32 Vassar Street Office 32G530
Cambridge, MA 02139
☎ +1 (415) 937 1806
✉ lgilpin@mit.edu
people.csail.mit.edu/lgilpin

Leilani H. Gilpin

Research Interests

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

Education

2020 (expected) **Ph.D., Electrical Eng. and Computer Science**, *MIT*.
Advisor: Gerald Jay Sussman, Panasonic Professor of Electrical Engineering

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

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

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*. International Foundation for Autonomous Agents and Multiagent Systems, 2019. [pdf].

DSAA 2018 **L.H. Gilpin**, D. Bau, B.Z. Yuan, A. Bajawal, M. Specter, and L. Kagal. "Explaining Explanations: An Overview of Interpretability of Machine Learning." *2018 IEEE 5th International Conference on data science and advanced applications (DSAA)*. IEEE, 2018. [pdf].

ACS 2018 **L.H. Gilpin**, J.C. Macbeth, and E. Florentine. "Monitoring Scene Understanders with Conceptual Primitive Decomposition and Commonsense Knowledge." *Advances in Cognitive Systems 6* (2018). [pdf].

Honors and Awards

2020 ACM FAT* 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] 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 aaai 2019 spring symposium series. *AI Magazine*, 40(3):59–66, 2019.

[2] **Leilani H. Gilpin**. Explaining possible futures for robust autonomous decision-making. *To appear in the Proceedings of the AAAI Fall Symposium on Anticipatory Thinking*, 2019.

[3] **Leilani H. Gilpin**. Monitoring opaque learning systems. *ICLR 2019 Debugging ML Models Workshop*, 2019.

[4] **Leilani H. Gilpin**, Tianye Chen, and Lalana Kagal. Learning from explanations for robust autonomous driving. In *ICML Workshop on AI for Autonomous Driving*, 2019.

[5] **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.

[6] **Leilani H. Gilpin**. Reasonableness monitors. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[7] **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.

[8] **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.

[9] **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.

[10] **Leilani H. Gilpin**, Cecilia Testart, Nathaniel Fruchte, and Julius Adebayo. Explaining explanations to society. *arXiv preprint arXiv:1901.06560*, 2018.

[11] **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.

[12] **Leilani H. Gilpin** and Ben Ze Yuan. Getting up to speed on vehicle intelligence. In *AAAI Spring Symposium Series*, 2017.

[13] Ayesha Bose, **Leilani Gilpin**, Jamin Agosti, and Quinn Dang. The veicl act: Safety and security for modern vehicles. *Willamette L. Rev.*, 53:137, 2016.

[14] 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.

[15] **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.

[16] Karianne Bergen and **Leilani Gilpin**. Negative news no more: Classifying news article headlines. Technical Report 11, 2012.

[17] **Leilani Gilpin**. The impact of topology and communication models on connectivity in networks, 2011.

Under Review

Leilani H. Gilpin et al. "Anomaly Detection Through Explanations"

Leilani H. Gilpin and Gerald Jay Sussman. "Should we Fear Intelligent Machines?"

Working Papers

Leilani H. Gilpin "Learning Symbolic Rules Through Explanation"

Talks

Invited Talks	Northwestern University	*2020*
	Idexx	*2020*
	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	AAAI Fall Symposium on Cognitive Systems for Anticipatory Thinking	*2019*
	17th International Conference on Artificial Intelligence and Law (ICAIL)	*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 forSmart Grids (EDF-INRIA).	*2013*
Poster	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*

Selected Press

- 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-present)

Currently working on incorporating commonsense reasoning into opaque algorithms. Processed and displayed explanations from simulated CAN bus logs.

Punya (2015)

Examined how to represent and present anomalous events in 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 [15] 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 website 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

MIT Explanatory Artificial Intelligence and Interpretability *IAP 2020*

Upcoming Seminar Course with tutorial and guest lectures on the main topics of explainability and interpretability.

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

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*

Co-advised with Lalana Kagal. Co-authored paper on rule-learning [4].

SuperUROP Evelyn Florentine *2017-2018*

Co-authored journal paper on monitoring opaque learning systems [8].

Zoe Lu *2017-2018*

MIT Research Project Students (6 month semester course)

UROP Vishnu S Penubarthi *Fall 2019-Spring 2020*

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*

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.

Service

Organizer	ACS Workshop on Story Enabled Intelligence. [link].	*2019*
	AAAI Spring Symposium 2019: Story-Enabled Intelligence. [link].	*2019*
	MIT Machine Learning Interpretability Reading Group	*2018*
	MIT AI and Ethics Reading Group. [link].	*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*
Reviewer	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.	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*
Mentor	Xerox ABI Mentoring Program	*2015*
Volunteer	UCSD Alumni Board	*2015-2019*

References

Gerald Jay Sussman
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
✉ gjs@mit.edu
☎ 617-253-5874

Daniel Weitzner
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
✉ djweitzner@csail.mit.edu
☎ 617-253-8036

Lalana Kagal
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
✉ lkagal@csail.mit.edu
☎ 617-253-5845

Luke Fletcher
Toyota Research Institute
✉ luke.fletcher@tri.global