Christine Herlihy

: crherlihy.github.io
: christine.herlihy@gmail.com
: christineherlihy
: github.com/crherlihy
: crherlihy

Education

University of Maryland, College Park — College Park, MD
Ph.D. in Computer Science — 2019–present

- *Coursework*: Algorithms in ML: Guarantees and Analyses, Applied Mechanism Design, Computational Linguistics I & II, Common-sense Reasoning and Natural Language Understanding, Program Analysis
- *Advised by*: Dr. John Dickerson | *Research interests*: sequential and combinatorial decision-making under uncertainty; algorithmic fairness; knowledge representation and reasoning in the clinical domain.

Georgia Institute of Technology — Atlanta, GA
M.S. in Computer Science | Concentration: Machine Learning — 2017–2018

- *Coursework*: Machine Learning, Artificial Intelligence, Reinforcement Learning, ML for Trading, HPC, Big Data Analytics for Healthcare, Data Analytics using Deep Learning, Info. Security, Computability and Algorithms

Georgetown University — Washington, DC
B.A. in Government; minor in Spanish — 2007–2011

- Received academic honors during every semester of attendance; studied abroad in Santiago, Chile.

Experience

Amazon Robotics — Remote
Summer Research Scientist — 05/2021 - present

- Graph-based deep reinforcement learning algorithm development.

Georgia Tech Research Institute — Atlanta, GA
Research Scientist II, High-Performance Computing & Data Analytics Branch — 01/2017 - 08/2019

- Technical task lead and core contributor for a range of research projects and proposals, including: patient-level predictive modeling; computational phenotyping; application of unsupervised learning and NLP techniques on unstructured text to develop machine phenotypes and detect spatially/temporally co-occurring machine failures; development of models to predict geopolitical conflict and detect fake news.

Econometrica, Inc. — Bethesda, MD
Research Associate II, Health Data Analytics — 08/2015 - 01/2017

- Patient and population-level predictive modeling; healthcare policy evaluation using Python, R, and Stata.
- Primary project was an impact evaluation contract that involved using various econometric methodologies, a genetic matching algorithm, and survey design schemes to detect and determine changes in quality of care, medical outcomes, Medicare costs, and unintended consequences associated with emerging bundled payment and gainsharing mechanisms in health care systems.

Frontier Strategy Group — Washington, DC
Macroeconomic Research Analyst, Latin America — 08/2012 - 08/2014

- Conducted econometric research and built forecast models using Python and R to advise C-suite executives at over 200 multinationals on resource allocation and risk management in LATAM.

Technical Skills

- **Proficient:** Python, R, Julia, SQL, LaTeX | **Familiar:** Java, C++, Scala, Spark, bash, Coq, Unity, MPI
- **Libraries & tools:** scikit-learn, pandas, numpy, OpenAI Gym, PyTorch, TensorFlow, spaCy/scispacy, nltk, gensim, textacy, AllenNLP, NetworkX, Neo4j, git, Postgres, Categorical Query Language (CQL), FHIR, OMOP, Tableau
- **Languages:** English (Native), Spanish (Fluent), Portuguese (Intermediate), Farsi (Beginner)

Preprints

[1] **C. Herlihy**, A. Prins, A. Srinivasan, and J. Dickerson, *Planning to Fairly Allocate: Probabilistic Fairness in the Restless Bandit Setting*, Under review, 2021. arXiv: 2106.07677 [cs.LG].

[2] **C. Herlihy** and R. Rudinger, *MedNLI Is Not Immune: Natural Language Inference Artifacts in the Clinical Domain*, Forthcoming at ACL-IJCNLP, 2021. arXiv: 2106.01491 [cs.CL].

Conference Papers

[1] M. Halter, **C. Herlihy**, and J. P. Fairbanks, "A Compositional Framework for Scientific Model Augmentation", in *Proceedings Applied Category Theory 2019, ACT 2019, University of Oxford, UK, 15-19 July 2019*, J. Baez and B. Coecke, Eds., ser. EPTCS, vol. 323, 2019, pp. 172–182. [Online]. Available: https://doi.org/10.4204/EPTCS.323.12.

