Robotics

Physically Plausible 4D Reconstruction from Monocular Videos

Gengshan Yang (CMU), Shuo Yang (CMU), John Z. Zhang (CMU), Zachary Manchester (CMU), Deva Ramanan (CMU)

Given monocular videos, we build models of articulated objects and environments whose 4D configurations satisfy dynamics and contact constraints. At its core, our method leverages differentiable physics simulation to aid visual reconstructions. We couple differentiable physics simulation with differentiable rendering via coordinate descent, which enables end-to-end optimization of, not only 4D reconstructions, but also physical system parameters from videos. We demonstrate the effectiveness of physics-informed reconstruction on monocular videos of quadruped animals and humans. It reduces reconstruction artifacts (e.g., scale ambiguity, unbalanced poses, foot skating) that are challenging to address by visual cues alone.

Syntax-Guided Transformers: Elevating Compositional Generalization and Grounding in Multimodal Environments

Danial Kamali (Michigan State University), Parisa Kordjamshidi (Michigan State University)

Compositional generalization, the ability of intelligent models to extrapolate understanding of components to novel compositions, is a fundamental yet challenging facet in AI research, especially within multi-modal environments. In this work, we address this challenge by exploiting the syntactic structure of language to boost compositional generalization. This paper elevates the importance of syntactic grounding, particularly through attention-masking techniques derived from text input parsing. We introduce and evaluate the merits of using syntactic information in the multi-modal grounding problem. Our results on grounded compositional generalization underscore the superior performance of dependency parsing across diverse tasks when utilized with Weight Sharing across the Transformer encoder. The results push the state-of-the-art in multi-modal grounding and parameter-efficient modeling and provide insights for future research.

LLM-Grounder: Open-Vocabulary 3D Visual Grounding with Large Language Model as an Agent

Jianing Yang (UM), Xuweiyi Chen (UM), Shengyi Qian (UM), Nikhil Madaan (independent researcher), Madhavan Iyengar (UM), David F. Fouhey (UM, NYU), Joyce Chai (UM)

3D visual grounding is a critical skill for household robots, enabling them to navigate, manipulate objects, and answer questions based on their environment. While existing approaches often rely on extensive labeled data or exhibit limitations in handling complex language queries, we propose LLM-Grounder, a novel zero-shot, open-vocabulary, Large Language Model (LLM)-based 3D visual grounding pipeline. LLM-Grounder utilizes an LLM to decompose complex natural language queries into semantic constituents and employs a visual grounding tool, such as OpenScene or LERF, to identify objects in a 3D scene. The LLM then evaluates the spatial and commonsense relations among the proposed objects to make a final grounding decision. Our method does not require any labeled training data and can generalize to novel 3D scenes and arbitrary text queries. We evaluate LLM-Grounder on the ScanRefer benchmark and demonstrate state-of-the-art zero-shot grounding accuracy. Our findings indicate that LLMs significantly improve the grounding capability, especially for complex language queries, making LLM-Grounder an effective approach for 3D vision-language tasks in robotics.

VLN-Trans: Translator for the Vision and Language Navigation Agent

Yue Zhang(Michigan State University), Parisa Kordjamshidi(Michigan State University)

Language understanding is essential for the navigation agent to follow instructions. We observe two kinds of issues in the instructions that can make the navigation task challenging: 1. The mentioned landmarks are not recognizable by the navigation agent due to the different vision abilities of the instructor and the modeled agent. 2. The mentioned landmarks are applicable to multiple targets, thus not distinctive for selecting the target among the candidate viewpoints. To deal with these issues, we design a translator module for the navigation agent to convert the original instructions into easy-to-follow sub-instruction representations at each step. The translator needs to focus on the recognizable and distinctive landmarks based on the agent's visual abilities and the observed visual environment. To achieve this goal, we create a new synthetic sub-instruction dataset and design specific tasks to train the translator and the navigation agent. We evaluate our approach on Room2Room (R2R), Room4room (R4R), and Room2Room Last (R2R-Last) datasets and achieve state-of-the-art results on multiple benchmarks.

Fairness & Inclusivity of AI

Optimal Probability Recalibration

Zeyu Sun (Electrical And Computer Enginnering, University of Michigan), Dogyoon Song (Electrical And Computer Enginnering, University of Michigan), Alfred Hero (Electrical And Computer Enginnering, University of Michigan)

Recalibrating probabilistic classifiers is vital for enhancing the reliability of machine learning models. Despite the rapid development of recalibration algorithms, there is still a lack of a comprehensive theory that integrates calibration and sharpness. We introduce the minimum-risk recalibration within the framework of mean-squared-error (MSE) decomposition, offering a principled approach for evaluating and recalibrating probabilistic classifiers. Using this framework, we analyze the uniform-mass binning recalibration method and establish a finite-sample risk upper bound of order O(B/n + 1/B^2) where B is the number of bins and n is the sample size. By balancing calibration and sharpness, we further determine that the optimal number of bins for UMB scales with $n^{1/3}$, resulting in a risk bound of approximately O(n^{-2/3}). Additionally, we tackle the challenge of label shift by proposing a two-stage approach. Our results show that transferring a calibrated classifier requires significantly fewer target samples compared to recalibrating from scratch. We validate our theoretical findings through numerical simulations, which confirm the tightness of the proposed bounds, the optimal number of bins, and the effectiveness of label shift adaptation.

