Engineering 2 Office 347B
1156 High Street
Santa Cruz, CA 95060
☎ *+1 (415) 937 1806*
✉ *lgilpin@ucsc.edu*
lgilpin.com
leilanigilpin

Leilani H. Gilpin

Last updated: October 10, 2022

Research Interests

The theories and methodologies towards monitoring, designing, and augmenting machines that can **explain** themselves for diagnosis, accountability, and liability.

Employment

2021-present **UC Santa Cruz**, *Assistant Professor of Computer Science and Engineering (CSE)*

2020-2021 **Sony AI**, *Research Scientist II*

2013-2015 **Palo Alto Research Center (PARC)**, *Member of Technical Staff*

Education

2015–2020 **Ph.D., Electrical Eng. and Computer Science**, *MIT*
Dissertation: Anomaly Detection through Explanations

2011–2013 **M.S., Computational and Mathematical Engineering**, *Stanford University*

2006–2011 **B.S., Computer Science and B.S., Mathematics**, *UC San Diego (UCSD)*
Highest Honors in Computer Science, Honors in Mathematics, Music Minor

Honors and Awards

2020 Rising Stars in EECS

2020 ACM FAccT Travel Award

2018 Nokia Bell Labs Prize Finalist

Finalist for prize that recognizes research that "changes the game" in the field of information and communications technologies by a factor of 10.

2018 AAAI Doctoral Consortium Travel Award

2017 Nokia Bell Labs Prize Semi-finalist

2016 USENIX Security Student Travel Award

2016-2020 MIT University Center for Exemplary Mentoring (UCEM) Sloan Scholar

2015 MIT ODGE Diversity Fellowship

2011-2013 National Science Foundation (NSF) Graduate Research Fellowship

2013 Stanford SSB Health IT Competition 1st Place
2011 Stanford School of Engineering Fellowship
2011 Yahoo! HackU All Stars Finalist
2011 Yahoo! HackU First Place
2011 Yahoo! Excellence Award
2010 CRA Outstanding Undergraduate Researcher Honorable Mention
2009-present Member of Tau Beta Pi and Eta Kappa Nu
2010 Tau Beta Pi Scholarship
2009 Gary C. Reynolds Memorial Scholarship
2009 BAE Scholarship Finalist

Publications

According to Google scholar: 1,791 citations, h-index: 8, i10-index: 7.

Refereed Journal Publications

[1] Ryan Jenkins, Kristian Hammond, Sarah Loehr, and Leilani Gilpin. "Separating Facts and Evaluation: Motivation, Account, and Learnings from a Novel Approach to Evaluating the Human Impacts of Machine Learning". In: *AI and Society*. 2022.

[2] Peter R Wurman, Samuel Barrett, Kenta Kawamoto, James MacGlashan, Kaushik Subramanian, Thomas J Walsh, Roberto Capobianco, Alisa Devlic, Franziska Eckert, Florian Fuchs, Leilani Gilpin, et al. "Outracing champion Gran Turismo drivers with deep reinforcement learning". In: *Nature* 602.7896 (2022), pp. 223–228.

[3] Michal Araszkiewicz, Ilaria Angela Amantea, Saurabh Chakravarty, Robert van Doesburg, Maria Dymitruk, Marie Garin, Leilani Gilpin, Daphne Odekerken, and Seyedeh Sajedeh Salehi. "ICAIL Doctoral Consortium, Montreal 2019". In: *Artif. Intell. Law* 28.2 (2020), pp. 267–280. DOI: 10.1007/s10506-020-09267-z. URL: https://doi.org/10.1007/s10506-020-09267-z.

[4] Ioana Baldini, Clark Barrett, Antonio Chella, Carlos Cinelli, David Gamez, Leilani Gilpin, Knut Hinkelmann, Dylan Holmes, Takashi Kido, Murat Kocaoglu, et al. "Reports of the AAAI 2019 Spring Symposium Series". In: *AI Magazine* 40.3 (2019), pp. 59–66.

