

Isaac Lage

Curriculum vitae

CONTACT INFORMATION	Harvard University Computer Science Cambridge, MA, USA 02138	isaaclage@harvard.edu isaaclage.github.io
---------------------	--	--

EDUCATION	Harvard University , Cambridge, MA School of Engineering and Applied Sciences Doctor of Philosophy, Computer Science	2017-Present
	New York University , New York, NY Bachelor of Arts, Computer Science, Social and Cultural Analysis • <i>Summa cum Laude</i> ; with High Honors in Computer Science; Phi Beta Kappa	2012-2016

SELECTED HONORS	<ul style="list-style-type: none">Selected as Pedagogy Fellow for Harvard’s School of EngineeringNSF Graduate Research Fellowship Recipient, Harvard UniversityNIH BD2K Training Grant Recipient, Harvard UniversityComputer Science Prize for Academic Excellence, New York University	2022-2023 2019-2022 2017-2019 2016
-----------------	--	---

RESEARCH EXPERIENCE	Harvard University , Cambridge, MA <i>Graduate Research Assistant</i> <ul style="list-style-type: none">Under supervision of Dr. Finale Doshi-VelezDeveloping methods for optimizing interpretable models with human feedbackEvaluating the interpretability of models through user studies Microsoft Research , Redmond, WA <i>Research Intern</i> <ul style="list-style-type: none">Under supervision of Dr. Ece KamarOptimized explanations of ML models for Human-AI teamwork Massachusetts Institute of Technology & New York University , Cambridge, MA & New York, NY <i>Jr. Research Assistant & Jr. Research Scientist/Research Software Engineer</i> <ul style="list-style-type: none">Under supervision of Dr. David SontagPredicted hospital readmissions from electronic health recordsModeled the disease progression of multiple myeloma	2017-Present Summer 2020 2016-2017
---------------------	---	--

TEACHING EXPERIENCE	Tufts University , Medford, MA <i>Part-Time Lecturer</i> <ul style="list-style-type: none">Instructor for Statistical Pattern RecognitionTaught this core machine learning offering with ~ 30 students Harvard University , Cambridge, MA <i>Pedagogy Fellow for the School of Engineering and Applied Sciences</i> <ul style="list-style-type: none">Training 35-40 new TAs a semesterCo-leading departmental pedagogy course (in spring 2023)Consulting with TAs about effective teaching strategiesOrganized a panel and orientation session on teaching for all 2nd year PhD students <i>Guest lecturer, Critical Thinking in Data Science</i> <ul style="list-style-type: none">Invited to co-lecture on methods and ethics for human-in-the-loop learning based on my research expertise <i>Guest lecturer, Interpretability and Explainability in ML</i> <ul style="list-style-type: none">Invited to give 2 guest lectures on human evaluation of interpretability based on my research expertise	Spring 2022 2022-2023 Spring 2021, 2022 Spring 2021
---------------------	--	--

<i>Teaching Fellow, Interpretability and Explainability in ML</i>	Fall 2019
<ul style="list-style-type: none"> • Supported all student research projects for the course • Co-designed portions of assignments • Taught a lecture on my research topic 	
<i>Teaching Fellow, Advanced Machine Learning</i>	Fall 2018
<ul style="list-style-type: none"> • Designed and lead 3 problem solving and review sections • Co-taught a lecture on sampling methods • Wrote portions of homework solutions 	
<i>Bok Teaching Certificate</i>	2017-2023
<ul style="list-style-type: none"> • Took coursework on public speaking, curriculum design, and STEM pedagogy • Reflected on student experiences through a teaching observation • Examined teaching values through structured writing assignments 	