Publications

[1] A. Moreland, **C. Herlihy**, M. Tynan, G. Sunshine, R. McCord, C. Hilton, J. Poovey, A. Werner, C. Jones, E. Fulmer, A. Gundlapalli, H. Strosnider, A. Potvien, M. García, S. Honeycutt, and G. Baldwin, "Timing of State and Territorial COVID-19 Stay-at-Home Orders and Changes in Population Movement - United States, March 1-May 31, 2020", *Morbidity and Mortality Weekly Report*, vol. 69, no. 35, Sep. 2020.

[2] D. Ruiz, D. Stout, and **C. Herlihy**, "Use of Genetic Matching in Program Evaluation: The Case of RAD", *Cityscape*, vol. 19, no. 2, pp. 337–350, 2017, ISSN: 1936007X. [Online]. Available: http://www.jstor.org/stable/26328344.

Select Projects

- Resource-Constrained Cohort Selection Using Multi-Arm Bandits | UMD (Python & Julia, 2020 - Present)
  Developing algorithms to incorporate equity constraints in the RMAB setting, and cope with vanishing arms in tiered MABs.
- A Category-Theoretic Approach to Computational Phenotyping | UMD & NIST (Python & Julia, 2019 - Present)
  Developing NLP and category-theoretic techniques to parse, represent, and compose clinical trial eligibility criteria.
- Task-Informed Document Retrieval on the CORD-19 Dataset Using Topic Modeling | UMD (Python, 2020)
  NMF and Jensen-Shannon divergence-based approach to unsupervised, COVID-19 task-informed document retrieval.
- A Compositional Approach to Representing and Manipulating Scientific Models | GTRI (Julia, 2018-2019)
  Developed program analysis and category-theoretic techniques for epi model extraction and composition from text and code.
- ClarityNLP: An open-source NLP framework for clinical phenotyping | GTRI & Celgene (Python, 2018-2019)
  Developed supervised and unsupervised methods for clinical text mining, feature engineering, and cohort identification.
- NLP Pipeline to Detect Cascading and Co-occurring Machine Failures | GTRI (Python & Neo4j, 2018-2019)
  Statistical analysis and topic modeling of work order notes; constructed KG to identify spatial/temporal failure patterns.

- Verifiable, Patient-Directed Provider-to-Provider EHR Record Sharing | GTRI (Python & Ethereum, 2018)
  Developed escrow smart contract prototype to facilitate patient-directed record sharing among providers and stakeholders.
- Chest X-ray Disease Diagnosis with Deep Convolutional Neural Networks | Georgia Tech (PyTorch, 2018)
  Used CNNs to detect and localize the 14 thoracic pathologies present in the NIH Chest X-ray dataset.
- Predicting Sepsis Onset in ICU Patients | GTRI (Python & R, 2017)
  Worked with clinical team to develop patient-level predictive models for sepsis and population-level triage models.

Teaching

- **Co-Instructor** at Georgia Tech Professional Education — Fall 2018
  Machine Learning Short Course
- **Teaching Assistant** at Johns Hopkins University — Spring 2017
  Algorithms and Data Structures
- **Teaching Assistant** at University of Maryland, College Park — 2014 - 2015
  Political Theory; International Relations
- **Volunteer Teacher** at Latino Student Fund — 2012-2014
  English as a Second Language (ESL) for Adult Learners

Extracurricular Activities

- Iribe Initiative for Inclusion & Diversity in Computing: Student Advisory Board Member — 2019–Current
  Bias, Education, Incident Reporting, and Professional Development committee member

Scholarships and Awards

- **Professional Research Experience Program (PREP) Researcher** | NIST — 2019–2021
- **Dean's Fellowship (CS)** | University of Maryland, College Park — 2019–2020
- **HIVE $25K Research Grant Winner** | Georgia Tech Research Institute — 2017
- **FIA-Deutsch $25K Seed Grant Fellow** | University of Maryland, College Park — 2014–2015
- **Dean's Fellowship (GVPT)** | University of Maryland, College Park — 2014–2015
- **Merit-based scholarship recipient; academic honors** | Georgetown University — 2007–2011
- **Gilman International Scholarship Recipient** | United States Department of State — 2010