[5] Leilani H. Gilpin, Jamie C. Macbeth, and Evelyn Florentine. "Monitoring Scene Understanders with Conceptual Primitive Decomposition and Commonsense Knowledge". In: *Advances in Cognitive Systems* 6 (2018).

[6] Ayesha Bose, Leilani Gilpin, Jamin Agosti, and Quinn Dang. "The Veicl Act: Safety and Security for Modern Vehicles". In: *Willamette L. Rev.* 53 (2016), p. 137.

[7] Juan Liu, Eric Bier, Aaron Wilson, John Alexis Guerra-Gomez, Tomonori Honda, Kumar Sricharan, Leilani Gilpin, and Daniel Davies. "Graph analysis for detecting fraud, waste, and abuse in healthcare data". In: *AI Magazine* 37.2 (2016), pp. 33–46.

Peer Reviewed Conference Publications

[8] Adam Amos-Binks, Dustin Dannenhauer, and Leilani H Gilpin. "Anticipatory Thinking Challenges in Open Worlds: Risk Management". In: *AAAI Spring Symposium on Designing Artificial Intelligence for Open Worlds*. 2022.

[9] Gregory Falco and Leilani H Gilpin. "A stress testing framework for autonomous system verification and validation (v&v)". In: *2021 IEEE International Conference on Autonomous Systems (ICAS)*. IEEE. 2021, pp. 1–5.

[10] Leilani H. Gilpin. "Anticipatory Thinking: A Testing and Representation Challenge for Self-Driving Cars". In: *2021 55th Annual Conference on Information Sciences and Systems (CISS)*. 2021, pp. 1–2. DOI: 10.1109/CISS50987.2021.9400212.

[11] Leilani H. Gilpin, Vishnu Penubarthi, and Lalana Kagal. "Explaining Multimodal Errors in Autonomous Vehicles". In: *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*. 2021, pp. 1–10. DOI: 10.1109/DSAA53316.2021.9564178.

[12] Leilani H. Gilpin. "System-wide Monitoring for Anomaly Detection". In: *Advances in Cognitive Systems* (2020).

[13] Jason R. Wilson, Leilani H. Gilpin, and Irina Rabkina. "A Knowledge Driven Approach to Adaptive Assistance Using Preference Reasoning and Explanation". In: *arXiv preprint arXiv:2012.02904* (2020).

[14] Leilani H. Gilpin. "Explaining Possible Futures for Robust Autonomous Decision-Making". In: *Proceedings of the AAAI Fall Symposium on Anticipatory Thinking* (2019).

[15] Leilani H. Gilpin and Lalana Kagal. "An Adaptable Self-Monitoring Framework for Opaque Machines". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems. 2019, pp. 1982–1984.

[16] Leilani H. Gilpin. "Reasonableness Monitors". In: *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.

[17] Leilani H. Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. "Explaining explanations: An overview of interpretability of machine learning". In: *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*. IEEE. 2018, pp. 80–89.

[18] Leilani H. Gilpin, Danielle M. Olson, and Tarfah Alrashed. "Perception of Speaker Personality Traits Using Speech Signals". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. 2018, LBW514.

[19] Leilani H. Gilpin, Cagri Zaman, Danielle Olson, and Ben Z Yuan. "Reasonable perception: Connecting vision and language systems for validating scene descriptions". In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2018, pp. 115–116.

[20] Leilani H. Gilpin and Ben Ze Yuan. "Getting Up to Speed on Vehicle Intelligence". In: *AAAI Spring Symposium Series*. 2017.

[21] Leilani Gilpin, Laurent Ciarletta, Yannick Presse, Vincent Chevrier, and Virginie Galtier. "Co-simulation solutions using AA4MM-FMI applied to smart space heating models". In: *Proceedings of the 7th International ICST Conference on Simulation Tools and Techniques*. ICST (Institute for Computer Sciences, Social-Informatics and ... 2014, pp. 153–159.

