

Baskin Engineering Building  
1156 High Street Office 347B  
Santa Cruz, CA 95060  
☎ +1 (415) 937 1806  
✉ [lgilpin@ucsc.edu](mailto:lgilpin@ucsc.edu)  
🌐 [lgilpin.com](http://lgilpin.com)  
🐦 [leilanigilpin](https://twitter.com/leilanigilpin)

# Leilani H. Gilpin

## 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

## Selected Publications

Full publication list at [\[http://lgilpin.com#publications\]](http://lgilpin.com#publications).

- DSAA 2021 **L.H. Gilpin**, Vishnu Penubarthi, and Lalana Kagal. “Explaining Multimodal Errors in Autonomous Vehicles.” *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, 2021, pp. 1-10, doi: 10.1109/DSAA53316.2021.9564178.
- ICAS 2021 Gregory Falco and **L.H. Gilpin**. “A stress testing framework for autonomous system verification and validation (v&v)”. *In IEEE International Conference on Autonomous Systems (IEEE, 2021)*.
- AAMAS 2019 **L. H. Gilpin** and Lalana Kagal. “An Adaptable Self-Monitoring Framework for Opaque Machines.” *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. [\[pdf\]](#).

## 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

---

## All Publications

- [1] 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)*, pages 1–5. IEEE, 2021.
- [2] **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)*, pages 1–2, 2021.
- [3] **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)*, pages 1–10, 2021.
- [4] Michal Araszkiwicz, 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. *Artif. Intell. Law*, 28(2):267–280, 2020.
- [5] **Leilani H. Gilpin**. Reconciling system-wide errors with symbolic explanations. *To appear in the Proceedings of the IJCAI Workshop on AI for Anomaly Detection*, 2020.
- [6] **Leilani H. Gilpin**. System-wide monitoring for anomaly detection. *Advances in Cognitive Systems*, 2020.

- [7] Jason R. Wilson, **Leilani H. Gilpin**, and Irina Rabkina. A knowledge driven approach to adaptive assistance using preference reasoning and explanation. *arXiv preprint arXiv:2012.02904*, 2020.
- [8] Ioana Baldini, Clark Barrett, Antonio Chella, Carlos Cinelli, David Gamez, **Leilani Gilpin**, Knut Hinkelmann, Dylan Holmes, Takashi Kido, Murat Kocaoglu, and others. Reports of the aai 2019 spring symposium series. *AI Magazine*, 40(3):59–66, 2019.
- [9] **Leilani H. Gilpin**. Explaining possible futures for robust autonomous decision-making. *Proceedings of the AAAI Fall Symposium on Anticipatory Thinking*, 2019.
- [10] **Leilani H. Gilpin**. Monitoring opaque learning systems. *ICLR 2019 Debugging ML Models Workshop*, 2019.
- [11] **Leilani H. Gilpin**, Tianye Chen, and Lalana Kagal. Learning from explanations for robust autonomous driving. In *ICML Workshop on AI for Autonomous Driving*, 2019.
- [12] **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*, pages 1982–1984. International Foundation for Autonomous Agents and Multiagent Systems, 2019.
- [13] **Leilani H. Gilpin**. Reasonableness monitors. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [14] **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)*, pages 80–89. IEEE, 2018.
- [15] **Leilani H. Gilpin**, Jamie C. Macbeth, and Evelyn Florentine. Monitoring scene understanders with conceptual primitive decomposition and commonsense knowledge. *Advances in Cognitive Systems*, 6, 2018.
- [16] **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*, page LBW514. ACM, 2018.
- [17] **Leilani H. Gilpin**, Cecilia Testart, Nathaniel Fruchte, and Julius Adebayo. Explaining explanations to society. *arXiv preprint arXiv:1901.06560*, 2018.
- [18] **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*, pages 115–116. ACM, 2018.
- [19] **Leilani H. Gilpin** and Ben Ze Yuan. Getting up to speed on vehicle intelligence. In *AAAI Spring Symposium Series*, 2017.
- [20] Ayesha Bose, **Leilani Gilpin**, Jamin Agosti, and Quinn Dang. The veicl act: Safety and security for modern vehicles. *Willamette L. Rev.*, 53:137, 2016.

- [21] 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. *AI Magazine*, 37(2):33–46, 2016.
- [22] **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*, pages 153–159. ICST (Institute for Computer Sciences, Social-Informatics and . . . , 2014.
- [23] Karianne Bergen and **Leilani Gilpin**. Negative news no more: Classifying news article headlines. Technical Report 11, 2012.
- [24] **Leilani Gilpin**. The impact of topology and communication models on connectivity in networks, 2011.