CONFERENCE
PUBLICATIONS

Joint first-authorship indicated with *

- [1] **Lage, I.**, McCoy, T. H., Perlis, R. H. & Doshi-Velez F. (2022) Efficiently identifying individuals at high risk for treatment resistance in major depressive disorder using electronic health records. *Journal of Affective Disorders Volume 306*. (**Editor's choice article**)
- [2] **Lage, I.**, Pradier, M. F., McCoy, T. H., Perlis, R. H. & Doshi-Velez F. (2022) Do clinicians follow heuristics in prescribing antidepressants? *Journal of Affective Disorders Volume 311*.
- [3] **Lage, I.***, Chen, E.*, He, J.*, Narayanan, M.*, Kim, B., Gershman, S. & Doshi-Velez, F. (2019) Human Evaluation of Models Built for Interpretability. *AAAI Conference on Human Computation and Crowdsourcing (HCOMP) 2019*. (**Honorable mention for best paper**)
- [4] **Lage, I.***, Lifschitz, D.*, Doshi-Velez, F. & Amir, O. (2019) Exploring Computational User Models for Agent Policy Summarization. *International Joint Conference on Artificial Intelligence (IJCAI) 2019*.
- [5] **Lage, I.**, Lifschitz, D., Doshi-Velez, F. & Amir, O. (2019) Toward Robust Policy Summarization—Extended Abstract. *International conference on Autonomous Agents and Multi-Agent Systems (AAMAS) 2019*.
- [6] **Lage, I.**, Ross, A.S., Kim, B., Gershman, S.J. & Doshi-Velez, F. (2018) Human-in-the-Loop Interpretability Prior. *Advances in Neural Information Processing Systems (NeurIPS) 2018*. (**Spotlight talk—3.5% of submitted papers**)

REFEREED
WORKSHOP
PUBLICATIONS

- [1] Mahinpei, A.*, Clark, J.*, **Lage, I.**, Doshi-Velez, F., Pan, W., (2021) Promises and Pitfalls of Black-Box Concept Learning Models. *ICML: Workshop on Theoretic Foundation, Criticism, and Application Trend of Explainable AI*. (**Mentoring role**)
- [2] Tabac A., Wigell, R., Wolf, K., **Lage, I.**, Landrum, S., Reyes Nieva, H., Bearnot B., Streed, C. (2020) Using Patterns of Missing EHR Data to Identify Care Disparities in Gender Diverse Patients. Abstract at *American Public Health Association*.
- [3] **Lage, I.***, Lifschitz, D.*, Doshi-Velez, F. & Amir, O. (2019) Exploring Computational User Models for Agent Policy Summarization. *IJCAI 2019 Workshop on Explainable Artificial Intelligence (XAI)*.

- [4] **Lage, I.**, Doshi-Velez, F. (2020) Human-in-the-Loop Learning of Interpretable and Intuitive Representations. *ICML Workshop on Human Interpretability in Machine Learning*.
- [5] **Lage, I.**, Chen, E., He, J., Narayanan, M., Kim, B., Gershman, S. J. & Doshi-Velez, F. (2018) An Evaluation of the Human-Interpretability of Explanation. *NeurIPS 2018 Workshop: Critiquing and Correcting Trends in Machine Learning*.
- [6] Ross, A. S., **Lage, I.**, Doshi-Velez, F. (2017) The Neural Lasso: Local Linear Sparsity for Interpretable Explanations. *NeurIPS 2017 Workshop: Transparent and interpretable machine learning in safety critical environments*.

MANUSCRIPTS IN PREPARATION

- [1] Narayanan, S., **Lage, I.** & Doshi-Velez, F. (2022) (When) Are Contrastive Explanations of Reinforcement Learning Policies Useful?. *Under preparation for Interactive User Interfaces 2022, and under submission to NeurIPS 2022 HILL workshop. (Mentoring role)*
- [2] **Lage, I.**, Parbhoo, S. & Doshi-Velez, F. (2022) Leveraging Human Features at Test Time. *Under submission to International Conference on Learning Representations 2022, and NeurIPS 2022 HILL workshop*
- [3] **Lage, I.**, Parbhoo, S. & Doshi-Velez, F. (2022) Feature-Level Synthesis of Human and ML Insights. *Under submission to NeurIPS 2022 HCAI workshop*
- [4] **Lage, I.**, Horvitz, E., Nushi, B. & Kamar, E. (2022) Investigations of Instance-Sensitive Explanations.
- [5] **Lage, I.** & Doshi-Velez, F. (2020) Learning interpretable concept-based models with human feedback. *ArXiv*.
- [6] McGrath, S.*, Mehta, P.*, Zyteck, A., **Lage, I.** & Lakkaraju, H. (2020) When does uncertainty matter?: Understanding the impact of predictive uncertainty in ML assisted decision making. *ArXiv. (Mentoring role)*

