Shannon Gallagher

 $\cdot\,$ Machine Learning $\,\cdot\,$ Data Science $\,\cdot\,$ Statistics $\,\cdot\,$ Ph.I

SEI Building, 4500 Fifth Ave., Pittsburgh, PA 15213

🛿 (724) 504-8990 | 🛛 skgallagher19@gmail.com, skgallagher@sei.cmu.edu | 🖀 skgallagher.github.io

Current

Carnegie Mellon University

AI Division | Software Engineering Institute Team Lead and Machine Learning Research Scientist

Education

Carnegie Mellon UniversityPittsburgh, PAPH.D. IN STATISTICS2014-2019M.S. IN STATISTICS2014-2015B.S. IN MATHEMATICAL SCIENCES (UNIVERSITY AND COLLEGE HONORS)2010-2014Doctoral Dissertation: "Catalyst: agents of change. Integration of compartment and agent-based models for use in infectious diseasemethodology." Advisor: William F. Eddy

In Preparation

Rallapalli, S., Mellinger A.O., **Gallagher, S.K.**, [and others.] "Quantifying the Efficacy of Fine-tuning Large Language Models."

Grimes K., Abidi C., Frank C., and Gallagher, S.K. "Improved Privacy Auditing for Machine Unlearning."

Turri, V., Robinson K-M., Gallagher, S.K., and Smith, C.J. "Persona: Analyzing Biases in Large Language Models."

Selected Publications and Reports

Gallagher, S.K., Ratchford J., [and others including Heim E., McMillan S., Rallapalli S., Smith C.J., and VanHoudnos, N.] "Assessing LLMs for High Stakes Applications." Accepted to International Conference in Software Engineering SEIP 2024.

Gallagher, S.K., Mellinger, A., Ratchford, J. and 9 others. "A Retrospective in Engineering Large Language Models for National Security." Software Engineering Institute, Digital Library, 2023.

Gallagher, S.K., Whisnant A., Hristozov, Anton D., and Vasudevan, Amit. "Reviewing the role of machine learning and artificial intelligence for remote attestation in 5G+ networks." IEEE Future Networks World Forum, 2022.

Gallagher, S.K., Klieber, W. E., and Svoboda, David. "LLVM intermediate representation for code weakness identification." Available at Defense Technical Information Center, 2022.

Gallagher, S.K., Wang, J., Lombard, K., Dodd, L., and Proschan, M. "Using Kaplan Meier estimates in non-inferiority hypothesis tests." *Statistics in Medicine*, 2022.

Gallagher, S.K. and Follmann, D. "Branching process models to identify risk factors for infectious disease transmission." *Journal of Computational and Graphical Statistics*, 2022.

Bohl, J. [and others including Fintzi, J., **Gallagher, S.K.**, and Manning, J.] "Discovering disease-causing pathogens in resource-scarce Southeast Asia using a global metagenomic pathogen monitoring system." *PNAS*, 2022.

Paules, C.I.[†], **Gallagher, S.K.**[†], Rapaka, R.R., [and others including Dodd, L.E., and Benson, C.A.] ([†] co-first author). *Clinical Infectious Diseases*, 2021.

Gallagher, S.K., Frisoli K., and Luby, A. "Opening up the court (surface) in tennis grand slams." *Journal of Quantitative Analysis in Sports*, 2021.

Azasi, Y.[†], **Gallagher, S.K.**[†], [and 11 others including Fay, Michael P., Miura, K., and Miller, Louis H.] († co-first author). "Bliss' and Loewe's additive and synergistic effects in Plasmodium falciparum growth inhibition by AMA1-RON2L, RH5, RIPR and CyRPA antibody combinations." *Scientific Reports*, 2020.

Pittsburgh, PA 2022-Present **Gallagher, S.K.** "Discussion of 'An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China.' "*Journal of Data Science*, 2020.

Gallagher, S.K., Chang, A., Eddy, W.F. "Exploring the nuances of R_0 : Eight estimates and application to 2009 pandemic influenza." Pre-print available at: https://arxiv.org/abs/2003.10442,2020.

