

Michael Stephen Saxon

saxon@ucsb.edu <https://saxon.me/>

I am seeking full-time opportunities in generative AI modeling, analysis, and assessment research. I am a 5th year PhD candidate working on the semantic analysis of multimodal language & text-to-image models. I build techniques for automatically and objectively characterizing GenAI system capabilities by inspecting their output behavior under carefully constructed assessment inputs. I'm an NSF Fellow with 10 first-author papers in NLP and ML venues. I am fluent in PyTorch & Hugging-Face, and have done 6 research internships in generative text and language understanding including at Amazon, Meta, AMD, and a clinical startup.

Education

University of California, Santa Barbara

Ph.D., Computer Science: 4.0/4.0

Advisors: William Yang Wang, Ph.D.

Santa Barbara, CA

9/2020–6/2025

Arizona State University

MS., Computer Engineering: 3.9/4.0

Advisors: Visar Berisha, Ph.D. & Sethuraman Panchanathan, Ph.D.

Tempe, AZ

8/2018–5/2020

Arizona State University

BSE., Electrical Engineering; Minor, Mathematics: *Magna Cum Laude*

Tempe, AZ


8/2014–8/2018

Publications

Archival (and publicly available, to-be-archival preprints)

○ Mentor ^ Representative ☆ Award

- [p4] **M. Saxon***, M. Khoshnoodi*, F. Jahara*, Y. Lu, A. Sharma, WY. Wang, “Who Evaluates the Evaluations? Assessing the Faithfulness and Consistency of Text-to-Image Evaluation Metrics with T2IScoreScore,” [arXiv:2403.11092](https://arxiv.org/abs/2403.11092) **NeurIPS 2024**, *Spotlight (5% of subs.)*, Dec 2024
- [p7] **M. Saxon***, A. Sharma*, WY. Wang, “Losing Visual Needles in Image Haystacks: Vision Language Models are Easily Distracted in Short and Long Contexts,” [arXiv:2406.16851](https://arxiv.org/abs/2406.16851), **EMNLP 2024**.
- [c23] **M. Saxon**, A. Holtzman, P. West, WY. Wang, N. Saphra, “Benchmarks as Microscopes: A Call for Model Metrology,” [arXiv:2407.16711](https://arxiv.org/abs/2407.16711) **COLM 2024**
- [p6] Q. Wu, H. Zhao, **M. Saxon**, T. Bui, WY. Wang, Y. Zhang, S. Chang, “VSP: Assessing the dual challenges of perception and reasoning in spatial planning tasks for VLMs,” *Under review*, 2024
- [ur] E. Tanwar, A. Chatterjee, **M. Saxon**, A. Albalak, WY. Wang, T. Chakraborty, “Do You Know About My Nation? An Investigation into Cultural Literacy of Multilingual Large Language Models,” *Under review*, 2024
- [p5] W. Feng, J. Li, **M. Saxon**, T. Fu, W. Chen, WY. Wang, “TC-Bench: Benchmarking Temporal Compositionality in Text-to-Video and Image-to-Video Generation,” [arXiv:2406.08656](https://arxiv.org/abs/2406.08656), June 2024.
- [c22] **M. Saxon***, Y. Luo*, S. Levy, C. Baral, Y. Yang, WY. Wang, “Lost in Translation? Translation Errors and Challenges for Fair Assessment of Text-to-Image Models on Multilingual Concepts,” **NAACL 2024**, *Oral (5% of subs.)* [arXiv:2403.11092](https://arxiv.org/abs/2403.11092) June 2024.
- [J21] L. Pan, **M. Saxon**, W. Xu, D. Nathani, X. Wang, WY. Wang, “Automatically Correcting Large Language Models: Surveying the landscape of diverse self-correction strategies,” [arXiv:2308.03188](https://arxiv.org/abs/2308.03188), **Trans. of the ACL (TACL)** May 2024.

