David M. Chan, Ph.D.

davidchan@berkeley.edu @_dmchan david-m-chan DavidMChan
http://dchan.cc/

Education

2024 **Ph.D., University of California, Berkeley** in Electrical Engineering and Computer Science.
Thesis title: *Understanding, Building, and Evaluating Models for Context Aware Conditional Natural Language Generation*. Advisor: John Canny.

2020 **M.Sc., University of California, Berkeley** in Electrical Engineering and Computer Science.
Thesis title: *TSNE-CUDA: GPU-Accelerated TSNE*. Advisor: John Canny.

2017 **B.Sc., University of Denver** in Mathematics and Computer Science.
Graduated Summa Cum Laude with Honors with Distinction, Phi Beta Kappa. GPA: 4.0.

Employment History

2024 – Present **Postdoctoral Scholar,** University of California, Berkeley.
Designed and lead projects on grounded multimodal representation learning and multimodal foundation models. Advised by Prof. Trevor Darrell.

2021 – 2024 **Student Researcher,** Amazon.
Developed and implemented large language models for context-driven speech understanding, achieving a 13.8% relative WER improvement. Researched model distillation to reduce hot-fixing domain-alignment time by over 1K GPU hours, saving $3K per training run. Designed multi-modal ASR model frameworks, improving performance on low-resource devices by 45%.

2020 **Intern,** Google.
Developed systems combining large language models (LLMs) and large multimodal models (LMMs) for image captioning, resulting in an 84% MRR improvement. Designed large-scale GPU training procedures for machine learning with up to 1K GPUs.

2019 **Intern,** Dropbox.
Architected and deployed Dropbox's first deep learning AI for multimodal content classification, achieving a 0.96 F1 score on classification tasks.

2018 **Intern,** NASA Jet Propulsion Laboratory.
Developed flight-ready provable AI systems for high-risk environments and contributed core software to a top-5 DARPA Subterranean Challenge team.

2014 – 2017 **Research Assistant,** Dreamface Technologies.
Engaged in research on cutting-edge AI technologies and supported various development projects.

Research Publications

Journal Articles

1. J.-w. Jung, D. Zhang, C.-H. H. Yang, S.-L. Wu, **D. M. Chan**, Z. Kong, R. Deng, Y. Zhou, R. Valle, and S. Watanabe, "Automatic audio captioning with encoder fusion, multi-layer aggregation, and large language model enriched summarization," *2024 IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2024.

2. C. Figueroa, T. Luo, A. Jacobo, A. Munoz, M. Manuel, **D. Chan**, J. Canny, and A. Aguilera, "Conversational physical activity coaches for spanish and english speaking women: A user design study. front. digit," *Digital health equity*, vol. 3, p. 14, 2023.

3. C. Figueroa, T. Luo, A. Jacobo, A. Munoz, M. Manuel, **D. Chan**, J. Canny, and A. Aguilera, "Conversational physical activity coaches for spanish and english speaking women: A user design study," *Frontiers in Digital Health*, p. 134, 2021.

4. **D. M. Chan**, R. Rao, F. Huang, and J. F. Canny, "Gpu accelerated t-distributed stochastic neighbor embedding," *Journal of Parallel and Distributed Computing*, vol. 131, pp. 1–13, 2019.

Conference Proceedings

1. R. Gupta, R. Corona, J. Ge, E. Wang, D. Klein, T. Darrell, and D. M. Chan, "Enough coin flips can make LLMs act Bayesian," in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar, Eds., Jul. 2025.

2. W.-H. Hsieh, E. Hsieh, D. Niu, T. Darrell, R. Herzig, and **D. M. Chan**, "Do what? teaching vision-language-action models to reject the impossible," in *Findings of the Association for Computational Linguistics: EMNLP 2025*, Association for Computational Linguistics, 2025.

3. B. Hu, **D. Chan**, T. Sorensen, X. Chen, H. Ji, Y. Choi, T. Darrell, and A. Basharat, "A roadmap for alignable algorithmic decision-makers in the medical triage domain," in *2025 IEEE Conference on Artificial Intelligence (CAI)*, IEEE, 2025, pp. 1179–1183.

4. M. Kang, S. Moon, S. H. Lee, A. Raj, J. Suh, and **D. Chan**, "Deep binding of language model virtual personas: A study on approximating political partisan misperceptions," in *Second Conference on Language Modeling*, 2025.