[Working Papers and Papers Under Review](#)

**Leilani H. Gilpin**, Mohammed A. Alam, Andrew R. Paley, Sarah Loehr, and Kristian J. Hammond. “Explanation is Not a Technical Term: The Problem of Ambiguity in XAI”

**Leilani H. Gilpin** and Gerald Jay Sussman. “Should we Fear Autonomous Machines?”

**Leilani H. Gilpin** and Greg Falco. “A Stress Testing Framework for Cyberphysical AI.”

**Talks**

Invited Talks	<p><a href="#">Explaining Errors in Autonomous Driving</a></p> <p>JHU Department of Civil Systems Engineering <span style="float: right;">2021</span></p> <p><a href="#">Anomaly Detection through Explanations</a></p> <p>55th Annual Conference on Information Sciences and Systems <span style="float: right;">2021</span></p> <p>FUZZ-IEEE <span style="float: right;">2021</span></p> <p>Cornell Computing and Information Science <span style="float: right;">2021</span></p> <p>Utah School of Computing <span style="float: right;">2021</span></p> <p>USC Computer Science <span style="float: right;">2021</span></p> <p>UC Santa Cruz CSE <span style="float: right;">2021</span></p> <p>Institute of Assured Autonomy at JHU <span style="float: right;">2020</span></p> <p>31st International Workshop on Principles of Diagnosis (DX) <span style="float: right;">2020</span></p> <p>USC - Information Science Institute <span style="float: right;">2020</span></p> <p>ICML Workshop on Monitoring ML Systems <span style="float: right;">2020</span></p> <p>Sony AI <span style="float: right;">2020</span></p> <p>Northeastern University - Experiential AI <span style="float: right;">2020</span></p> <p>Northwestern University - Computer Science <span style="float: right;">2020</span></p> <p>Salesforce Research (canceled due to Covid-19) <span style="float: right;">2020</span></p> <p>Rochester Data Science Consortium (postponed due to Covid-19) <span style="float: right;">2020</span></p>
---------------	---

	Idexx (postponed due to Covid-19)	2020
	<a href="#">Explaining Explanations in AI</a>	
	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
	<a href="#">The Car Can Explain!</a>	
	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
Conference	Advances in Cognitive Systems (ACS)	2020
	AAAI Fall Symposium on Cognitive Systems for Anticipatory Thinking	2019
	17th International Conference on Artificial Intelligence and Law (ICAAIL)	2019
	AAAI Spring Symposium on Story Enabled Intelligence	2019
	NeurIPS Workshop on Ethical, Social and Governance Issues in AI	2018
	The 5th IEEE Conference on Data Science and Advanced Analytics (DSAA)	2018
	Advances in Cognitive Systems (ACS)	2018
	The Twenty-Third AAAI/SIGAI Doctoral Consortium	2018
	SIMUTOOLS:Conference on Simulation Tools and Techniques	2014
	Workshop MS4SG: Multisimulation for Smart Grids (EDF-INRIA).	2013
Poster	Women in Data Science Cambridge Conference	2020
	ICLR Workshop on Debugging ML models	2019
	CSAIL-Toyota Meeting	2019
	CSAIL Alliances Meeting	2019
	Women in Data Science Cambridge Conference	2019
	MIT College of Computing Poster Session	2019
	MIT QI Symposium on Robust, Interpretable Deep Learning Systems	2018
	CSAIL-Toyota Meeting	2018
	CSAIL Alliances Meeting	2018
	New England Machine Learning Day	2018
	2nd NorthEast Computational Health Summit AI in Healthcare (NECHS)	2018
	SDSCon (Statistics and Data Science Center Conference)	2018
	MIT IQ Launch	2018
	Workshop on Human Centric AI for Intelligent Machines	2017
	The Cambridge Cyber Security Summit	2016
	USENIX Security	2016

---

## Grants

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

2017

---

## Selected Press

- 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) [\[Student profile\]](#).

---

## Research Experience

### Research Assistant

MIT CSAIL **The Car Can Explain!** (2015-2020)

Developing techniques to allow self-driving cars and other AI-driven systems to explain behaviors and failures. [\[project webpage\]](#).

**Punya** (2015)

Examined how to represent and present anomalous events when using sensitive (PII) data.

Stanford **Geometric Computing Group** (2011-2013)

Worked on developing maps to understand brain geometry in medicine. [\[group alumni webpage\]](#).

