Workshop on Spurious Correlation and Shortcut Learning: Foundations and Solutions

Workshop Summary

Overview

Despite the remarkable advancements towards generalizability and autonomy in AI systems, persistent challenges such as spurious correlations and shortcut learning continue to hinder the robustness, reliability, and ethical deployment of machine learning systems [1, 2, 3, 4]. These challenges arise from the statistical nature of machine learning algorithms and their implicit or inductive biases at all stages, including data preprocessing, architectures, and optimization. As a result, models rely on spurious patterns rather than understanding underlying causal relationships, making them vulnerable to failure in real-world scenarios where data distributions involve under-represented groups or minority populations. The foundational nature and widespread occurrence of reliance on spurious correlations and shortcut learning make it an important research topic and a gateway to understanding how deep models learn patterns and the underlying mechanisms responsible for their effectiveness and generalization.

This workshop aims to foster a collaborative community to address these critical issues by bringing together experts from diverse fields and pushing the boundaries of current research. We will focus on promoting three key avenues: (i) the development of comprehensive evaluation benchmarks and the exploration of under-examined facets of the problem, (ii) the creation of novel solutions for building robust models that effectively tackle spurious correlations in real-world applications, and (iii) shedding light on lesser-explored aspects to deepen our understanding of the nature of these phenomena.

Objectives

Current benchmarks based on group labels offer limited guarantees of robustness, addressing only a few known spurious correlations. Additionally, human annotation of groups is not a scalable solution and may overlook spurious correlations that do not align with human perceptions. Current evaluations do not inform us about the scenarios when the spurious correlation is unknown or annotations are missing [5]. Thus, there is a notable lack of rigorous evaluation benchmarks for assessing robustness to spurious correlations. Developing comprehensive benchmarks and also automated methods for detecting spurious correlations could significantly advance progress in this field.

Moreover, many facets of developing robust models to combat spurious correlations remain inadequately explored. The investigation of spurious correlations in learning paradigms beyond supervised learning has been particularly limited. As foundation models continue to gain prominence, it becomes necessary to leverage these models not only as tools for tackling spurious correlation challenges [6] but also as subjects of study to better understand the spurious correlations they may manifest [7].

While the impacts of and solutions for robustness to spurious correlation and shortcut learning have been targeted more frequently, attention has recently shifted to their foundations. Recent works focus on the origins of reliance on spurious correlation and shortcut learning in DNNs. Factors such as the tendency to maximize margins [8], biases introduced during training with SGD [9][10], and the time difference in learning core versus spurious patterns [11][12] are a few works towards a fundamental understanding of this phenomenon in deep learning. However, lots of open questions regarding the mechanism behind learning biases in various paradigms of AI and in different architectures and algorithms remain open.

Overall, the topics of interest for the workshop include, but are not limited to, the following:

- Introducing new spurious correlation benchmarks for various fields and modalities, including multimodal data (image, text, audio, video, graph, time series, etc.)
    - Examining foundational large language models (LLMs) and large multimodal models (LMMs) in terms of robustness to spurious correlations
    - Creating new datasets to evaluate the robustness of multi-modal models
    - Developing new benchmarks focusing on different types of features (depending on their modality) as shortcuts
    - Constructing new robustness benchmarks for various applications (medical, social, industrial, geographical, etc.)
    - Designing new tasks and environments to study spurious correlations in reinforcement learning
    - Presenting new real-world scenarios and benchmarks that challenge reliance on spurious correlations and shortcut learning
- Proposing new robustification methods
    - Finding solutions for the efficient robustification of LLMs and LMMs
    - Introducing new robustification methods for various paradigms, such as reinforcement learning, contrastive learning, and self-supervised learning
    - Proposing new algorithms for causal representation learning
    - Investigating novel solutions for robustness to spurious correlations in less-explored areas, such as optimization algorithms and data gathering and preprocessing schemes

    - Finding solutions for robustness to spurious correlation when information regarding spurious feature is completely or partially unknown
    - Introducing methods for robustness to spurious correlations in specific applications (medical, social, industrial, geographical, etc.)
- Exploring the foundations of spurious correlations and shortcut learning
    - Presenting mathematical formulations that describe the issue and its origins
    - Studying the role of widely used gradient-descent-based optimization methods in reliance on shortcuts and improvement solutions
    - Exploring the effect of shortcuts and spurious features on the loss landscape

We're working to build a strong, collaborative community of researchers and experts to tackle these challenges together. We invite researchers to participate in the workshop by submitting papers, attending talks from experts who have worked on these questions, and engaging in panel discussions with some of the leading senior researchers in the field. The workshop is structured to promote these goals, aiming for a better, safer, and fairer AI, and to enhance our understanding of machine learning. Our workshop's website draft: https://scslworkshop.github.io/.