You Are What You Annotate: Towards Better Models through Annotator Representations

Naihao Deng (University of Michigan), Siyang Liu (University of Michigan), Xinliang Frederick Zhang (University of Michigan), Winston Wu (University of Michigan), Lu Wang (University of Michigan), Rada Mihalcea (University of Michigan)

Annotator disagreement is ubiquitous in natural language processing (NLP) tasks. There are multiple reasons for such disagreements, including the subjectivity of the task, difficult cases, unclear guidelines, and so on. Rather than simply aggregating labels to obtain data annotations, we instead try to model the diverse perspectives directly and propose to explicitly account for the annotator idiosyncrasies and leverage them in the modeling process. We create representations for each annotator (annotator embeddings) and also their annotations (annotation embeddings). Our approach helps the model learn significantly better from disagreements on six different NLP datasets while increasing model size by fewer than 1% parameters. By capturing the unique tendencies and subjectivity of individual annotators through embeddings, our representation primes AI models to be inclusive of diverse viewpoints.

Visual Geo-Diversity Annotations on a Budget

Oana Ignat, Longju Bai, Joan Nwatu, Rada Mihalcea

Current AI foundation models have shown impressive performance across various tasks. However, several studies have revealed that these models are not effective for everyone due to the imbalanced geographical and economic representation of the data used to train them. Most of this data comes from Western countries, leading to poor results for underrepresented countries. To address this issue, more data needs to be collected from these countries. However, the cost of annotating this data is a significant bottleneck. In this paper, we propose methods to identify the most effective data to annotate. Our approach involves finding countries with images of topics (objects and actions) that are visually distinct from those in most current training data for vision and vision-language foundation models. Additionally, we identify countries that visually depict topics similarly and show that using data from these countries to supplement the training data improves model performance and reduces annotation costs.

Mitigating the Effects of Label Bias: An Expectation-Maximization Approach

Trenton Chang (CSE, University of Michigan), Jenna Wiens (CSE, University of Michigan)

Disparate censorship arises when access to ground truth labels used to train machine learning (ML) models varies across subgroups. Such differences can emerge due to biases in decisions to collect labels. For example, in healthcare, labels used in ML analyses often depend on laboratory testing decisions. In hiring decisions, labels could depend on screening tools that determine who is invited to interview. In both settings, ground truth labels are only available for a potentially-biased subset of the cohort of interest. Many approaches do not explicitly address such sources of label bias, and assume that individuals without confirmed ground truth labels are negative during model training (e.g., those who never receive a test are disease-free, those eliminated prior to being interviewed are poor job candidates), which may exacerbate biases in labeling decisions. Thus, we propose Disparate Censorship-based Expectation-Maximization (DCEM), a novel approach for mitigating bias in label collection decisions under disparate censorship. On synthetic data, we demonstrate that DCEM mitigates bias effectively while maintaining competitive performance compared to the best baseline (median ROC gap & range: 3.1 [1.2 - 6.0] vs. 4.4 [2.0 - 8.3] and median AUC & range: 0.787 [0.768 - 0.823] vs. 0.815 [0.623 - 0.867]). We demonstrate similar bias mitigation results on a pseudosynthetic sepsis classification task based on real data (ROC gap: 0.115 [0.071 - 0.139] vs. 0.162 [0.096 - 0.213]). The results demonstrate that our method can partially mitigate the impacts of biased decision-making on ML models.

Bridging the Digital Divide: Performance Variation across Socio-Economic Factors in Vision-Language Models

Joan Nwatu (University of Michigan), Oana Ignat (University of Michigan), Rada Mihalcea (University of Michigan)

Despite the impressive performance of current AI models reported across various tasks, performance reports often do not include evaluations of how these models perform on the specific groups that will be impacted by these technologies. Among the minority groups under-represented in AI, data from low-income households are often overlooked in data collection and model evaluation. We evaluate the performance of a state-of-the-art vision-language model (CLIP) on a geo-diverse dataset containing household images associated with different income values (Dollar Street) and show that performance inequality exists among households of different income levels. Our results indicate that performance for the poorer groups is consistently lower than the wealthier groups across various topics and countries. We highlight insights that can help mitigate these issues and propose actionable steps for economic-level inclusive AI development.