Peer Reviewed Workshoop Publications

[22] Leilani H. Gilpin. "Reconciling System-wide Errors with Symbolic Explanations". In: *Proceedings of the IJCAI Workshop on AI for Anomaly Detection* (2020).

[23] Leilani H. Gilpin. "Monitoring Opaque Learning Systems". In: *ICLR 2019 Debugging ML Models Workshop* (2019).

[24] Leilani H. Gilpin, Tianye Chen, and Lalana Kagal. "Learning From Explanations for Robust Autonomous Driving." In: *ICML Workshop on AI for Autonomous Driving*. 2019.

[25] Leilani H. Gilpin, Cecilia Testart, Nathaniel Fruchte, and Julius Adebayo. "Explaining Explanations to Society". In: *arXiv preprint arXiv:1901.06560* (2018).

Other Publications

[26] Leilani Hendrina Gilpin. "Anomaly detection through explanations". PhD thesis. Massachusetts Institute of Technology, 2020.

[27] Karianne Bergen and Leilani Gilpin. *Negative News No More: Classifying News Article Headlines*. Tech. rep. 11. 2012, p. 2017.

[28] Leilani Gilpin. *The Impact of Topology and Communication Models on Connectivity in Networks*. 2011.

Invited Talks and Panels

Accountability Layers

Talks

University of Birmingham (joint with R. Marinescu)	*2022*
University College London	*2022*
UCSC Applied Mathematics Seminar (scheduled)	*2022*

Explaining Errors in Autonomous Driving

JHU Department of Civil & Systems Engineering	*2021*

Anomaly Detection through Explanations

55th Annual Conference on Information Sciences and Systems	*2021*
FUZZ-IEEE	*2021*
Cornell Computing and Information Science	*2021*
Utah School of Computing	*2021*
USC Computer Science	*2021*
UC Santa Cruz CSE	*2021*
Institute of Assured Autonomy at JHU	*2020*

	31st International Workshop on Principles of Diagnosis (DX)	*2020*
	USC - Information Science Institute	*2020*
	ICML Workshop on Monitoring ML Systems	*2020*
	Sony AI	*2020*
	Northeastern University - Experiential AI	*2020*
	Northwestern University - Computer Science	*2020*
	Salesforce Research (canceled due to Covid-19)	*2020*
	Rochester Data Science Consortium (postponed due to Covid-19)	*2020*
	Idexx (postponed due to Covid-19)	*2020*
	Explaining Explanations in AI	
	Statistics Canada (STATCAN)	*2021*
	TU Berlin	*2020*
	U. of Cambridge Distributed Info. and Automation Lab (DIAL) XAI Day	*2020*
	Stanford University - Knowledge Graphs Seminar	*2020*
	The Car Can Explain!	
	UC San Diego - Halicioglu Data Science Institute	*2019*
	CSAIL-Toyota Meeting	*2018*
	MIT Museum	*2018*
	Columbia Law School: Software Freedom Law Center	*2018*
	CSAIL-Toyota Meeting	*2017*
Panelist	ICLR Workshop on AI for Earth and Space Sciences	*2022*
	ESIP Winter Meeting visionary panel on knowledge graphs	*2021*
	Northwestern ML Impact Initiative: Intelligibility, Fairness, Transparency	*2021*
	Columbia Law School: Software Freedom Law Center	*2019*

Grants

PI **Underwriters Laboratory (subaward under Northwestern University)**. *Adversarial Examples to Test Explanation Robustness*. March 1, 2022-April 1, 2023. Award Amount: $125,477.00.

UCSC Office of Research Seed Funding for Early Stage Initiatives. *Robustifying Machine Learning for Safe and Secure Autonomous Vehicles*. March 15, 2022 - March 31, 2023. Alvaro Cardenas (Co-PI), Daniel Fremont (Co-PI), and Cihang Xie (Co-PI). Award Amount: $40,000.00.