Call For Paper

We invite researchers in machine learning and related fields to submit their recent work on spurious correlations and shortcut learning to SCSL main track. Accepted papers will be showcased as posters on the workshop day. Key topics of interest are outlined in the summary section.

Formatting Instructions

The main text must be between **6 and 10 pages** (inclusive). The page limit applies to both the initial and final camera ready version. The list of references does not count towards the page limit, and unlimited additional pages are allowed for the bibliography/references. Authors may use as many pages of appendices (after the bibliography) as they wish, but reviewers are not required to read the appendix.

Provided LaTeX style for ICLR 2025 should be used for submission: https://github.com/ICLR/Master-Template/raw/master/iclr2025.zip

Short Papers Track

We're looking to make our workshop at ICLR as inclusive and impactful as possible! That's why we're inviting submissions for the tiny track paper from newcomers to the field, as well as under-represented, under-resourced, and budding researchers. This will help foster

late-breaking developments and provide valuable feedback while opening the doors to a broader range of potential authors beyond the usual ML conference circuit. Tiny papers are also a platform to propose counter-intuitive results, proof of concepts for an idea, firming new perspectives, etc., that are not still in the stage of being published as a full paper.

Short papers follow the same formatting instructions as main track papers, except that they are limited to **4 pages** at submission time. Accepted tiny papers have an extra page for the camera-ready version of their work.

Reviewing Process

1. **Review Process:** All submissions to the workshop undergo a double-blind review process, ensuring impartiality and anonymity for both authors and reviewers. We aim to provide a fair, rigorous, and constructive review process that upholds the standards of the ICLR workshops. Submissions uploaded on OpenReview. Each paper will be assigned to at least two reviewers to ensure a balanced evaluation. A senior meta reviewer makes the final decision after a discussion with reviewers during the discussion period.
2. **Selection Criteria:** Submissions will be evaluated on the following criteria:
    - Novelty
    - Technical Quality
    - Relevance to the Workshop Theme
    - Clarity and Presentation
    - Potential Impact
3. **Reviewer Guidelines:** We will essentially follow the same guidelines for our reviewers as those used by the main conference, as outlined here.
4. **Reviewers:** We will issue a call to recruit reviewers who are experts in the relevant fields and the specific focus area of the workshop. We will release a Google form to invite reviewers. We plan to leverage social media, along with our personal networks, to publicly promote the form. Reviewers will be chosen based on their expertise, experience, and willingness to engage constructively with the submitted work. In case of emergency, if we cannot recruit enough reviewers through this process, we have already spoken to several graduate students from the University of Maryland, New York University, and Sharif University of Technology, as well as some senior researchers in the field. Including these contacts and the core organizing team, we already have over 25 reviewers available to assist.
5. **Review outcomes:** We will have meta-reviewers, who are more senior researchers, responsible for summarizing the reviews and making final decisions. Submissions will be categorized as:
    - Accepted
    - Rejected

6. **Post-Review:** Accepted papers will be presented during the workshop poster sessions. We will select multiple high-impact and interesting papers for oral presentation on the workshop day.

Timeline

- Submission deadline (both tracks): 3 February 2025 AOE
- Reviewer Assignment: 6 February 2025
- Review Period: 6 February 2025 to 20 February 2025
- Discussion Period: 21 February 2025 to 27 February 2025
- Acceptance Notification: 3 March 2025
- Camera-ready deadline: 12 April 2025

Conflict of Interest

To ensure fairness and transparency in the evaluation process, we will implement strict measures to manage conflicts of interest (COI) for our workshop. Organizers and program committee members will not be involved in assessing submissions from individuals with whom they share any potential conflicts, including those from the same organization, collaborators, or recent co-authors. We will employ a COI management system where reviewers will be required to declare any potential conflicts before being assigned papers. In cases where a conflict is identified, alternative reviewers without such conflicts will be assigned. These measures will ensure that the assessment of submissions is impartial, maintaining the integrity of the review process.

