Yao Qin

Email: yaoqin@ece.ucsb.edu

Education

University of California, San Diego *2015.09 - 2020.01*
Doctor of Philosophy, Department of Computer Science and Engineering
Advisor: Prof. Garrison Cottrell

University of California, San Diego *2015.09 - 2017.12*
Master of Science, Department of Computer Science and Engineering
Advisor: Prof. Garrison Cottrell

Dalian University of Technology *2011.09 - 2015.06*
Bachelor of Science, Department of Electrical Engineering
Advisor: Prof. Huchuan Lu

Research Experience

Assistant Professor, Department of ECE, University of California, Santa Barbara, USA *2023.01 - present*

Research Scientist, Google Research, New York, USA *2020 - 2023*

Publications (**Note**: * below denotes equal contribution) Google Scholar

Preprints

13. A. Hua, C. Gu, J. Gu, E. Wong and **Y. Qin**. Prompt-Insensitive Evaluation: Generalizing LLM Evaluation Across Prompts Through Fine-Tuning. *Under Review*, 2025.
12. K. Tang, Y. Li and **Y. Qin**. ChainReaction: A Precise, Customizable, and Iterative Image Editing Workflow. *Under Review*, 2025.
11. C. Gu, A. Hua, J. Gu and **Y. Qin**. Improving Adversarial Transferability in MLLMs via Dynamic Vision-Language Alignment Attack. *Under Review*, 2025.
10. M. Dhaliwal*, K. Tang*, E. M. Aiello, D. P. Zaharieva, R. A. Lal, C. Summers, B. Arbiter, K. Watson, M. J. Connolly, L. E. Figg, I. Balistreri, A. L. Cortes, R. S. Kingman, B. Suh, M. C. Riddell and **Y. Qin**. Variation in Hypoglycemia Risk with Real-World Physical Activity in Adults with Type 1 Diabetes: Insights from the Type 1 Diabetes Exercise Initiative. *Under Review*, 2025.
9. L. Liu, R. Pourreza, S. Panchal, A. Bhattacharyya, **Y. Qin** and R. Memisevic. Enhancing Hallucination Detection through Noise Injection. *Under Review*, 2025.
8. Y. Yoon, D. Hu, I. Weissburg, **Y. Qin** and H. Jeong. Detecting Out-of-Distribution through the Lens of Neural Collapse. *Under Review*, 2025.
7. L. Liu and **Y. Qin**. Detecting Out-of-Distribution through the Lens of Neural Collapse. *Under Review*, 2025.
6. A. Balashankar, X. Ma, A. Sinha, A. Beirami, **Y. Qin**, J. Chen and A. Beutel. Improving Few-shot Generalization of Safety Classifiers via Data Augmented Parameter-Efficient Fine-Tuning. *Under Review*, 2025.

5. J. Gu, A. Beirami, X. Wang, A. Beutel, P. Torr and **Y. Qin**. Towards Robustness of In-Context Learning on Vision-language Models. *Under Review*, 2024.

4. J. Gu, Z. Han, S. Chen, A. Beirami, B. He, G. Zhang, R. Liao, **Y. Qin**, V. Tresp and P. Torr. A Systematic Survey of Prompt Engineering on Vision-Language Foundation Models. *Under Review*, 2023.

3. M. Song, X. Wang, T. Biradar, **Y. Qin** and M. Chandraker. Acute Zero-Shot Imitation Learning with Task Prompting. *Under Review*, 2023.

2. **Y. Qin**, N. Frosst, C. Raffel, G. Cottrell and G. Hinton. Deflecting Adversarial Attacks. *Preprints*, 2019.

1. Ian Goodfellow, **Yao Qin**, David Berthelot. Evaluation Methodology for Attacks Against Confidence Thresholding Models. *Preprints*, 2018.

Conferences & Journals

24. A. Hua*, M. Dhaliwal*, L. Pullela, R. Burke and **Y. Qin**. NutriBench: A Dataset for Evaluating Large Language Models in Nutrition Estimation from Meal Descriptions (**ICLR**), 2025.

23. K. Tang, P. Song, **Y. Qin**, X. Yan. Creative and Context-Aware Translation of East Asian Idioms with GPT-4. *Findings of Empirical Methods in Natural Language Processing* (**Findings of EMNLP**), 2024.

22. L. Liu and **Y. Qin**. Fast Decision Boundary based Out-of-distribution Detection. *International Conference on Machine Learning* (**ICML**), 2024.

21. M. Dhaliwal, K. Tang, E. Aiello, D. Zaharieva, R. Lal, C. Summers, B. Arbiter, K. Watson, L. Figg, I. Balistreri, R. Kingman, B. Suh and **Y. Qin**. Understanding Hypoglycemia Risk in Unstructured Real-World Physical Activities in Adults with Type 1 Diabetes. *American Diabetes Association* (**ADA**), 2024.

20. M. Dhaliwal, K. Tang, E. Aiello and **Y. Qin**. Glycemic Effect of Free-Living Activities in Adults with Type 1 Diabetes. *American Diabetes Association* (**ADA**), 2024.