Gallagher, S.K. and Eddy, W.F. "A hybrid compartment/agent-based model for infectious disease modeling." ENAR Epidemiology section student paper contest submission, 2019.

Gallagher, S.K., Richardson L.F., Ventura S.L., and Eddy, W.F. "SPEW: Synthetic Populations and Ecosystems of the World." *Journal of Computational and Graphical Statistics*, 2018.

Selected Presentations, Panels, Posters, and Other Media

Software Engineering Institute	YouTube
Podcast	January 2024
"Using Large Language Models in the National Security Realm"	
Al for Defense	National Harbor, MD
Panelist	September 2023
"Applications of Generative AI Tools for Civilian and Defense Operations"	
Pittsburgh Chapter of SWE Professional Development Day	Pittsburgh, PA
Invited Presentation	April 2023
"The Near Future of AI: Transformer-Based Learning"	
Software Engineering Institute	YouTube
WEBCAST	August 2022
"What are Deepfakes, and How Can We Detect Them?"	
Joint Statistical Meetings	Washington, DC
Speed Presentation	August 2022
"A Statistical Framework for Deepfake Detection."	
Software Engineering Institute	YouTube
Podcast	July 2022
"A Dive into Deepfakes"	
Military Operations Research Society 90th Symposium	Quantico, VA
TUTORIAL	June 2022
"Intro to Data Science for Digital Forensics."	
ISACA Huntsville Chapter Meeting	Virtual Meeting
Invited Presentation	May 2022
"Deepfake Generation & Detection: How easy are they to make and detect?"	
Global Platforms Austin Workshop: SESIP's Role in the IoT Ecosystem.	Virtual Conference.
Presentation	April 2022
"Trust Assessment Models for Next Generation Networks."	
5th Annual DoD Artificial Intelligence/Machine Learning TEM	Virtual Conference
Presentation	October 2021
"Using Transformer Machine Learning to Identify Code Weaknesses."	
George Washington University	Washington D.C.
GUEST LECTURE	February 2020 & 2021
"A brief survey of statistical models to analyze the transmission of infectious diseases."	
Los Alamos National Laboratory	Virtual Seminar
Invited Presentation	August 2020
"Use of multiple covariates in branching processes for modeling the spread of Tuberculosis."	
Carnegie Mellon University	Virtual Class
GUEST LECTURE	September 2020
"A foray into infectious disease modeling: compartment models."	
Joint Statistical Meetings	Virtual Conference
Poster/Presentation	August 2020
"Branching processes with covariates to model the spread of Tuberculosis."	

ENAR	Virtual Conference
Presentation	March 2020
"A Hybrid Compartment/Agent-Based Model for Infectious Disease Modeling."	
Dissertation Defense	Pittsburgh, PA
Presentation	July 2019
"Catalyst: agents of change. Integration of compartment and agent-based models for use in infectious disease epie	demiology."
Carnegie Mellon Sports Analytics Conference	Pittsburgh, PA
Presentation – Honorable Mention	October 2018
"Opening up the (court) surface in tennis grand slams." Joint work with Kayla Frisoli and Amanda Luby.	
International Conference on Synthetic Populations	Lucca, Italy
Presentation – Invited Speaker	February 2017
"Generating Synthetic Ecosystems: A Tutorial" Joint work with Lee Richardson, Samuel Ventura, and William Eddy.	
MIDAS National Conference	Washington D.C.
Presentation	May 2016
"Services for the MIDAS Network: Visualization and Synthetic Ecosystems." Joint work with Lee Richardson, Samue	el Ventura, and William Eddy.
UP-STAT	Buffalo, NY
Presentation – 2nd place	March 2016

PRESENTATION - 2ND PLACE

"From forecasting the Flu to Predicting the 'Next' Disease." Joint work with Roni Rosenfeld, Ryan Tibshirani, Lee Richardson, Samuel Ventura, and William Eddy.