5. D. Kutscher, **D. M. Chan**, Y. Bai, T. Darrell, and R. Gupta, "Reordering patches improves vision models," in *Advances in Neural Information Processing Systems 38*, 2025.

6. H. Lee, J. Ge, T.-H. Wu, M. Kang, T. Darrell, and **D. M. Chan**, "Puzzled by puzzles: When vision-language models can't take a hint," in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 2025.

7. J. Park, S. Takamichi, **D. M. Chan**, S. Kando, Y. Saito, and H. Saruwatari, "Analysing the language of neural audio codecs," in *2025 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2025.

8. J. Quenum, W.-H. Hsieh, T.-H. Wu, R. Gupta, T. Darrell, and **D. M. Chan**, "Lisat: Language-instructed segmentation assistant for satellite imagery," in *Advances in Neural Information Processing Systems 38*, 2025.

9. T.-H. Wu, G. Biamby, J. Quenum, R. Gupta, J. E. Gonzalez, T. Darrell, and **D. Chan**, "Visual haystacks: A vision-centric needle-in-a-haystack benchmark," in *The ThirteCenth International Conference on Learning Representations*, 2025.

10. T.-H. Wu, J. E. Gonzalez, T. Darrell, and **D. M. Chan**, "Clair-a: Leveraging large language models to judge audio captions," in *2025 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2025.

11. T.-H. Wu, H. Lee, J. Ge, J. E. Gonzalez, T. Darrell, and **D. M. Chan**, "Generate, but verify: Reducing hallucination in vision-language models with retrospective resampling," in *Advances in Neural Information Processing Systems 38*, 2025.

12. K. Cai, C. Liu, and **D. M. Chan**, "Anim-400k: A large-scale dataset for automated end-to-end dubbing of video," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2024.

13. **D. M. Chan**, S. Ghosh, H. Tulsiani, A. Rastrow, and B. Hoffmeister, "Task oriented dialogue as a catalyst for self-supervised automatic speech recognition," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2024.

14 **D. M. Chan**, Y. Ni, D. A. Ross, S. Vijayanarasimhan, A. Myers, and J. Canny, "Distribution aware metrics for conditional natural language generation," in *LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, 2024.

15 **D. M. Chan**, S. Petryk, A. Kachinthaya, H. Zou, J. Canny, J. E. Gonzalez, and T. Darrell, "Aloha: A new measure for hallucination in captioning models," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, vol. 2, 2024, pp. 342–357.

16 **D. M. Chan**, H. Tulsiani, S. Ghosh, G. Lalwani, P. Pandey, A. Bansal, S. Garimella, A. Rastrow, and B. Hoffmeister, "An efficient self-learning framework for interactive spoken dialog systems," in *International Conference on Machine Learning, 2024*, 2024.

17 Y. Jain, **D. Chan**, P. Dheram, A. Khare, O. Shonibare, V. Ravichandran, and S. Ghosh, "Multi-stage multi-modal pre-training for automatic speech recognition," in *LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, 2024.

18 S. Moon, M. Abdulhai, M. Kang, J. Suh, W. Soedarmadji, E. K. Behar, and **D. M. Chan**, "Virtual personas for language models via an anthology of backstories," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2024.

19 J. Suh, S. Moon, M. Kang, and **D. M. Chan**, "Rediscovering the latent dimensions of personality with large language models as trait descriptors," in *NeurIPS 2024 Workshop on Behavioral Machine Learning*, 2024.

20 T.-H. Wu, G. Biamby, **D. Chan**, L. Dunlap, R. Gupta, X. Wang, J. E. Gonzalez, and T. Darrell, "See say and segment: Teaching lmms to overcome false premises," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 13 459–13 469.

21 **D. Chan**, S. Petryk, J. E. Gonzalez, T. Darrell, and J. Canny, "Clair: Evaluating image captions with large language models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2023, pp. 13 638–13 646.

22 **D. M. Chan**, S. Ghosh, A. Rastrow, and B. Hoffmeister, "Domain adaptation with external off-policy acoustic catalogs for scalable contextual end-to-end automated speech recognition," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023, pp. 1–5.

23 **D. M. Chan**, A. Myers, S. Vijayanarasimhan, D. A. Ross, and J. Canny, "Ic3: Image captioning by committee consensus," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2023, pp. 8975–9003.

