Isaac Lage

Curriculum vitae

Contact Information

Colby College
Computer Science
Waterville, ME, 04901

ilage@colby.edu
isaaclage.github.io

Education

Harvard University, Cambridge, MA — 2017-2023
School of Engineering and Applied Sciences
Doctor of Philosophy, Computer Science

New York University, New York, NY — 2012-16
Bachelor of Arts, Computer Science, Social and Cultural Analysis
- *Summa cum Laude*; with High Honors in Computer Science; Phi Beta Kappa

Selected Honors

- NSF Graduate Research Fellowship Recipient, Harvard University — 2019-2022
- NIH BD2K Training Grant Recipient, Harvard University — 2017-2019
- Computer Science Prize for Academic Excellence, New York University — 2016

Professional Experience

Colby College, Waterville, ME — 2023-Present
Assistant Professor of Computer Science

Microsoft Research, Redmond, WA — Summer 2020
Research Intern
Under supervision of **Dr. Ece Kamar**

Massachusetts Institute of Technology & New York University, Cambridge, MA & New York, NY — 2016-2017
Jr. Research Assistant & Jr. Research Scientist/Research Software Engineer
Under supervision of **Dr. David Sontag**

Teaching Experience

Colby College, Waterville, ME
Assistant Professor of Computer Science — Fall 2023
- Lecture and lab instructor for 1 section of CS 152–Computational Thnking: Science
- Lab instructor for 2 sections of CS 231–Data Structures and Algorithms

Tufts University, Medford, MA
Part-Time Lecturer — Spring 2022
- Instructor for Statistical Pattern Recognition, a core offering with $\sim$ 30 students.

Harvard University, Cambridge, MA
Pedagogy Fellow for the School of Engineering and Applied Sciences — 2022-2023
- Training 35-40 new TAs a semester
- Co-leading departmental pedagogy course (in spring 2023)
- Consulting with TAs about effective teaching strategies
- Organized a panel and orientation session on teaching for all 2nd year PhD students

Guest lecturer, Critical Thinking in Data Science — Spring 2021, 2022
- Invited to co-lecture on methods and ethics for human-in-the-loop learning

Guest lecturer, Interpretability and Explainability in ML — Spring 2021
- Invited to give 2 guest lectures on human evaluation of interpretability

Teaching Fellow, Interpretability and Explainability in ML — Fall 2019
- Holding office hours to support student research projects
- Co-designing portions of assignments
- Teaching a lecture
- Grading assignments

Teaching Fellow, Advanced Machine Learning Fall 2018

- Holding office hours
- Co-teaching a lecture
- Grading assignments

Bok Teaching Certificate 2017-2023

- Coursework in public speaking, curriculum design, and STEM pedagogy
- Guided reflection through classroom observation
- Reflective writing assignments on teaching values

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] Narayanan, S., **Lage, I.** & Doshi-Velez, F. (2022) (When) Are Contrastive Explanations of Reinforcement Learning Policies Useful?. *NeurIPS Workshop on Human in the Loop Learning.* (**Mentoring role**)

[2] **Lage, I.**, Parbhoo, S. & Doshi-Velez, F. (2022) Leveraging Human Features at Test Time. *NeurIPS Workshop on Human in the Loop Learning.*

[3] **Lage, I.**, Parbhoo, S. & Doshi-Velez, F. (2022) Feature-Level Synthesis of Human and ML Insights. *NeurIPS Workshop on Human Centered AI.*

[4] 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**)

[5] 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.*

[6] **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).*

[7] **Lage, I.**, Doshi-Velez, F. (2020) Human-in-the-Loop Learning of Interpretable and Intuitive Representations. *ICML Workshop on Human Interpretability in Machine Learning.*

[8] **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.*

[9] 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 and Pre-prints

[1] Narayanan, S., **Lage, I.** & Doshi-Velez, F. (2022) (When) Are Contrastive Explanations of Reinforcement Learning Policies Useful?. *ArXiv.* (**Mentoring role**)

[2] **Lage, I.**, Parbhoo, S. & Doshi-Velez, F. (2022) Incorporating Human Features into Machine Learning Predictions at Test-Time. *Under preparation for resubmission*

[3] **Lage, I.** & Doshi-Velez, F. (2020) Learning interpretable concept-based models with human feedback. *ArXiv.*

[4] McGrath, S.*, Mehta, P.*, Zytek, A., **Lage, I.** & Lakkaraju, H. (2020) When does uncertainty matter?: Understanding the impact of predictive uncertainty in ML assisted decision making. *ArXiv.* (**Mentoring role**)

Professional Mentoring and Service

Colby College, Waterville, ME

Computer Science Faculty Search Committee Member Fall 2023
- Read application materials and participated in interviews

Co-Planned Computer Science Department Social Event Fall 2023
- Planned food and activities with 1 other colleague for an event with around 40 student attendees

Jan-Plan Internship Faculty Sponsor Fall 2023
- Faculty sponsor for 1 student for a Jan-Plan internship

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 I designed to build AI and research skills, and explore research interests
- Program was geared towards students from underrepresented groups in STEM

Mentor for Harvard IACS Ph.D. student working group Fall 2021
- Conducted mock interviews and read application statements for students applying to Ph.D. programs
- Program was geared towards students from underrepresented groups in STEM

Mentor for Women in Machine Data Science, Cambridge workshop Spring 2021
- Mentored teams of participants in a hands-on data-science workshop
- 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 with history
- Program was geared towards lower income students

Diversity Inclusion and Belonging Committee 2017-2019
- Worked in a small team to design and analyze of a school-wide climate survey
- Results informed future DIB initatives 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 Agen 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.