ICMI GENEA Workshop
- Grant Reviewer: Army Research Office (ARO)
- CMU Graduate Applicant Support Program Volunteer: 2020
- CMU AI Undergraduate Research Mentor: 2020, 2021
- CMU Graduate Student Association Representative for Language Technologies Institute: 2017

Skills

- Languages: English (fluent), Hindi (native), Spanish (Limited Working)
- Programming Languages: Python, C, MATLAB, CSS, HTML, LaTeX
- Frameworks: Numpy, Pandas, PyTorch, Scikit-Learn, Scipy, Tensorflow, Theano

Relevant Graduate Coursework

- Structured Prediction for Language and Other Discrete Data (CMU 10-763): T. Berg-Kirkpatrick, B. Raj *Spring 2018*
- Deep Reinforcement Learning (CMU 10-703): R. Salakhutdinov, K. Fragkiadaki *Spring 2017*
- Statistical Machine Learning (CMU 10-702): L. Wasserman, R. Tibshirani *Spring 2017*
- Deep Learning (CMU 10-707): R. Salakhutdinov *Fall 2016*
- Intermediate Statistics (CMU 10-705): L. Wasserman *Fall 2016*
- Advanced Multimodal Machine Learning (CMU 11-777): L.-P. Morency *Spring 2016*
- Machine Learning (PhD) (CMU 10-701): T. Mitchell *Spring 2016*
- Human Communication and Multimodal ML (CMU 11-776): L.-P. Morency *Fall 2015*
- Algorithms for NLP (CMU 10-702): C. Dyer *Fall 2015*

References

Available on request

Last updated: February 4, 2026