24 E. Kosoy, **D. M. Chan**, A. Liu, J. Collins, J. Hamrick, S. H. Huang, N. R. Ke, E. R. Reagan, J. Canny, and A. Gopnik, "Towards understanding how machines can learn causal overhypotheses," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 45, 2023.

25 V. Lialin, S. Rawls, **D. Chan**, S. Ghosh, A. Rumshisky, and W. Hamza, "Scalable and accurate self-supervised multimodal representation learning without aligned video and text data," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 390–400.

26 A. S. Sundar, C.-H. H. Yang, **D. M. Chan**, S. Ghosh, V. Ravichandran, and P. S. Nidadavolu, "Multimodal attention merging for improved speech recognition and audio event classification," in *ICASSP 2024 Workshop on Self-supervision in Audio, Speech and Beyond*, IEEE, 2023.

27 **D. M. Chan** and S. Ghosh, "Content-context factorized representations for automated speech recognition," in *Proceedings of Interspeech 2022*, International Speech Communication Association, 2022, pp. 61–65.

28 **D. M. Chan**, S. Ghosh, D. Chakrabarty, and B. Hoffmeister, "Multi-modal pre-training for automated speech recognition," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 246–250.

29 **D. M. Chan**, A. Myers, S. Vijayanarasimhan, D. A. Ross, B. Seybold, and J. F. Canny, "What's in a caption? dataset-specific linguistic diversity and its effect on visual description models and metrics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4740–4749.

30 S. Gurram, **D. Chan**, A. Fang, and J. Canny, "Lava: Language audio vision alignment for data-efficient video pre-training," in *First Workshop on Pre-training: Perspectives, Pitfalls, and Paths Forward at ICML 2022*, 2022.

31 E. Kosoy, A. Liu, J. L. Collins, **D. Chan**, J. B. Hamrick, N. R. Ke, S. Huang, B. Kaufmann, J. Canny, and A. Gopnik, "Learning causal overhypotheses through exploration in children and computational models," in *Conference on Causal Learning and Reasoning*, PMLR, 2022, pp. 390–406.

32 K. Wang, S. Z. Zhao, **D. Chan**, A. Zakhor, and J. Canny, "Multimodal semantic mismatch detection in social media posts," in *2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, 2022, pp. 1–6.

33 **D. M. Chan**, S. Vijayanarasimhan, D. A. Ross, and J. F. Canny, "Active learning for video description with cluster-regularized ensemble ranking," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

34 E. Kosoy, J. Collins, **D. M. Chan**, S. Huang, D. Pathak, P. Agrawal, J. Canny, A. Gopnik, and J. B. Hamrick, "Exploring exploration: Comparing children with rl agents in unified environments," in *ICLR 2020 Workshop on Bridging AI and Cognitive Science (BAICS)*, 2020.

35 **D. M. Chan** and A.-a. Agha-mohammadi, "Autonomous imaging and mapping of small bodies using deep reinforcement learning," in *IEEE Aerospace Conference 2019*, 2019.

36 D. Seita, **D. Chan**, R. Rao, C. Tang, M. Zhao, and J. Canny, "Zpd teaching strategies for deep reinforcement learning from demonstrations," in *NeurIPS 2019 Workshop on Deep Reinforcement Learning Workshop*, 2019.

37 **D. M. Chan**, R. Rao, F. Huang, and J. F. Canny, "T-sne-cuda: Gpu-accelerated t-sne and its applications to modern data," in *2018 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*, IEEE, 2018, pp. 330–338.

38 L. Cohen, G. Wagner, **D. Chan**, H. Choset, N. Sturtevant, S. Koenig, and T. Kumar, "Rapid randomized restarts for multi-agent path finding solvers," in *Proceedings of the International Symposium on Combinatorial Search*, vol. 9, 2018, pp. 148–152.

39 T. Walker, **D. Chan**, and N. Sturtevant, "Using hierarchical constraints to avoid conflicts in multi-agent pathfinding," in *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 27, 2017, pp. 316–324.

40 A. Mollahosseini, **D. Chan**, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter conference on applications of computer vision (WACV)*, IEEE, 2016, pp. 1–10.

41 A. Mollahosseini, B. Hasani, M. J. Salvador, H. Abdollahi, **D. Chan**, and M. H. Mahoor, "Facial expression recognition from world wild web," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 58–65.