Keynote Speakers

1. **Pavel Izmailov** *(NYU / Anthropic)* **(Confirmed)**
   Email: pi390@nyu.edu
   Website: https://izmailovpavel.github.io/
   Experience: Pavel Izmailov is a researcher at Anthropic, focusing on reasoning, AI alignment, and AI for scientific discovery. He is set to join NYU in Fall 2025 as an Assistant Professor in the Tandon CSE department, with a courtesy appointment in the Courant CS department. With a Ph.D. from New York University advised by Andrew Gordon Wilson, Pavel has experience at OpenAI, xAI, and interned at Google and Amazon. His research spans core machine learning topics such as interpretability, probabilistic deep learning, out-of-distribution generalization, and improving problem-solving capabilities in AI systems.

2. **David Krueger** *(University of Cambridge)* **(Confirmed)**
   Email: david.scott.krueger@gmail.com

Homepage: https://davidscottkrueger.com/
Experience: David Scott Krueger is an Assistant Professor at the University of Cambridge, affiliated with the Department of Engineering's Information Engineering Division, where he is a member of the Computational and Biological Learning Lab (CBL). His research focuses on deep learning, AI alignment, and AI safety, with an interest in addressing existential risks posed by advanced AI systems. David earned his Ph.D. in Computer Science from the University of Montreal under the supervision of Aaron Courville, focusing on AI alignment and generalization in deep learning. His work has been recognized through significant grants, including from the Open Philanthropy Project, and he holds several affiliations with prominent AI research institutions such as the Center for the Study of Existential Risk (CSER), Quebec AI Institute (Mila), European Laboratory for Learning and Intelligent Systems (ELLIS) and the Center for Human Compatible AI (CHAI).

3. **Stefano Sarao Mannelli** *(Chalmers University/ Gothenburg University)* **(Confirmed)**
Email: stefano.sarao.mannelli@gu.se
Website: https://stefsmlab.github.io/
Experience: Stefano Sarao Mannelli is a tenure-track Assistant Professor in the Data Science and AI division of the Computer Science department at Chalmers University of Technology and Gothenburg University. His research is centered on developing a fundamental understanding of learning in machine learning systems, with a particular focus on bias generation and amplification. Before this role, Stefano completed postdoctoral research with Andrew Saxe at University College London and the University of Oxford, following his Ph.D. in Physics applied to Machine Learning from the University of Paris-Saclay, under the supervision of Lenka Zdeborova. He has expertise in machine learning theory, statistical physics, cognitive science, and modeling, with an emphasis on the statistical physics of learning and the dynamics of learning processes in machine learning systems.

4. **Baharan Mirzasoleiman** *(UCLA)* **(Confirmed)**
Email: baharan@cs.ucla.edu
Webpage: https://baharanm.github.io/
Experience: Baharan Mirzasoleiman is an Assistant Professor in the Computer Science Department at UCLA, where she leads the BigML research group. Her research focuses on enhancing the sustainability, reliability, and efficiency of machine learning systems. Specifically, she develops theoretically rigorous methods to select high-quality data for robust and efficient learning, improving big data quality. Her work also extends to improving machine learning models and algorithms, with applications in diverse areas such as medical diagnosis and environmental sensing. Prior to joining UCLA, Baharan was a postdoctoral research fellow at Stanford University, working with Jure Leskovec, and earned her Ph.D. in Computer Science

from ETH Zurich under the supervision of Andreas Krause. She has been recognized with an ETH medal for Outstanding Doctoral Thesis, was named a Rising Star in EECS by MIT, and received the prestigious NSF Career Award.

5. **Katherine Hermann** *(Google DeepMind)* **(Confirmed)**
   Email: hermannk@google.com
   Website: https://scholar.google.com/citations?user=owcAYmEAAAAJ&hl=en&oi=ao
   Experience: Katherine Hermann is a Senior Research Scientist at Google DeepMind, where she specializes in artificial intelligence and computer vision. She earned her Ph.D. from Stanford University, where her dissertation, titled "Understanding Feature Use Divergences Between Human and Machine Vision," explores the intricate differences in perception between humans and machines. Katherine's research contributions have been recognized through prestigious honors, including the NSF Graduate Research Fellowship Program (GRFP) Fellowship and the Best Poster Award at the SVRHM workshop during NeurIPS 2019. Prior to her current role, she was a Research Scientist at Google and completed research internships at Facebook AI and Google Brain, focusing on self-supervised computer vision models. An Outstanding Graduate (BA) in the College of Arts and Sciences at the University of Colorado, Katherine continues to work at the forefront of AI research.