- [c20] V. Himakunthala*, A. Ouyang*, D. Rose*, R. He*, A. Mei, Y. Lu, C. Sonar, **M. Saxon**, WY. Wang,  “Let’s Think Frame by Frame with VIP: A Video Infilling and Prediction Dataset for Evaluating Video Chain-of-Thought,” **EMNLP 2023**, [arXiv:2305.13903](#), Dec 2023
- [c19] X. Wang, W. Zhu, **M. Saxon**, M. Steyvers, WY. Wang, “Large Language Models Are Implicitly Topic Models: Explaining and Finding Good Demonstrations for In-Context Learning,” **NeurIPS 2023**, [arXiv:2301.11916](#), Dec 2023
- [c18] **M. Saxon**, WY. Wang, “Multilingual Conceptual Coverage in Text-to-Image Models,” **ACL 2023**;  **FAccT 2023 Oral** [arXiv:2306.01735](#), [[oral presentation link](#)] Jul 2023.
- [c17] Y. Tuan, A. Albalak, W. Xu, **M. Saxon**, C. Pryor, L. Getoor, WY. Wang, “CausalDialogue: Modeling Utterance-level Causality in Conversations,” **ACL 2023 F** [arXiv:2212.10515](#), Jul 2023.
- [p2] D. Rose*, V. Himakunthala*, A. Ouyang*, R. He*, A. Mei, Y. Lu, **M. Saxon**, C. Sonar, D. Mirza, WY. Wang,  “Visual Chain of Thought: Bridging Logical Gaps with Multimodal Infillings,” *preprint*, [arXiv:2305.02317](#), May 2023.
- [p1] **M. Saxon***, A. Mei*, S. Chang, ZC. Lipton, WY. Wang, “Users are the North Star for AI Transparency,” *preprint*, [arXiv:2303.05500](#), Mar 2023.
- [c16] **M. Saxon**, X. Wang, W. Xu, WY. Wang, “PECO: Examining Single Sentence Label Leakage in Natural Language Inference Datasets,” **EACL 2023** [arXiv:2112.09237](#), May 2023.
- [c15] M. Ho*, A. Sharma*, J. Chang*, **M. Saxon**, S. Levy, Y. Lu, WY. Wang, “WikiWhy: Answering and Explaining Cause-and-Effect Questions,” **ICLR 2023** *Oral (top 5%)* [arXiv:2210.12152](#), May 2023.
- [c14] X. Wang, **M. Saxon**, J. Li, H. Zhang, K. Zhang, WY. Wang, “Causal Balancing for Domain Generalization,” **ICLR 2023** [arXiv:2206.05263](#), May 2023.
- [c13] W. Xu, Y. Tuan, Y. Lu, **M. Saxon**, L. Li, WY. Wang, “Not All Errors are Equal: Learning Text Generation Metrics using Stratified Error Synthesis,” **EMNLP 2022 F** [arXiv:2210.05035](#), Dec 2022.
- [c12] W. Xu, **M. Saxon**, M. Sra, WY. Wang, “Self-Supervised Knowledge Assimilation for Expert-Layman Style Transfer,” **AAAI 2022** [arXiv:2110.02950](#), Jan 2022.
- [c11] X. Wang, W. Chen, **M. Saxon**, WY. Wang, “Counterfactual Maximum Likelihood Estimation for Training Deep Networks,” **NeurIPS 2021** [arXiv:2106.03831](#), Dec 2021.
- [c10] **M. Saxon**, S. Levy, X. Wang, A. Albalak, WY. Wang, “Modeling Disclosive Transparency in NLP Application Descriptions,” **EMNLP 2021** *Oral (8% of subs.)* [arXiv:2101.00433](#), pp. 2023–2037.
- [c9] **M. Saxon**, S. Choudhary, J. McKenna, A. Mouchtaris, “End-to-End Spoken Language Understanding for Generalized Voice Assistants,” **Interspeech 2021**, pp. 4738–4742.
- [c8] S. Levy, **M. Saxon**, WY. Wang, “The Truth is Out There: Investigating Conspiracy Theories in Text Generation,” [arXiv:2101.00379](#), **Findings of ACL 2021**, pp. 4718–4729.
- [j7] **M. Saxon**, A. Tripathi, Y. Jiao, J. Liss, V. Berisha, “Robust Estimation of Hypernasality in Dysarthria,” **IEEE Trans. on Audio, Speech, and Language Processing** 2020, Vol. 28, pp. 2511–2522.
- [c6] **M. Saxon***, J. McKenna*, S. Choudhary*, G. Strimel, A. Mouchtaris, “Semantic Complexity in End-to-End Spoken Language Understanding,” **Interspeech 2020**, pp. 4273–4277.
- [c5] M. Moore, P. Papreja, **M. Saxon**, V. Berisha, S. Panchanathan, “UncommonVoice: A Crowdsourced Dataset of Dysphonic Speech,” **Interspeech 2020**, pp. 2532–2536.
- [c4] M. Moore, **M. Saxon**, H. Venkateswara, V. Berisha, S. Panchanathan, “Say what? A dataset for exploring the error patterns that two ASR engines make,” **Interspeech 2019**, pp. 2528–2532.
- [c3] **M. Saxon**, J. Liss, V. Berisha, “Objective Measures of Plosive Nasalization in Hypernasal Speech,” 2019 **IEEE ICASSP 2019**, pp. 6520–6524.