Towards Algorithmic Fidelity: Mental Health Representation across Demographics in Synthetic vs. Human-generated Data

Shinka Mori(University of Michigan), Andrew Lee(University of Michigan), Oana Ignat(University of Michigan), Rada Mihalcea(University of Michigan)

We explore the capabilities of GPT-3 as a synthetic data generator for mental health datasets. Using GPT-3, we develop HeadRoom, a synthetic dataset of 3,120 posts about depression-triggering stressors, by controlling for race, gender, and time frame (before and after COVID-19). We conduct semantic and lexical analyses to (1) identify the predominant stressors for each demographic group; and (2) compare our synthetic data to a human-generated dataset. Our findings show that synthetic data mimics some of the human-generated data distribution for the predominant depression stressors across diverse demographics.

Causal NLP Research and the Way towards Social Good

Zhijing Jin (Max Planck Institute & ETH)

This poster will be a compilation of my research building socially responsible LLMs using causal and moral principles. Specifically, I use causal inference to benchmark existing LLMs' reasoning ability, analyze the failure modes of LLMs, and interpret the relation between data collection and model learning. Further, I combine interdisciplinary knowledge from moral philosophy to design socially-important moral questions to test LLMs, and propose standards for future models to be more morally safe.

Interpretability & Transparency of AI

DeckFlow: A Card Game Interface for Exploring Generative Model Flows

Gregory Croisdale, Emily Huang, John Chung, Xu Wang, Anhong Guo

Recent Generative AI models have been shown to be substantially useful in different fields, often bridging modal gaps, such as text-prompted image or human motion generation. However, their accompanying interfaces do not sufficiently support iteration and interaction between models, and due to the computational intensity of generative technology, can be unforgiving to user errors and missteps. We propose DeckFlow, a no-code interface for multimodal generative workflows which encourages rapid iteration and experimentation between disparate models. DeckFlow emphasizes the persistence of output, the maintenance of generation settings and dependencies, and continual steering through user-defined concept groups. Taking design cues from Card Games and Affinity Diagrams, DeckFlow is aimed to lower the barrier for non-experts to explore and interact with generative AI.

Human Inspired Progressive Alignment and Comparative Learning for Grounded Word Acquisition

Yuwei Bao (University of Michigan), Barrett Martin Lattimer (University of Michigan, ASAPP), Joyce Chai (University of Michigan)

Human language acquisition is an efficient, supervised, and continual process. In this work, we took inspiration from how human babies acquire their first language, and developed a computational process for word acquisition through comparative learning. Motivated by cognitive findings, we generated a small dataset that enables the computation models to compare the similarities and differences of various attributes, learn to filter out and extract the common information for each shared linguistic label. We frame the acquisition of words as not only the information filtration process, but also as representation-symbol mapping. This procedure does not involve a fixed vocabulary size, nor a discriminative objective, and allows the models to continually learn more concepts efficiently. Our results in controlled experiments have shown the potential of this approach for efficient continual learning of grounded words.

Measuring the Impact of AI in the Diagnosis of Hospitalized Patients: A Web-based Randomized Survey Vignette Multicenter Study

Sarah Jabbour (University of Michigan), David Fouhey (New York University), Stephanie Shepard (University of Michigan), Thomas S. Valley (Michigan Medicine), Ella A. Kazerooni (Michigan Medicine), Nikola Banovic (University of Michigan), Jenna Wiens (University of Michigan), Michael W. Sjoding (Michigan Medicine)

Artificial intelligence (AI) could support clinicians when diagnosing hospitalized patients. However, ML models trained on clinical data are prone to making biased predictions based on clinically irrelevant findings. Left unchecked, models could exacerbate biases widespread in healthcare, and errors in clinical judgement influenced by AI could have severe consequences such as patient harm. Furthermore, recent regulatory guidance has called for AI models to include explanations to mitigate errors made by models, but the effectiveness of this strategy has not been established. Through a randomized clinical vignette survey study administered across 14 US states, we evaluate the impact of systematically biased AI on clinician diagnostic accuracy, and determine if image-based AI model explanations can mitigate errors made by such models. The results showed that clinicians' diagnostic accuracy improved when shown predictions from a standard AI model. However, when presented with systematically biased AI predictions, clinician diagnostic accuracy decreased.