19. A. Hua, J. Gu, Z. Xue, N. Carlini, E. Wong and **Y. Qin**. Initialization Matters for Adversarial Transfer Learning. *Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2024.

18. S. Niazi, N. Aadit, M. Mohseni, S. Chowdhury, **Y. Qin** and K. Camsari. Training Deep Boltzmann Networks with Sparse Ising Machines. ***Nature Electronics***, 2024.

17. X. Zhang, S. Li, X. Yang, C. Tian, **Y. Qin** and L. Petzold. Enhancing Small Medical Learners with Privacy-preserving Contextual Prompting. *International Conference on Learning Representations* (**ICLR**), 2024.

16. B. Puranik, A. Beirami, **Y. Qin**, U. Madhow. Improving Robustness via Tilted Exponential Layer: A Communication-Theoretic Perspective. *Artificial Intelligence and Statistics* (**AISTATS**), 2024.

15. A. Balashankar, X. Wang, **Y. Qin**, N. Thain, B. Packer, E. Chi and A. Beutel. Improving Robustness through Pairwise Generative Counterfactual Data Augmentation. *Findings of Empirical Methods in Natural Language Processing* (**Findings of EMNLP**), 2023.

14. Z. Shi, N. Carlini, A. Balashankar, L. Schmidt, C. Hsieh, A. Beutel and **Y. Qin**. Effective Robustness against Natural Distribution Shifts for Models with Different Training Data. *Advances in Neural Information Processing Systems* (**NeurIPS**), 2023.

13. **Y. Qin**, X. Wang, B. Lakshminarayanan, E. Chi, A. Beutel. What are Effective Labels for Augmented Data? Improving Robustness with AutoLabel. *IEEE Conference on Secure and Trustworthy Machine Learning* (**SaTML**), 2023.

12. J. Zhao, X. Wang, **Y. Qin**, J. Chen, K. Chang. Investigating Ensemble Methods for Model Robustness Improvement of Text Classifiers. *Findings of Empirical Methods in Natural Language Processing* (**Findings of EMNLP**), 2022.

11. **Y. Qin**, C. Zhang, T. Chen, B. Lakshminarayanan, A. Beutel, X. Wang. Understanding and Improving Robustness of Vision Transformers through Patch-based Negative Augmentation. *Advances in Neural Information Processing Systems* (**NeurIPS**), 2022.

10. J. Gu, V. Tresp, **Y. Qin**. Are Vision Transformers Robust to Patch-wise Perturbations? *European Conference on Computer Vision* (**ECCV**), 2022.

9. **Y. Qin**, X. Wang, A. Beutel, E. Chi. Improving Uncertainty Estimates through the Relationship with Adversarial Robustness. *Advances in Neural Information Processing Systems* (**NeurIPS**), 2021.

8. T. Wang, X. Wang, **Y. Qin**, B. Packer, K. Li, J. Chen, A. Beutel, E. Chi. CAT-Gen: Improving Robustness in NLP Models via Controlled Adversarial Text Generation. *Conference on Empirical Methods in Natural Language Processing* (**EMNLP**), 2020.

7. **Y. Qin***, N. Frosst*, S. Sabour, C. Raffel, G. Cottrell and G. Hinton. Detecting and Diagnosing Adversarial Examples with Class-Conditional Capsule Reconstructions. *International Conference on Learning Representations* (**ICLR**), 2020.

6. **Y. Qin**, N. Carlini, I. Goodfellow, G. Cottrell and C. Raffel. Imperceptible, Robust and Targeted Adversarial Example for Automatic Speech Recognition. *International Conference on Machine Learning* (**ICML**), 2019.

5. **Y. Qin**, S. Ancha, J. Nanavati, G. Cottrell, A. Criminisi and A. Nori. Autofocus Layer for Semantic Segmentation. *International Conference on Medical Image Computing & Computer Assisted Intervention* (**MICCAI**), 2018. (**Oral presentation**, 4% acceptance rate)

4. **Y. Qin***, M. Feng*, H. Lu and G. Cottrell. Hierarchical Cellular Automata for Visual Saliency. *International Journal of Computer Vision* (**IJCV**), 2017

3. **Y. Qin**, D. Song, H. Chen, W. Cheng, G. Jiang and G. Cottrell. A Dual- Stage Attention-Based Recurrent Neural Network for Time Series Prediction. *International Joint Conference on Artificial Intelligence* (**IJCAI**), 2017

2. Q. Pan, **Y. Qin**, Y. Xu, M. Tong and M. He. Opinion Evolution in Open Community. *International Journal of Modern Physics C, 1750003*, 2016.

1. **Y. Qin**, H. Lu, Y. Xu and H. Wang. Saliency Detection via Cellular Automata. In *Conference on Computer Vision and Pattern Recognition* (**CVPR**), 2015

Patents

1. **Y. Qin**, X. Wang, B. Lakshminarayanan, E. Chi, A. Beutel. What are Effective Labels for Augmented data? Improving Robustness with AutoLabel.

2. D. Song, H. Chen, G. Jiang, **Y. Qin**. Dual Stage Attention based Recurrent Neural Network for Time Series Prediction.