- [w2] **M. Saxon***, S. Bhandari*, L. Ruskin, G. Honda, "Word Pair Convolutional Model for Happy Moment Classification," **2nd Workshop on Affective Content Analysis, AAAI 2019**, pp. 111–119. (Workshop Oral; CL-Aff Shared task runner up, 2/47)
- [c1] T. Houghton, **M. Saxon**, Z. Song, H. Nyugen, H. Jiang and H. Yu, "2D Grating Pitch Mapping of a through Silicon Via (TSV) and Solder Ball Interconnect Region Using Laser Diffraction" **IEEE 66th Electronic Components and Technology Conference (ECTC) 2016**, pp. 2222–2227. (Texas Instruments Best Student Interactive Paper Award)

Select Non-archival Presentations

- [n4] **M. Saxon**, WY. Wang, "Disparities in Text-to-Image Model Concept Possession Across Languages," **FAccT 2023 Oral** (Non-archival) [OpenReview: 5H2m3tCEaQ](#), Jun 2023.
- [n3] **M. Saxon**, X. Wang, W. Xu, WY. Wang, "Automated Cheating Feature Semantic Identification for NLI Datasets," **SoCal NLP 2022**, Nov 2022.

Professional Experience

AMD (Open Source Generative AI) Bellevue, WA
Research Intern 6/2024–10/2024
Assessing the impact of multimodal regularization and regression objectives in vision-language model representation space training dynamics and end performance.

Meta (Facebook Conversational AI) Menlo Park, CA
Research Intern 6/2022–10/2022
Investigating how different performance metrics metrics for task-oriented dialogue diverge under catastrophic forgetting in continual learning for conversational agents.

Amazon (Alexa Web-based Question Answering) Manhattan Beach, CA
Applied Science Intern 6/2021–9/2021
Training ASR-error-robust retrieval models with mixed phoneme/lexeme tokenization.

Amazon (Alexa Edge ML) Pittsburgh, PA
Applied Science Intern (2x) 5/2019–8/2019, 1/2020–8/2020
Demonstrated how task difficulty predictable distorts benchmark performance for spoken language understanding; trained end-to-end SLU models through differentiable ASR/NLU interface.

Aural Analytics Scottsdale, AZ
Research Engineer Intern 12/2018–4/2019
ASR implementation for speech-based clinical neurological health assessment product.

Invited Talks

Rising Stars in Generative AI Workshop, University of Massachusetts, Amherst 9/2024
Allen Institute for AI, Company Talk 9/2024
Stanford University SALT Group Presentation 8/2024
University of Maryland, College Park UMD CLIP Seminar 5/2024
Georgetown University NLP Group Presentation 5/2024
University of Maryland, Baltimore Perception, Prediction, and Reasoning Seminar 4/2024

Arizona State University *Active Perception Group Presentation* 11/2023
USC Information Sciences Institute *Natural Language Processing Seminar* 11/2023

Service

Program Co-Chair, 2022 Southern California NLP Workshop (SoCalNLP) 11/2022
Reviewer, AACL, EMNLP, ACL, EACL, NeurIPS, ICLR, ACM FAccT, ICASSP, Interspeech 2020–*present*
Mentor, FIRST Robotics Team 2478 (Westwood Robotics), Mesa, AZ 2014–2016

Mentoring

Mahsa Khoshnoodi	2023–2024	Fatima Fellowship Mentee	⇒	PhD., Stony Brook University
Fatima Jahara	2023–2024	Fatima Fellowship Mentee	⇒	PhD., Rutgers University
Matthew Ho	2021–2024	UCSB Undergrad	⇒	Ph.D., UC San Diego

Honors, Fellowships, Scholarships

Rising Star in Generative AI *1/9 selectees, UMass Amherst Rising Stars Workshop* 2024
Google PhD Fellowship Nominee *One of 4 selected by UCSB* 2024
Neal Fenzi—Resonant Founder Fellowship *University of California, Santa Barbara* 2024
Outstanding Reviewer Award, ACL 2023 6/2023
National Science Foundation Graduate Research Fellowship *(NSF GRFP)* 2020
Center for Responsible Machine Learning Fellowship *University of California, Santa Barbara* 2020
Graduate Division Central Fellowship *University of California, Santa Barbara* 2020
Presidential Scholarship (Full Tuition) *Arizona State University* 2014