Co-author CSAIL-TRI: "The Car Can Explain!" *2017*

Tutorials

AAAI — Trustworthy and Responsible AI — *2023*
Co-lead with Yilun Zhou, Jieyu Zhao, Harsha Nori, and Besmira Nushi.

US2TS — Knowledge-based commonsense reasoning and explainability — *2022*
Co-lead with Filip Ilievski of USC ISI.

Selected Press

- Robustifying AVs Project Chosen for Seed Funding. [UCSC News] .
- Mob.ly App Makes Driving Safer by Changing How Drivers Navigate. [AI Pulse Report] .
- MIT CSAIL Student Spotlight. [Student Profile]
- MIT student lead AI and Ethics Reading Group. [MIT News].
- MIT Internet Policy Research Initiative (IPRI) [Student profile].

Teaching

Lead Instructor

UCSC — Responsible Data Science (CSE 246) — *Fall 2022*
Artificial Intelligence (CSE 140) — *Winter 2021, Spring 2022*

MIT — Artificial Intelligence and Global Risks — *IAP 2018*
Developed, taught, managed a new course on the risks of AI from a global perspective. [course webpage].

Stanford — SMASH Institute - Calculus — *Summer 2015*
Planned and lead weekly lectures to teach a semester-long calculus class over the summer.

Guest Lectures

UCSC — Projects in Artificial Intelligence: CSE 247 — *Spring 2022*
Research & Teaching in CSE: CSE 200 — *Winter 2022*

Emmanuel — Practical Machine Learning (Invited talk on GANs and XAI) — *Spring 2022*

Northwestern — CS 496: AI Perspectives (Invited talk on XAI) — *Spring 2021/2022*

Stanford — CS 520: Knowledge Graphs (Invited talk on XAI) — *Spring 2020*

MIT — 6.905/6.945: Large-scale Symbolic Systems — *Spring 2019*
6.S978: Privacy Legislation in Practice: Law and Technology — *Spring 2017*

Teaching Assistant

MIT — 6.905/6.945: Large-scale Symbolic Systems — *Spring 2019*

Stanford — CS 348A: Geometric Modeling (PhD Level Course) — *Spring 2013*

UCSD	COGS 5A (beginning java)	
	CSE 8A/8B (beginning java)	
	CSE 5A (beginning C)	
	CSE 21 (discrete mathematics)	
	CSE 100 (Advanced Data Structures)	
	CSE 101 (Algorithms)	

Mentoring

PhD students (primary supervisor)

	Oliver Chang	*2022-present (incoming)*

Visiting PhD students (primary supervisor)

Sapienza U.	Biagio La Rosa	*2022-present (incoming)*

PhD student committee service (other supervisor)

Adv. Committee	Michael Briden	*2022*
	Anuj Kaul	*2022*

Master's Students

Thesis Advisor	Shengjie Xu	*2021-present*
	Sainaya Brid	*2022-present*
Project Advisor	Rohit Das	*2022-present*
	Olivia Anastassov	*2022-present*
	Zhihang Zhou	*2022*
Project Reader	Rahul Bharadwaj	*2022-present*

Undergraduate Students (UCSC)

Primary Advisor	Tanisha Khemka	*2022-present*
	Suneet Katrekar	*2022-present*
	Owen Shi	*2022-present*
	Nick Wang	*2022-present*
	Sahil Gupta	*2022-present*
	Ariel Kamen	*2022-present*
	Batuhan Salih	*2022-present*
	Lan Mi	*2022*
	Radhika Gadre	*2022*
	Aylin Akkus (visiting student)	*2022*
	Christopher Oey	*2022*
	Aaja Ouellette	*2022*
	Isaac Plotkin	*2022*

Co-supervisor	Coen Adler	*2022-present*
Reading Committee	Nishanth Jayram	*2022*

MIT Thesis Students (12+ month fulltime student)

