LESIA SEMENOVA

lesia.semenova@duke.edu | https://users.cs.duke.edu/~ls305

My ***theoretical*** work creates a foundation for the existence of accurate simpler, including interpretable, machine learning models. I introduced a new simplicity measure of a learning problem and proposed first methods to measure the Rashomon set (the set of equally well-performing models), which enabled the shift in the machine learning community towards model multiplicity and underspecification. My ***applied*** work with immunologists has led to understanding of how cannabis affects the immune system of people with HIV. Student teams that I've coached have won the ASA Data Challenge Expo (twice) and placed third in a competition on a scholarly document processing. My work has over 700 citations.

RESEARCH INTERESTS

Responsible and Trustworthy AI, Machine Learning, Human-Centered Design, Reinforcement Learning, Reasoning, AI in Healthcare

EDUCATION

Duke University, Durham, NC, USA
Ph.D. in Computer Science — 2016 – 2024 (expected)
Advisors: Cynthia Rudin, Ronald Parr

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
MS in Applied Mathematics — 2012 – 2014
BS in Applied Mathematics — 2008 – 2012

PUBLICATIONS
Google Scholar
(* denotes equal contributions)

1 **Lesia Semenova**, Harry Chen, Ronald Parr, Cynthia Rudin. **A path to simpler models starts with noise.** *Neural Information Processing Systems (NeurIPS), 2023.*

2 Siong Thye Goh*, **Lesia Semenova***, Cynthia Rudin. **Sparse density trees and lists: an interpretable alternative to high-dimensional histograms**. *INFORMS Journal on Data Science (IJDS), 2024.*

3 Shane D Falcinelli, Alicia Volkheimer, **Lesia Semenova**, Ethan Wu, Alexander Richardson, Manickam Ashokkumar, David M Margolis, Nancie M Archin, Cynthia D Rudin, David Murdoch, Edward P Browne. **Impact of cannabis use on immune cell populations and the viral reservoir in people with HIV on suppressive antiretroviral therapy.** *The Journal of Infectious Disease (JID), 2023.*

4 **Lesia Semenova**, Cynthia Rudin, Ronald Parr. **On the existence of simpler machine learning models**. *ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2022.*

5 Cynthia Rudin, Chaofan Chen*, Zhi Chen*, Haiyang Huang*, **Lesia Semenova***, Chudi Zhong*. **Interpretable machine learning: fundamental principles and 10 grand challenges**. *Statistics Surveys, 2022.*

6 Dennis Tang, Frank Willard, Ronan Tegerdine, Luke Triplett, Jon Donnelly, Luke Moffett, **Lesia Semenova**, Alina Jade Barnett, Jin Jing, Cynthia Rudin, Brandon Westover. **ProtoEEGNet: An interpretable approach for detecting interictal epileptiform discharges**. *Medical Imaging meets NeurIPS workshop, 2023 (oral).*

7 Gaurav Rajesh Parikh, Jenny Huang, Albert Sun, **Lesia Semenova**, Cynthia Rudin. **Moving towards a more equal world, one ride at a time: studying public transportation initiatives using interpretable causal inference.** *NeurIPS Workshop on Causality for Real-world Impact, 2022.*
— Won 2022 American Statistical Association Data Challenge Expo Student Competition

8 Alex Oesterling, Angikar Ghosal, Haoyang Yu, Rui Xin, Yasa Baig, **Lesia Semenova**, Cynthia Rudin. **Multitask learning for citation purpose classification**. *Second Workshop on Scholarly Document Processing, NAACL, 2021.*
— Won third place in 3C Shared Task Competition

9 **Lesia Semenova**, Yingfan Wang, Shane Falcinelli, Nancie Archin, Alicia D Cooper-Volkheimer, David M Margolis, Nilu Goonetilleke, David M Murdoch, Cynthia D Rudin, Edward P Browne. **Machine learning approaches identify immunologic signatures of total and intact HIV DNA during long-term antiretroviral therapy.** *Accepted to eLIFE.*

PREPRINTS

10 Chloe Qinyu Zhu, Muhang Tian, **Lesia Semenova**, Jiachang Liu, Jack Xu, Joseph Scarpa, Cynthia Rudin. **Fast and interpretable mortality risk scores for critical care patients**. *Submitted to Journal of the American Medical Informatics Association (JAMIA).*