Panelists

- **Soheil Feizi** *(UMD)* **(Confirmed)**
  Email: sfeizi@cs.umd.edu
  Website: https://www.cs.umd.edu/~sfeizi/
  Experience: Soheil Feizi is a faculty member and the director of Reliable AI Lab in the Computer Science department at University of Maryland, College Park (UMD). Currently on leave from his academic position, he is the Founder and CEO of RELAI , a startup dedicated to advancing AI reliability. He holds a Ph.D. from MIT and completed postdoctoral research at Stanford University . He has published over 100 peer-reviewed papers and given more than 50 invited talks. He has received multiple awards for his work including the ONR's Young Investigator Award, the NSF CAREER Award, the ARO's Early Career Program Award, two best paper awards, the Ernst Guillemin Thesis Award, a Teaching Award, and more than fifteen research awards from national agencies such as NSF, DARPA, ARL, ONR, DOE, NIST as well as industry such as Meta, IBM, Amazon, Qualcomm and Capital One. His work has been featured by major outlets such as the Washington Post, BBC, MIT Technology Review, Bloomberg, and the Wire . Recently, he testified before the U.S. House's Bipartisan Task Force on AI, reflecting his commitment to ensuring AI is developed with safety, accuracy, and reliability in mind. He is committed to promoting diversity in STEM and

has mentored several high school, undergraduate, and graduate students through various programs.

- **Andrew Gordon Wilson** *(NYU)* **(Confirmed)**
  Email: andrewgw@cims.nyu.edu
  Website: https://cims.nyu.edu/~andrewgw/
  Experience: Andrew Gordon Wilson is a Professor at the Courant Institute of Mathematical Sciences and the Center for Data Science at New York University, where he engages in pioneering research at the intersection of learning and decision-making. His work focuses on developing intelligent systems through the discovery of scientifically interpretable structures in data, with particular emphasis on probabilistic deep learning, scalable Gaussian processes, and AI alignment. Andrew's research has practical applications across diverse fields such as time series analysis, natural language processing, public policy, and medicine. He has contributed significantly to the academic community, introducing several software libraries to promote open and reproducible research. Andrew received his Ph.D. from Trinity College, University of Cambridge, and has held faculty positions at Cornell University and Carnegie Mellon University. His excellence in research has been recognized with numerous awards, including the NSF Career Award, Best Paper Award at the ICML Theoretical Foundations Workshop, Outstanding Paper Award at ICML for his work on Bayesian Model Selection, and the Amazon Machine Learning Research Award. Additionally, he has received the Outstanding Area Chair Award at ICLR, Best Paper Award at the NeurIPS Time Series Workshop, and Outstanding Ph.D. Dissertation from G-Research, among others, underscoring his influential contributions to the field.

Tentative Schedule

In addition to a diverse lineup of talks, our workshop will actively promote discussion and collaboration through various avenues. We have scheduled a panel session with leading experts to provide deeper insights into the topic, along with dedicated networking opportunities to foster interaction among participants. Each talk will also be followed by a 15-minute discussion period to ensure ample time for engaging with the audience's questions and ideas.

Start Time	Plan	Details
9:00	Opening	
9:10	Invited Talk 1	35 + 15 minutes for Q&A and discussion
10:00	Invited Talk 2	35 + 15 minutes for Q&A and discussion

10:50	Break, Informal Discussion, and Networking	20 minutes break
11:10	Invited Talk 3	35 + 15 minutes for Q&A and discussion
12:00	Poster Session 1	
12:30	Lunch Break	1 hour break
13:30	Invited Talk 4	35 + 15 minutes for Q&A and discussion
14:20	Invited Talk 5	35 + 15 minutes for Q&A and discussion
15:10	Break, Informal Discussion and Networking	
15:30	Oral Presentation 1	10 + 10 minutes for Q&A and discussion
15:50	Oral Presentation 2	10 + 10 minutes for Q&A and discussion
16:10	Oral Presentation 3	10 + 10 minutes for Q&A and discussion
16:30	Interactive Panel Discussion	
17:30	Poster Session 2	

Modality and Accessibility

This workshop is designed as a hybrid event, allowing participants to join either in-person or virtually. It will feature four main types of events: invited talks, panel discussions, contributed talks, and poster sessions. Abstracts for accepted papers and talks will be shared at least one week prior to the workshop, and accepted papers will be published on OpenReview.