Teaching & Mentoring

Instructor

1. ECE180: Introduction to Deep Learning (Spring 2024), UC Santa Barbara
2. ECE194: Adversarial Robustness in Machine Learning (Winter 2024, 2025), UC Santa Barbara
3. ECE594: Robustness in Machine Learning (Winter 2023, Fall 24, Winter 25), UC Santa Barbara

Teaching Assistant

1. CSE253: Neural Networks for Pattern Recognition (Winter 2019), UC San Diego
2. CSE190: Neural Networks and Deep Learning (Fall 2017), UC San Diego

Student Mentorship

∗ **Current PhD Students**

1. Mehak Dhaliwal (PhD at UCSB)
2. Andong Hua (PhD at UCSB)
3. Kenan Tang (PhD at UCSB)
4. Youngseok Yoon (PhD at UCSB)

∗ **Previous Students/Interns**

1. Zhouxing Shi (PhD at UCLA)
2. Jieyu Zhao (PhD at UCLA → Assistant Prof. at USC)
3. Ananth Balashankar (PhD at NYU → Research Scientist at Google)
4. Jindong Gu (PhD at University of Munich → Postdoc at University of Oxford)
5. Tianlu Wang (PhD at UVA → Research Scientist at FAIR)

Selected Awards

∗ Regents' Junior Faculty Fellowship Award *2024*

∗ UCSB Faculty Research Grant Award *2023*

∗ Adobe Faculty Research Award *2023*

∗ AI2000 Most Influential Scholar Honorable Mention in AAAI/IJCAI *2022*

∗ Rising Star in EECS *MIT, 2021*

∗ UCSD GSA Travel Grant *UC San Diego, 2019*

∗ MICCAI Travel Award *MICCAI, 2018*

- NIPS Women in Machine Learning Travel Award *NIPS WiML, 2017, 2016*
- Departmental Fellowship *UC San Diego, 2015*
- Outstanding Undergraduate Student Award *Liaoning Province, China, 2015*
- HIWIN Elite Scholarship (*top 15 students university-wide*) *China, 2014*
- Honorable Mention of Mathematical Contest in Modeling *International, 2013*
- National Scholarship *China, 2013, 2012*

Selected Invited Talks

- Data-Driven Machine Learning: Unlocking the Future of Closed-Loop Diabetes Care
  @ NIH-NIDDK Fifth Artificial Pancreas Workshop, 2024
  @ Sansum Diabetes Research Institute, 2024
  @ Endocrine Society AI in Healthcare Summit, 2024
- Effective Robustness against Natural Distribution Shifts for Models with Different Training Data
  @ UCSB Center of Responsible ML Summit, 2023
- Improving Robustness through Safe Data Augmentation *@ ITA Workshop, 2023*
- Invited panelist: Robustness in Machine Learning *@ LatinX in AI Researchp at NeurIPS 2022, 2023*
- Leading a Breakout Session: Robustness of Machine Learning *@ WiML Un-Workshop at ICML 2022, 2023*
- Improving Calibration through the Relationship with Adversarial Robustness *@ ITA Workshop, 2022*
- Understanding and Improving Robustness of Machine Learning Models *@ UCSB/CMU/USC/MPI, 2022*
- What are Effective Labels for Augmented Data? Improving Robustness with AutoLabel *@ UCSD, 2020*
- Detecting, Diagnoising, Deflecting and Designing Adversarial Attacks *@ Google/FAIR/Amazon/Apple, 2019*
- Imperceptible, Robust and Targeted Adversarial Example for ASR *@ Salesforce, 2019*

Professional Services

Area Chair/Workshop Organizer/Conference Reviewer

- (Workshop Organizer) AIM-FM: Advancements In Medical Foundation Models: Explainability, Robustness, Security, and Beyond at NeurIPS *2024*
- (Workshop Organizer) The 3rd New Frontiers in Adversarial Machine Learning at NeurIPS *2024*
- (Summit Organizer) Department of ECE Summit at UCSB *2023*
- (Summit Organizer) Responsible Machine Learning Summit at UCSB *2023*
- (Workshop Organizer) Robustness of Zero/Few-shot Learning in Foundation Models at NeurIPS *2023*
- (Local Arrangement co-Chair) Knowledge Discovery and Data Mining (KDD) *2023*
- (Workshop Organizer) Southern California Data Science Day at KDD *2023*

* (Area Chair) International Conference on Machine Learning (ICML) *2024-2025*
* (Area Chair) International Conference on Learning Representations (ICLR) *2023-2025*
* (Area Chair) International Conference on Computer Vision (ICCV) *2023, 2025*
* (Area Chair) Conference on Computer Vision and Pattern Recognition (CVPR) *2025*
* (Senior Program Committee) AAAI Conference on Artificial Intelligence (AAAI) *2025*
* (Area Chair) Workshop for Women in Machine Learning (WiML) *2019-2022*

Journal Reviewer

* (Reviewer) IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
* (Reviewer) Transaction of the International Society for Music Information (TISMIR))