Preprints

1. J. Ge, G. Luo, H. Lee, N. Malpani, L. Lian, X. Wang, A. Holynski, T. Darrell, S. Min, and **D. M. Chan**, "Constantly improving image models need constantly improving benchmarks," *arXiv preprint arXiv:2510.15021*, 2025.
2. T.-H. Wu, M. Miroyan, **D. M. Chan**, T. Darrell, N. Norouzi, and J. E. Gonzalez, "Are large reasoning models interruptible?" *arXiv preprint arXiv:2510.11713*, 2025.
3. **D. M. Chan**, R. Corona, J. Park, C. J. Cho, Y. Bai, and T. Darrell, "Analyzing the language of visual tokens," *arXiv preprint arXiv:2411.05001*, 2024.
4. K. Wang, **D. Chan**, S. Z. Zhao, J. Canny, and A. Zakhor, "Misinformation detection in social media video posts," *arXiv preprint arXiv:2202.07706*, 2022.
5. A. Jiwani, S. Ganguly, C. Ding, N. Zhou, and **D. M. Chan**, "A semantic segmentation network for urban-scale building footprint extraction using rgb satellite imagery," *arXiv preprint arXiv:2104.01263*, 2021.
6. B. Jiang, **D. M. Chan**, T. Zhang, and J. F. Canny, "Diagnostic visualization for deep neural networks using stochastic gradient langevin dynamics," *arXiv preprint arXiv:1812.04604*, 2018.
7. P. S. Negi, **D. Chan**, and M. Mahoor, "Leveraging class similarity to improve deep neural network robustness," *arXiv preprint arXiv:1812.09744*, 2018.

Supervised Theses

1. D. Jhamb, **D. Chan**, J. F. Canny, and A. Zakhor, "Hallucination is all you need: Using generative models for test time data augmentation," 2022.
2. I. Herzi, **D. Chan**, and J. F. Canny, "Exploring the effects of view transforms on self-supervised video representation learning techniques," 2021.

Organized Workshops/Tutorials

2026 **CVPR –** What's Next in Multimodal Foundation Models (MMFM5)

2025 **IEEE CAI –** Human Alignment in AI Decision-Making Systems: An Inter-disciplinary Approach towards Trustworthy AI.

ICCV – What's Next in Multimodal Foundation Models (MMFM4)

ASRU (Tutorial) – Multimodal Speech Modeling: from Understanding to Generation

Invited Talks and Presentations

2026 CV4Edu CVPR Workshop, Panel Member

Edge AI and Vision Alliance Innovation Forum, Beyond Words: Teaching AI to Converse With the Physical World

Hippocratic AI, Hallucination by Design: Building Models That Know What to Say When They're Wrong

2025 University of Tokyo, Multimodal Pragmatics for Foundation Models

ICCV MMFM4, Panel Moderator: "What's Next in Multimodal Foundation Models"

TTIC Summer Workshop on Speech Foundation Models: "Metrics for Spoken Language Models"

BayLearn: "Dementor: Stealing the Soul of an LLM"

2024 EMNLP, Panel Member: "LLM Agents for Audio and Continuous Signals"

Invited Talks and Presentations (continued)

- KAUST, "Understanding, Building, and Evaluating Models for Context-Aware Conditional Natural Language Generation."

2023
- Berkeley Workshop on Generative Modeling, "IC3: Image Captioning by Committee Consensus."

2022
- PyTorch Lightning Developer's Conference, "Multimodal Language Modeling."
- CVPR 2022 (Workshop on Vision Datasets Understanding), "What's in a Caption."
- Google, "What's in a Caption."
- Amazon, "Multimodal Automated Speech Recognition."

2021
- Google, "Vision and Language Pre-Training for Visual Description."

2020
- Google, "Active Learning for Visual Description."

Teaching

2025
- Co-Instructor: *CS294-43: Large Vision and Language Models*

2024
- Co-Instructor: *CS294-43: Large Vision and Language Models*

2020
- Lead GSI: *CS182: Designing, Visualizing and Understanding Deep Neural Networks*

2019
- Initiating GSI: *CS182: Designing, Visualizing and Understanding Deep Neural Networks*

2017
- Co-Instructor: *COMP 3004: Foundations in Discrete Structures and Algorithms*
- Undergraduate TA: *COMP 3003: Foundations in Computer Systems*
- Undergraduate TA: *COMP 3004: Foundations in Discrete Structures and Algorithms*

2016
- Undergraduate TA: *COMP 3004: Foundations in Discrete Structures and Algorithms*