1. **Invited Talks and Panel Discussions**: All invited talks and the panel discussion will be recorded, captioned, and made accessible to ICLR attendees, ensuring inclusivity across time zones.
2. **Contributed Talks (Oral Presentations)**: The Program Committee will select several submissions to be presented as oral contributed talks. Papers that have been previously published will not be considered for oral presentations.
3. **Poster Sessions**: All accepted papers must be accompanied by a poster. Our schedule includes poster sessions to encourage interaction among attendees. All posters and camera-ready versions of the accepted papers will be publicly available by the day of the workshop.

For remote participants, we will live-stream the invited talks, panel discussions, and contributed talks via Zoom and provide support for remote Q&A. Additionally, we will host the accepted papers and posters on our website. To further foster discussion, we will create a Discord channel for all attendees to engage in ongoing conversations throughout the workshop.

Organizers

Our team brings a wealth of experience in organizing high-impact workshops, seminars, and challenges, making us well-suited to host this workshop on spurious correlations. The workshop will be organized by researchers from New York University (NYU), University of Maryland (UMD), and Sharif University of Technology (SUT), specifically RIML and MLL Labs at SUT. These labs have a strong track record of organizing successful events. Notable among these are the International Winter Seminar Series (WSS), which has hosted both in-person and virtual seminars covering a broad spectrum of topics in artificial intelligence, and the Rayan Challenge, which focused on the trustworthiness of deep learning models. Moreover, one of our team members, Aahlad Puli, co-organized the highly relevant workshop *Spurious Correlations, Invariance, and Stability (SCIS)* at ICML 2023.

Team members (sorted alphabetically):

- **Hesam Asadollahzadeh**
    - Hesam Asadollahzadeh is a Ph.D. student at New York University (NYU). He recently completed his B.Sc. in Computer Engineering at the University of Tehran, where he received the best undergraduate thesis award for his work on out-of-distribution generalization and mitigating spurious correlations and shortcut learning. Additionally, he has worked as a research scientist at the Wellcome Sanger Institute, University of Cambridge, under the supervision of Dr. Mo Lotfollahi. Hesam's research focuses on trustworthy machine learning, with an emphasis on the distributional and adversarial robustness of deep networks. His interests also extend to deep generative models, computational biology, and self-supervised learning.
- **Mahdi Ghaznavi**
    - Mahdi Ghaznavi is a last-year M.Sc. student in AI & Robotics at Sharif University of Technology (SUT), Iran. He is co-supervised by Dr. M.H. Rohban and Dr. M. Soleymani Baghshah at Robust and Interpretable Machine Learning (RIML) and Machine Learning Lab (MLL). His research focuses on out-of-distribution generalization, robustness to spurious correlation, and shortcut learning. He is interested in how deep models learn spurious correlations and shortcuts, their impacts on feature learning in DNNs, the role of optimization algorithms and models' inductive biases in learning shortcuts, and how to build models and algorithms that are robust to spurious correlations with minimal supervision. He obtained his bachelor's in both computer engineering and mathematics and applications (majoring in pure mathematics) at SUT. He has been a research assistant, working on interpretable molecular optimization with Dr. V. Garg at Aalto University, Finland.
- **Parsa Hosseini**
    - Parsa Hosseini is a Ph.D. student in Computer Science at the University of Maryland, where he is advised by Professor Soheil Feizi. His research focuses on addressing spurious correlations in AI models, with an emphasis on developing methods to detect and mitigate their effects. Parsa earned his BSc in Computer Engineering from Sharif University of Technology in 2023 before pursuing his doctoral studies. His work aims to enhance the reliability and fairness of machine learning systems by tackling one of the key challenges in model generalization.
- **Polina Kirichenko**
    - Polina Kirichenko is a machine learning researcher with a strong focus on generalization, robustness, and fairness in AI. She recently completed her

Ph.D. at New York University's Center for Data Science, where she worked with Professor Andrew Gordon Wilson on probabilistic machine learning and Bayesian deep learning, particularly investigating uncertainty estimation and robustness to distribution shifts. Polina has also contributed to cutting-edge research as a Visiting Researcher at FAIR Labs (Meta AI) and through internships at Meta AI, Google DeepMind, and Cold Spring Harbor Laboratory. She holds a Bachelor's degree in Computer Science from the Higher School of Economics, where she collaborated with Professor Dmitry Vetrov on Bayesian methods. Polina has been recognized as both a Google Generation Scholar and a DeepMind Fellow.