OUTREACH AND SERVICE

- Harvard University**, Cambridge, MA
- Panelist in Seminar on Effective Research Practices & Academic Culture* 2022
- Participated as a panelist in panels on time management and applying for fellowships
 - Course is geared towards all first year PhD students in the department
- Try AI Research Internship Mentor* Fall 2021
- Mentored an early career college student in a short AI research project
 - Designed research project to build AI and research skills, and explore interests
 - Program was geared towards students from underrepresented minorities in STEM
- Mentor for Harvard IACS Ph.D. student working group* Fall 2021
- Conducted mock graduate school interviews
 - Read and gave feedback on application statements materials
 - Program was geared towards students from underrepresented backgrounds in STEM
- Mentor for Women in Machine Data Science, Cambridge workshop* Spring 2021
- Mentored teams of participants in hands-on data exploration methods
 - Gave personalized career advice to interested participants
 - Program was geared towards women and non-binary people exploring data science careers
- Tutor through Harvard initiative for local high school students* Spring 2021
- Tutored a high school student in history
 - Program was geared towards lower income students

Diversity Inclusion and Belonging Committee 2017-2019

- Worked in a small team to design and analyze a school-wide climate survey
- Results informed future DIB initiatives for the school of engineering

INVITED TALKS

- [1] “Exploring Computational User Models for Agent Policy Summarization” (Spring 2020) Science of Intelligence working group. Harvard university.
- [2] “Teaching AI To Think Like People, and People To Think Like AI” (Spring 2020) Joint with Felix Sosa. Science in the News Lecture Series. Harvard University. (Talk was canceled after substantial preparation due to COVID-19 pandemic.)

CONFERENCE TALKS

Joint first-authorship indicated with *
Presenting author indicated with ★

- [1] **Lage, I.*★**, Chen, E.*, He, J.*, Narayanan, M.*, Kim, B., Gershman, S. & Doshi-Velez, F. (2019) Human Evaluation of Models Built for Interpretability. *AAAI Conference on Human Computation and Crowdsourcing (HCOMP) 2019*. Washington, USA. 28-30 October 2019.
- [2] **Lage, I.*★**, Lifschitz, D.*, Doshi-Velez, F. & Amir, O. (2019) Exploring Computational User Models for Agent Policy Summarization. *International Joint Conference on Artificial Intelligence (IJCAI) 2019*. Macau, China. 10-16 August 2019.
- [3] **Lage, I.*★**, Ross, A.S., Kim, B., Gershman, S.J. & Doshi-Velez, F. (2018) Human-in-the-Loop Interpretability Prior. *Advances in Neural Information Processing Systems (NeurIPS) 2018*. Montreal, Canada. 3-8 December 2018.

WORKSHOP TALKS

- [1] **Lage, I.*★**, Doshi-Velez, F. (2020) Human-in-the-Loop Learning of Interpretable and Intuitive Representations. *ICML Workshop on Human Interpretability in Machine Learning*.
- [2] **Lage, I.*★**, Lifschitz, D., Doshi-Velez, F. & Amir, O. (2019) Toward Robust Policy Summarization—Extended Abstract. *AAMAS 2019 Workshop: EXTRAAMAS - International Workshop on Explainable Transparent Autonomous Agents and Multi-Agent Systems*.
- [3] **Lage, I.*★**, Sontag, D., Turning Insurance Claims Data into Insights for Multiple Myeloma. *2016 Moore-Sloane Data Science Summit*. New Paltz, New York. 23-26 October 2016.

OTHER PROFESSIONAL ACTIVITIES

Reviewer

- Conferences: AISTATS 2018, NeurIPS 2020
- Journals: AIJ in 2020, 2022
- Workshops: AAMAS Workshop on EXplainable and TRANSPARENT AI and Multi-Agent Systems 2020-2022, ICLR Workshop on Debugging Machine Learning Models 2019, IJCAI Workshop on Explainable Artificial Intelligence 2019-2022, ICML workshop on Algorithmic Recourse 2021, ICML Workshop on Human Interpretability in Machine Learning (WHI) 2020, ICAPS Workshop on eXplainable AI Planning 2021.

Participant in Queer Health Hackathon

Fall 2018

- Worked in a team to analyze EHR data of queer and trans patients
- Analysis revealed additional missing data for transgender compared to cisgender patients
- Team was runner up for first prize

Volunteer for AFS Intercultural Programs

2010-2014

- Conducted workshops on cross cultural communication for high school foreign exchange students