2015
- Undergraduate TA: *COMP 3004: Foundations in Discrete Structures and Algorithms*

Mentorship

PhD
- Suzanne Petryk (2024, Apple), Tsung-Han Wu (2025), Jiaxin Ge (2025)

MS
- Dhruv Jhamb (2020, ServiceNow), Illian Herzi (2020, Apple)

Undergraduate
- Karen Liu (2019, KiwiBot), Xue Bofan (2019, Bytedance), Dipika Khullar (2021, Amazon), Aatif Jiwani (2020, Amazon), Shubhara Ganguly (2020, LiveKit), Oliver Bryniarski (2021, Remedy Robotics), Sumanth Gurram (2021, Meta), Yiming Ni (2022, Stanford MS), May Liu (2022, UT Austin MS), Kevin Cai (2022, Bytedance), Haodi Zou (2023, UCSD PhD), Anish Kachinthaya (2023, UC Berkeley MS), Wen-Han Hsieh (2024), Heekyong Lee (2024), Joonyong Park (2024), Haoming Chen (2025)

Service

Departmental Service

2022-2024
- UC Berkeley Faculty Search Committee Member.

2021 – Present
- UC Berkeley Incoming Grad Student Peer Mentor.
- UC Berkeley BAIR Webmaster.

2020 – Present
- UC Berkeley EECS Graduate Admissions Committee Member.

2021 – Present
- BAIR Coffee Procurement Manager

Academic Service

Outstanding Reviewer
- NeurIPS Datasets/Benchmarks (22), CVPR (23,25), ECCV (24), EMNLP (24)

Service (continued)

Area Chair/Action Editor	ICASSP (26), *CL ARR (25,26), ECCV (26)
Reviewer	WACV (19,20,21,22,23,24,25), NeurIPS (23,24,25) ICLR (25,26), ICCV (21, 23, 25), ECCV (22, 24), ACCV (24), CVPR (21,22,23,24,25,26), EMNLP (22,23,24,25), *CL ARR (24,25), ICASSP (24,25), Interspeech (24,25), AMLC (22,23)

Awards and Grants

Awards (Conference)

- 2024 — Judge's Award, DCASE 2024, "Automatic audio captioning with encoder fusion, multi-layer aggregation, and large language model enriched summarization"
- 2018 — Best Paper, HPML 2018 "Gpu accelerated t-distributed stochastic neighbor embedding"

Awards (Academic)

- 2022-2024 — Evergreen Award for Undergraduate Mentorship.
- 2016 – 2017 — Herbert J. Greenberg Award for Excellence in Mathematics.
- 2017 — DU Undergraduate Student Researcher of the Year.
- DU Computer Science Departmental Service Award.
- 2015 – 2017 — Outstanding Computer Science and Mathematics Major, University of Denver.

Grants

- 2024-2025 — Amazon Gift Funding, "Audio-Visual Representation Learning for Multi-Modal Language Model Alignment", $150K, Co-PI
- 2020 – 2023 — Google BAIR Commons Funded Project, "Assistive Video Description," $450K, Student Co-PI.
- 2021 – 2022 — Amazon BAIR Commons Funded Project, "Weakly Supervised Multi-modal Pre-training," $150K, Student Co-PI.
- 2020 — UC Berkeley CTSP Fellowship, "Conversational Coaches for Diverse Individuals with Low Literacy," $5K.
- CITRIS Institute Tech Innovation Grant, "Conversational Coaches for Diverse Individuals with Low Literacy," $10K.
- 2015 — University of Denver Undergraduate Research Grant, "Deep Neural Networks for Facial Expression Recognition," $3K.
- 2016 — University of Denver PINS Grant, "Bayesian Priors for Object Recognition," $1.5K.

Skills

Programming Languages	Python, C++, CUDA, TypeScript, C, Rust, Bash, R, Matlab, C#, CSS, HTML, JavaScript.
Frameworks and Technologies	PyTorch, TensorFlow, Jax, MPI, Spark, Kubernetes, AWS, GCP.
Distributed/Parallel Computing	MPI, OpenMP, Slurm, Spark.
Cloud Computing	AWS, GCP, Kubernetes.
Development Tools	Bash, Git, Bazel, MS Office, Photoshop, Illustrator, In-Design.
AI/ML	Transformers, Large Language Models (LLMs), Multimodal Models (LMMs), Model Distillation.

References

Available on Request