- **Arash Marioryad**
    - Arash Marioriyad is a master's student specializing in Artificial Intelligence at the Sharif University of Technology, Tehran, Iran. His research centers on compositional generation in text-to-image models and modularity in deep learning. By drawing on cognitive insights, Arash seeks to enhance the compositional generation capabilities of generative models, aiming to advance their creative potential.
- **Nahal Mirzaie**
    - Nahal Mirzaie is a Ph.D. candidate in AI and Robotics at Sharif University of Technology (SUT), Iran, where she is supervised by Dr. M.H. Rohban in the Robust and Interpretable Machine Learning (RIML) lab. Her research focuses on enhancing group robustness, addressing spurious correlations, and exploring shortcut learning in machine learning models. Nahal's earlier work involved drug discovery, and she has also served as a research assistant at Aalto University, Finland, where she collaborated with Dr. Vikas Garg on investigating the expressivity of graph neural networks (GNNs).
- **Aahlad Puli**
    - Aahlad Puli is a Faculty Fellow at the Center for Data Science at NYU. Motivated by issues in healthcare, he develops techniques for causal inference and out-of-distribution generalization. He uses these techniques to develop clinical risk factors and models that transport across populations. His work has appeared in machine learning conferences such as ICLR, AISTATs, and NeurIPS. He has co-organized the SCIS workshop at ICML 2022 and 2023. Aahlad finished his Ph.D. at the Courant Institute at NYU, advised by Rajesh Ranganath, and is a recipient of the Apple Scholars in AI/ML Ph.D. fellowship for 2022.
- **Shikai Qiu**
    - Shikai Qiu is a Ph.D. student in Computer Science at NYU Courant, working under the supervision of Professor Andrew Gordon Wilson. He has worked on data-efficient group robustness, probabilistic models, and understanding and improving neural scaling laws. Shikai is also a Student Researcher at Google

DeepMind, where he is supervised by Jeffrey Pennington. He has previously contributed to Meta AI's protein team, collaborating with Alex Rives and Tom Sercu, and worked at Amazon AWS with Boran Han and Danielle Robinson. Shikai holds a degree in Physics and Computer Science from UC Berkeley, where he developed equivariant deep learning models for high energy physics and biochemistry, working with Jennifer Listgarten, Haichen Wang, and Ben Nachman.

- **Mohammad Hossein Rohban**
    - Mohammad Hossein Rohban is an Assistant Professor in the Department of Computer Engineering at Sharif University of Technology, where he earned his BS, MS, and Ph.D. degrees in Computer Engineering. As the Principal Investigator of Robust and Interpretable Machine Learning Lab (RIML) at Sharif, his research spans interpretable and robust machine learning, anomaly detection, and computational biology, with an additional focus on out-of-distribution generalization and mitigating the impact of spurious correlations in AI models. Prior to his current role, Mohammad Hossein was a Postdoctoral Associate at the Broad Institute of Harvard and MIT, where he contributed to groundbreaking research at the intersection of machine learning and image-based computational biology. He has also held positions as a part-time lecturer at Rochester Institute of Technology and conducted postdoctoral research at Boston University.
- **Mahdieh Soleymani Baghshah**
    - Mahdieh Soleymani Baghshah is an Associate Professor in the Department of Computer Engineering at Sharif University of Technology, where she earned her B.Sc., M.Sc., and Ph.D. degrees in 2003, 2005, and 2010, respectively. As the Principal Investigator of the Machine Learning Lab (MLL) at Sharif, her research focuses on machine learning and deep learning, with particular interests in reinforcement learning, meta-learning, out-of-distribution generalization, and mitigating shortcut learning. Mahdieh is a leading expert in her field, working to develop more robust and generalizable AI models that can perform reliably in diverse and challenging environments.