**Autonomous Systems Laboratory** (2013)

Worked on queueing for the last mile problem in autonomous systems in cities.

UCSD **Geometric Mechanics Group** (2009-2011)

Worked in the geometric mechanics group on robotic networks and optimization of numerical methods. Completed honors thesis on distributed algorithms for communication networks and robotic networks.

### Member of Technical Staff

PARC **Intelligent Systems Laboratory** (2013-2015)

Integrated Python and R-scripts into the automatic extract-transform-load (ETL) process. Started preliminary work on reason codes and explanations for medical anomalies

### Research Intern

INRIA **MADYNES Group** (Summer 2013)

Worked as part of the MADYNES group on smart space models. Project Title: The Impact of Communication Models for Demand Response in Smart Grid Co-simulation. Published and presented a first-author paper [22] with results.

DIMACS **Communication Networks** (Summer 2010)

Completed research project on convergence guarantees for communication models. Attended the Midsummer Combinatorial Workshop in Prague with the DIMACS/DIMATIA exchange program.

NEES **San Diego Supercomputer Center-NEESIT Intern** (Summer 2009)

Completed project on data visualization of earthquake test data. Developed an earthquake test site application using the Google Maps API.

CAIDA **San Diego Supercomputer Webmaster Assistant** (2008-2009)

Performed research on web-based applications and assisted with web infrastructure.

### Technical Experience

Salesforce **Data.com Software Engineering Intern** (Summer 2012)

Worked as part of the Data.com group to develop a statistical classifier and machine learning algorithm for detecting fraud in contact data.

---

## Teaching Experience

### Lead Instructor

UCSC Intro to AI (CSE 140) *Winter 2021*

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.*

### Lectures

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

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

### MIT Thesis Students (12+ month fulltime student)

MEng	Tianye Chen	2018-2019
	<i>Co-advised with Lalana Kagal. Co-authored paper on rule-learning [11].</i>	
SuperUROP	Evelyn Florentine	2017-2018
	<i>Co-authored journal paper on monitoring opaque learning systems [15].</i>	
	Zoe Lu	2017-2018

### MIT Research Project Students (6+ month semester course)

UROP	Shayda Moezzi	Fall 2020-present
	Vishnu S Penubarthi	Fall 2019-present
	<i>Co-authored paper on multimodal explanations [3].</i>	
	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

### Other MIT Advising

UROP	Yunxing (Lucy) Liao	IAP 2019
Mentor	6.805 (Foundations of Information Policy)	Fall 2017
	6.805 (Foundations of Information Policy)	Fall 2016
	<i>Project mentor for introductory policy class. Met weekly with teams to give high level feedback on ideas, implementations, and writing. Several groups went onto publish their projects.</i>	

### Professional Activities

Organizer	AAAI Fall Symposium on Anticipatory Thinking <a href="#">[link]</a>	2020/2021
	NeurIPS Workshop on XAI for debugging and diagnosis <a href="#">[link]</a> .	2021
	ACS Workshop on Story Enabled Intelligence. <a href="#">[link]</a> .	2019
	AAAI Spring Symposium 2019: Story-Enabled Intelligence. <a href="#">[link]</a> .	2019
	MIT Machine Learning Interpretability Reading Group	2018-present
	MIT AI and Ethics Reading Group. <a href="#">[link]</a> .	2018-present
	MIT IPRI Privacy, Security and Policy (PSP) Meeting	2018-2019
	MIT Path of Professorship Workshop	2018
	MIT EECS Visit Days and Orientation	2016
Local Chair	Advances in Cognitive Systems	2019
PC	Semantic Web Journal issue on Commonsense Knowledge and Reasoning	2021
	ACS Workshop on Goal Reasoning	2021
	DX-2020	2020



	AAMAS	2020
Reviewer	HRI	2020,2021
	ACS Workshop on Goal Reasoning	2021
	AAAI	2020
	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
	<i>Met with the EECS Visiting Committee and gave a personal perspective on the EECS Department, student life, and diversity.</i>	
	MIT Grad Rat	2017-2019
Mentor	Black in AI	2020
	MIT EECS GAAP	2020
	<i>The Graduate Application Assistance Program (GAAP) is a student-run, volunteer-based program which provides assistance to EECS PhD applicants from underrepresented groups, including students from groups underrepresented in STEM and students with non-traditional academic backgrounds.</i>	
	Xerox ABI Mentoring Program	2014-2015
Volunteer	UCSD Alumni Board	2015-2019