Honors & Awards

2023	Nomination, Leading and Advancing as a Member of the 5G Team	Pittsburgh, PA
2022	Nomination, Newcomer of the Year at SEI	Pittsburgh, PA
2022	Certificate, Awarded CERT Spotlight for Outstanding Achievement on a project	Pittsburgh, PA
2020 F	Selected Carnegie Mellon University Student Speaker for 2020 Commencement, Conferral of Degrees.	Virtual
	Full program available at: https://www.cmu.edu/commencement/.	VITLUAI
2018	Honorable Mention, Carnegie Mellon University Sports Analytics Conference Reproducible Paper	Dittaburah DA
2010	Competition. \$1,000 award.	FILISDUIGII, FA
Honorable Mention, Gertrude M. Cox Scholarship. ASA Committee on Women in Statistics and the G		
2018, 2014	for Women in Statistics.	
2018	Scholarship Recipient, Summer Institute in Statistics and Modeling. Tuition and travel stipend.	Seattle, WA
2017	Selected Presenter, AT&T Labs Graduate Student Symposium. One of fourteen PhD students out of 79	New York, NY
	applicants selected. Awarded \$800 in travel funding.	
2016	Hackathon Champion, MIDAS MISSION Public Health Hackathon. Awarded \$3,000 prize.	Pittsburgh, PA
2016	2nd place, Student presentation at UP-STAT conference.	Buffalo, NY
2014	Judith A. Resnik Award for Outstanding Women in the Sciences, Carnegie Mellon University.	Pittsburgh, PA
2013	Phi Beta Kappa Honor Society, Fall induction.	Pittsburgh, PA

Software_____

2023-	Mayflower, Brooks, T., Mellinger, A.O., Winski N., [and others including Gallagher, S.K.] Internal SEI	
Present	repository for LLM assessment, experimentation, and validation.	
2022	Deepfake Detection Pipeline, Gallagher, S.K., Ross, D., Mellon, J., and Bernaciak, C. Internal SEI	
2022	repository.	
2010 2021	EpiCompare, Gallagher, S.K. and Leroy, B. Software for simulation and analysis of disease data via ternary	
2019-2021	plots. Available at https://skgallagher.github.io/EpiCompare/index.html.	
	InfectionTrees R package, Gallagher, S.K. and Follmann, D. Analysis, vignettes, and code for studying	
2021	transmisison trees and incorporating information from individuals. Available at	
	https://skgallagher.github.io/InfectionTrees/articles/getting-started.html.	
	loewesadditivity R Package , Gallagher, S.K. and Fay, M. P. Software for modelling synergy, antagonism,	
2019-2020 or Loewe additivity between varying dose combinations of different compounds. Available at		

www.github.com/skgallagher/loewesadditivity.