MEng	Tianye Chen	*2018-2019*
	Co-advised with Lalana Kagal. Co-authored paper on rule-learning [24].	
SuperUROP	Evelyn Florentine	*2017-2018*
	Co-authored journal paper on monitoring opaque learning systems [5].	
	Zoe Lu	*2017-2018*

MIT Research Project Students (6+ month semester course)

UROP	Shayda Moezzi	*Fall 2020-2021*
	Vishnu S Penubarthi	*Fall 2019-2021*
	Co-authored paper on multimodal explanations [11].	
	Marla E. Odell	*Spring 2019*
	Elizabeth Han	*Spring 2019*
	Obada Alkhatib	*IAP/Spring 2018*
	Michal Reda	*IAP/Spring 2018*
	Ishan Pakuwal	*IAP/Spring 2018*
UAP	Matthew Kalinowski	*Spring 2017*

University Service

Dept. Service	Graduate Admissions Committee	*2022-present*
	AI Teaching Faculty Search	*2022*
Committee	External Member: HCI Associate Prof. Search (UCSC CM Dept.)	*2022*

Professional Activities

Program Chair	Advances in Cognitive Systems (ACS) [link]	*2022*
Organizer	NeurIPS workshop on Building Trustworthy Embodied AI	*2022*
	AAAI Fall Symposium on Anticipatory Thinking [link]	*2020/2021*
	NeurIps Workshop on XAI for debugging and diagnosis [link].	*2021*
	ACS Workshop on Story Enabled Intelligence. [link].	*2019*
	AAAI Spring Symposium 2019: Story-Enabled Intelligence. [link].	*2019*
	MIT Machine Learning Interpretability Reading Group	*2018-2020*
	MIT AI and Ethics Reading Group. [link].	*2018-2020*
	MIT IPRI Privacy, Security and Policy (PSP) Meeting	*2018-2019*
	MIT Path of Professorship Workshop	*2018*

	MIT EECS Visit Days and Orientation	*2016*
Advisory Board	Center for Advancing Safety of Machine Intelligence (CASMI)	*2022-present*
	Governance Advisory Committee. [CASMI Leadership]	
Local Chair	Advances in Cognitive Systems	*2019*
Senior PC	AAAI	*2022-present*
PC	Advances in Cognitive Systems (ACS)	*2022-present*
	Semantic Web Journal issue on Commonsense Knowledge and Reasoning	*2021*
	AAAI	*2021-2022*
	ACS Workshop on Goal Reasoning	*2021*
	AAMAS	*2021*
	DX-2020	*2020*
Reviewer	AI4HRI	*2022*
	Information Issue on Foundations and Challenges of Interpretable ML	*2021*
	HRI	*2020-present*
	ACS Workshop on Goal Reasoning	*2021*
	Artificial Intelligence Review	*2020*
	IEEE Transactions on Cybernetics	*2019*
	NeurIPS	*2019*
	AAAI Spring Symposium	*2019*
	Slovak-Israeli Scientific Research Program	*2018*
	MIT MITES	*2018*
	HRI Late Breaking Reports (LBR)	*2018*
	AAAI (Guest Reviewer)	*2015*
Student Rep.	Stanford ICME	*2011-2013*
	MIT EECS Visiting Committee	*2017*
	Met with the EECS Visiting Committee and gave a personal perspective on the EECS Department, student life, and diversity.	
	MIT Grad Rat	*2017-2019*
Volunteer	UCSD Alumni Board	*2015-2019*

Diversity, Equity and Inclusion

	UCSC ACM-W Board Faculty Sponsor	*2022-present*
	Santa Cruz County Science & Engineering Fair Judge	*2022*
Mentor	Women in ML (WiML) PhD Mentor	*2022-present*
	Black in AI	*2020*
	MIT EECS GAAP	*2020*

The Graduate Application Assistance Program (GAAP) is a student-run, volunteer-based program which provides assistance to EECS PhD applicants from under-represented groups, including students from groups underrepresented in STEM and students with non-traditional academic backgrounds.

Xerox ABI Mentoring Program *2014-2015*