11 Allan Guo, Eric Song, Gaurav Rajesh Parikh, Harry Chen, **Lesia Semenova**, Cynthia Rudin. **Weed and violence: The impact of marijuana legalization on crime in California.** *To be submitted to Harvard Data Science Review (HDSR).*

— Won 2023 American Statistical Association Data Challenge Expo Student Competition

INVITED TALKS

Conference on Information Sciences and Systems (CISS) 2023
JSM, Near-Optimization Topic-Contributed session 2021
INFORMS Annual Meeting 2020

CO-ADVISING AND MENTORING

Harry Chen, undergraduate, Duke University
Flora Shi, undergraduate, Duke University (now PhD student at MIT)

Co-instructor for Duke Data Science teams that participated in various Data Science competitions (9 teams total, 29 Duke undergraduate students)

TEACHING

Certificate in College Teaching, Duke Graduate School
Formal pedagogical training in the college teaching.

Institute of Advanced Study, Teaching Assistant
2022 Program for women and mathematics: "The Mathematics of Machine Learning"
Terng Lecture Course on Interpretable Machine Learning May 2022

Duke University, Teaching Assistant
CS474, Data Science Competition SP23, SP22, SP21
CS571, Probabilistic Machine Learning SP18
— TA Award Honorable Mention
CS101, Introduction to Computer Science SP17

SELECTED AWARDS AND HONORS

Rising Stars in Computational and Data Science, University of Texas at Austin 2024
SAMSI Fellowship 2018
Duke CS Department Fellowship 2016 – 2018
Scholarship of Mayor of Kyiv 2012 – 2014
Scholarship of Academic Council of Taras Shevchenko National University 2011 – 2012
Scholarship of President of Ukraine 2008 – 2009

SELECTED PROFFESSIONAL EXPERIENCE

Research Intern, Pinterest Labs, Palo Alto, CA, USA Summer 2021
Applied Science Team
Spammy User Detection. Proposed interpretable and explainable models for the spammy user detection, which enabled team to provide reasons for the account freeze to the customers.

Research Intern, Pinterest Labs, Palo Alto, CA, USA Summer 2020
Applied Science Team
User modeling. Added long-term user activity to Pinterest's PinnerSage user embedding framework. Experimented with different ML models (linear, XGBoost, NN, LSTM) and after extensive evaluation concluded that simple models perform similarly to black boxes.

Software Engineer, Samsung R&D Institute, Kyiv, Ukraine 2012 –2014
Interaction Lab, Augmented Reality Team
Developed a representative image selection algorithm based on similarities for image clustering. As a part of the team, developed a barcode engine for QR-code and barcode encoding, decoding, and tracking.

SKILLS

Python, C++, Keras, TensorFlow, PyTorch, Scikit-learn, OpenCV, MATLAB, SQL, Hive, Hadoop, LaTeX, Git

SERVICE

Reviewer: JMLR, Patterns, DeepMath 2022, 2023, ACM Journal on Responsible Computing

PC member: International Workshop on Advances in Interpretable Machine Learning and Artificial Intelligence — 2021, 2022, 2023

Volunteer procurement manager at "Razom for Ukraine" nonprofit — 2022-2023

Co-lead the "Artificial Intelligence for Art and Fun" capstone event as a part of Duke's FEMMES+ (Females and Allies Excelling More in Math, Engineering, and Science) outreach program to introduce young female students (4th-6th grade) to math, science, and engineering — 2021

Student Assistant at NSF-sponsored Seamless/Seamful Human Technology Interaction (HTI) Workshop — May 2021

Co-organized the discussion series "Controversial Topics in Precision Medicine and Learning" as a collaboration between SAMSI and Duke Computer Science — 2019

Graduate student committee member for the Faculty Search and Prospective Student Visit for the Department of Computer Science, Duke University — 2016–2019

Conference volunteer — KDD 2017, ICML 2019

Alumni of ComSciCon Triangle (science communication workshop for graduate students) — 2018

Volunteer tutor at Study Zone at King County Library System (provided homework help to K-12 students) — 2015

Head of Scientific Association of Students and Postgraduates of Taras Shevchenko National University of Kyiv — 2012 – 2013