Resea	arch, Teaching, and Work Experience	
AI Divisio	on, Software Engineering Institute, Carnegie Mellon University arning Research Scientist	Pittsburgh, Pr 2022-Presen
 Instantia Develop Headed Coordina Explored Research Explored 	ated, fine tuned, orchestrated, and assessed LLMs for high stakes applications ed prototype experiment and UI for human assessment of LLM responses Mayflower Project for Office of Director of National Intelligence to research the use of LLMs for National Security ated and oversaw research for related projects including analyzing biases in LLMs and machine unlearning. d adversarial attacks and defenses on and for LLMs. hed penultimate embeddings of computer vision foundation models with respect to data drift d, modeled, and visualized time series categorical data for customer.	
AI Divisio	on, Software Engineering Institute, Carnegie Mellon University	Pittsburgh, PA
TEAM LEAD C	OF AI Engineering Machine Learning Scientists	2023-Presen
 Managed Organized Co-created mation 	d and led six machine learning research scientists in AI Engineering ed monthly internal seminar for AI division, "Timely Topics in AI" to discuss relevant topics red and co-led biweekly seminar for AI Engineering Research Scientists to discuss academic journal articles and oth	ner relevant infor
Data Scie	ence Team, CERT, Software Engineering Institute, Carnegie Mellon University	Pittsburgh, PA
DATA SCIENT	rist	2021-202.
Led deepExploredConsultePresente	pfake detection internal research project that implements a pipeline of reproducible statistical and ML models d machine learning models and existing data for code weakness identification ed Department of Defense (DoD) agencies in AI/ML use in emerging technology ed work at DoD conferences	
National	Institute of Allergy and Infectious Disease	Rockville, ME
Розт-Досто	DRAL FELLOW	2019-202
 Worked Devised Analyzed Helped I 	with Dean Follmann to analyze the effect of smear status on spread of Tuberculosis and implemented statistical model to analyze the synergy of antibody pair combinations for Malaria vaccine efforts d survival probability under different loss to follow up scenarios ead a retrospective analysis of the COVID-19 ACTT-1 Clinical Trial to identify subgroups that respond to remdesivir	
Carnegie	Mellon University	Pittsburah. PA
RESEARCH A	nd Teaching Assistant	2014-2019
 Develop Generate Oversaw Statistica 	ed and presented material for the Summer Undergraduate Research Experience in Statistics ed high-resolution synthetic ecosystem of the U.S. and 70+ countries for use in agent-based models for transmission / lab for 100 students, organized and led review sessions for a variety of statistics and mathematics classes includir al Computing, Intro to Probability, Advanced Undergraduate Research, Concepts of Mathematics, and Multi-dimensio	of disease ng Epidemiology onal Calculus
PNC		Pittsburgh, PA
GRADUATE IN	NTERN	2015
ScrapedParalleliz	and analyzed social media data for sentiment analysis zed code via Hadoop	
Profe	ssional Service	
2023-	Co-leader, SEI LLM Research Lunch. Organized monthly meeting across the three SEI divisions to coordinate	
Present	and share research for large language models	
2023- Present	Member, Association for Computing Machinery	
2017- Present	Member, American Statistical Association	

catalyst, Gallagher, S.K.. Software for simulation, testing, and analysis of compartment and agent-based

spew, Richardson L., Gallagher, S.K., Ventura, S., and Eddy, W.F. R package for synthetic ecosystem

spewview, Gallagher, S.K. and Richardson L. R. Shiny application for infectious disease visualization.

models. Available at www.github.com/skgallagher/catalyst.

generation. Available at www.github.com/lrichardson/spew.

Available at www.github.com/skgallagher/hackathon.

Reviewer, Statistics in Medicine, Clinical Infectious Diseases, Statistics and Public Policy, Journal of 2016-Present Quantitative Analysis in Sports, Journal of Data Science, Digital Threats: Research and Practice

2019

2018

2016

Mentor to students, National Insitute of Allergy and Infectious Diseases

- Statistical mentor for a student's research study to complete her Masters of Health Sciences in Clinical Re-2020-2021 search Training Program at Duke School of Medicine
 - Undergraduate student research co-advisor to a Carnegie Mellon University statistics student along with Professor Joel Greenhouse
- PI, ProSeed/Crosswalk recipient for \$1600 to seed a mentorship program across all levels of students within 2018-2019

the Stat&DS community

President, Carnegie Mellon University Women in Statistics.

- Organized Women in Data Science Pittsburgh @CMU as an Executive Committee Member. Inivted speakers
 and sponsors, helped organize venue logistics, sent out invitations for for attendance, and created the 2018
 website
- 2018-2019 Maintained the Women in Statistics Website from 2017-2018
 - Organized a seminar by former PhD student about her experiences as a post-doc at Harvard Biostatistics (2017)
 - Organized a panel about applying to graduate school for 30+ undergraduate and masters students (2016)
 - Organized dinner with new dean of Mellon College of Science (2016)
- **Co-Organizer**, Pittsburgh useR. Organized meet-ups for 30+ members on a variety of topics including 2016-2018

cross-language coding and integrating R with github

2016-2017 Judge and volunteer, Tartan Data Science Cup - three separate events.

- 2016-2017 Vice President, CMU Women in Statistics.
 - 2016 **Presenter**, Coding for Girls

Relevant Course Work

- Machine Learning I and II (Grad)
- Statistical Computing (Grad)
- Modern Regression (**Grad**)
- Hierarchical Models (Grad)

- Multivariate Methods and Data Mining
- Data Matching and Record Linkage
- Advanced Methods for Data Analysis
- Epidemiology