Previous Related Workshops

In recent years, various workshops have explored challenges related to spurious correlations, distribution shifts, and out-of-distribution (OOD) generalization in machine learning. Among them, the **Spurious Correlations, Invariance, and Stability (SCIS)** workshops at ICML 2022 and 2023 stand out as the only events explicitly focused on spurious correlations and shortcut learning. These workshops addressed how spurious correlations arise, their impact on generalization, and methods such as invariance and stability for mitigating their effects.

Other workshops have addressed related topics, albeit from broader perspectives. The **Distribution Shifts (DistShift)** workshops at NeurIPS 2022 and 2023, for example, focused on the challenges posed by distribution shifts, which are often linked to spurious correlations but not their primary focus. Additionally, several workshops on OOD generalization and adaptation, such as **Out Of Distribution Generalization in Computer Vision (OOD-CV)** at ECCV 2024, **OOD Generalization and Adaptation in Natural and Artificial Intelligence** at NeurIPS 2021, **What do we need for successful domain generalization** at ICLR 2023, and **Generalization beyond the training distribution in brains and machines** at ICLR 2021, have explored how models can generalize beyond training data in various settings.

While the SCIS workshops at ICML 2022 and 2023 played a crucial role in establishing foundational discussions on spurious correlations and shortcut learning, two years have passed, and the landscape of machine learning has evolved significantly. Notably, the rise of foundation models has brought new challenges and opportunities related to how these large-scale models handle spurious correlations and shortcut learning. Since SCIS, several groundbreaking works, including those published by some of our confirmed invited speakers, have provided fresh insights into these issues.

Given the rapid progress in the field, it is imperative to revisit this topic with a broader and more up-to-date perspective. Our proposed workshop aims to build upon previous efforts, while also addressing emerging questions in the context of foundation models, ensuring that the community continues to critically engage with the core challenges and solutions surrounding spurious correlations and their impact on generalization. This timely re-examination will foster deeper collaboration and innovation, advancing both the theoretical understanding and practical approaches to these problems.

Audience: Advertisement Plan and Anticipated Size

The following advertisement plan outlines a comprehensive strategy that leverages multiple channels to ensure the workshop's success and attract a wide audience of relevant researchers and professionals.

1. **Social Media Platforms:** Social media platforms provide an effective way to reach a large, targeted audience quickly. The workshop will be promoted on the Facebook, r/MachineLearning, LinkedIn, and X (formerly Twitter) platforms. Our workshop’s X account ID is @scslworkshop.
2. **Email Outreach Campaign:** Sending direct email invitations and announcements to relevant researchers and professionals in the field will be used to create personalized engagement.
3. **Academic Platforms and Announcement Boards:** A strong presence on academic platforms where researchers regularly visit to find conferences and workshops is essential to attract attention from the academic community. The workshop will be promoted on the ResearchGate Platform and the workshop's venue page in OpenReview.

4. **Workshop Website:** We have designed a dedicated workshop website that serves as the central hub for information (Link: https://scslworkshop.github.io).

Based on previous workshop experience on similar topics, we anticipate **about 100 in-person and 500 virtual attendees**.

Diversity Commitment

Our workshop organizers represent a diverse group of 4 nationalities (Iranian, Chinese, Russian, Indian) from 3 different institutions (Sharif University of Technology, New York University, University of Maryland) across 2 countries (Iran, U.S.A). The team is comprised of 3 women and 7 men.
Our invited speakers come from various nationalities (U.K., Iran, Italy, Russia, U.S.A) and different institutions across 3 countries (Sweden, U.S.A, U.K.), including 2 women and 3 men. We are committed to diversity in our advertising plan to reach researchers from different backgrounds, ethnicities, and more. Additionally, we have a tiny track paper category to encourage submissions from newcomers to the field, as well as under-represented, under-resourced, and budding researchers.

Conclusion

This workshop on spurious correlations and shortcut learning in AI systems comes at a crucial time in the field's development. By bringing together researchers to explore new benchmarks, robustification methods, and fundamental theoretical understanding across multiple modalities and learning paradigms, this workshop aims to advance our ability to build AI systems that rely on meaningful, generalizable patterns rather than spurious correlations. This work is essential for ensuring the reliability, fairness, and trustworthiness of AI systems as they continue to be deployed in critical real-world applications.