Metabolism-inspired Mechanistic Deep Learning for Treating Drug-resistant Infections

Harkirat Singh Arora (Biomedical Engineering), Sriram Chandrasekaran (Biomedical Engineering)

Antibiotic resistance (AR) is a pressing global health concern; new treatments are urgently needed. Drug combinations are a promising solution to this problem, but they are designed empirically, driven by clinical intuition, leading to suboptimal results and increased AR. The combinatorial explosion further aggravates the problem. Therefore, there is a need for an efficient data-driven approach, to facilitate the development of these treatments. We have developed a mechanistic approach that combines multi-omics data with artificial neural networks (ANN) to design effective drug combinations. Our approach (a) accurately predicts multi-way drug interactions in E. coli (R = 0.58, p~10^-14) and M. tuberculosis (R = 0.42, p~10^-8), (b) accommodates strains of critical drug-resistant bacterial pathogens, S. typhimurium and P. aeruginosa, (c) provides insights into a diverse set of pathways predictive of drug combinations at the molecular level. The mechanistic approach identified Alternate Carbon Metabolism in E. coli and Fatty Acid Metabolism in M. tuberculosis along with Transport mechanisms in both, as the most important pathways governing drug resistance, in agreement with literature evidence.

Using Persuasive Writing Strategies to Explain and Detect Health Misinformation

Danial Kamali (Michigan State University), Joseph Romain (Michigan State University), Huiyi Liu (Michigan State University), Wei Peng (Michigan State University), Jingbo Meng (Ohio State University), Parisa Kordjamshidi (Michigan State University)

Nowadays, the spread of misinformation is a prominent problem in society. To address the issue, the AI research community across different sub-fields has been actively engaged in developing automated solutions for misinformation detection. Our research focuses on aiding the identification of misinformation by analyzing the persuasive strategies employed in textual documents. The persuasive strategies can serve as valuable insights and explanations, enabling other models or even humans to make more informed decisions regarding the trustworthiness of the information. To achieve our objective, we introduce a novel annotation scheme that encompasses common persuasive writing tactics. Additionally, we provide a dataset on health misinformation, thoroughly annotated by humans utilizing our proposed scheme. Our contribution includes proposing a new task of classifying pieces of text with their persuasive writing strategy types. We develop language model-based baselines for both persuasive strategy labeling and misinformation detection. We delve into the effects of employing persuasive strategies as intermediate labels in the context of misinformation detection. Our results show the automatic analysis and labeling of those strategies not only enhances accuracy but also improves the explainability of misinformation detection models. We are committed to promoting further research in this area by making our newly annotated resource and baseline models publicly accessible.

Reliability & Safety of AI

Reinforcement Learning for Red-Teaming LLMs

Aylin Gunal (University of Michigan), Namho Koh (University of Michigan)

In this work-in-progress, we experiment with using reinforcement learning to modify prompts in such a way that we can effectively red-team LLMs.

GRAPHITE: Generating Automatic Physical Examples for Machine-Learning Attacks on Computer Vision Systems

Ryan Feng (University of Michigan), Neal Mangaokar (University of Michigan), Jiefeng Chen (University of Wisconsin-Madison), Earlence Fernandes (University of Wisconsin-Madison), Somesh Jha (University of Wisconsin-Madison), Atul Prakash (University of Michigan)

This paper investigates an adversary's ease of attack in generating adversarial examples for real-world scenarios. We address three key requirements for practical attacks for the real-world: 1) automatically constraining the size and shape of the attack so it can be applied with stickers, 2) transform-robustness, i.e., robustness of a attack to environmental physical variations such as viewpoint and lighting changes, and 3) supporting attacks in not only white-box, but also black-box hard-label scenarios, so that the adversary can attack proprietary models. In this work, we propose GRAPHITE, an efficient and general framework for generating attacks that satisfy the above three key requirements. GRAPHITE takes advantage of transform-robustness, a metric based on expectation over transforms (EoT), to automatically generate small masks and optimize with gradient-free optimization. GRAPHITE is also flexible as it can easily trade-off transform-robustness, perturbation size, and query count in black-box settings. On a GTSRB model in a hard-label black-box setting, we are able to find attacks on all possible 1,806 victim-target class pairs with averages of 77.8% transform-robustness, perturbation size of 16.63% of the victim images, and 126K queries per pair.

Privacy & Safety of AI

SafeAR: Towards Safer Algorithmic Recourse by Risk-Aware Policies

Haochen Wu (University of Michigan), Shubham Sharma (J.P. Morgan AI Research), Sunandita Patra (J.P. Morgan AI Research), Sriram Gopalakrishnan (J.P. Morgan AI Research)

With the growing use of machine learning (ML) models in critical domains such as finance and healthcare, the need to offer recourse for those adversely affected by the decisions of ML models has become more important. Sequential algorithmic recourse provide recommendations on actions to take for improving individual's situation and thus receiving a favorable decision. However, it is undesirable if a recourse could result in a worse situation from which recovery requires an extremely high cost. It is essential to incorporate risks when computing and evaluating recourse. We call the recourse computed with such risk considerations as Safer Algorithmic Recourse (SafeAR). The objective is to empower people to choose a recourse based on their risk tolerance. We discuss and show how existing recourse desiderata can fail to capture the risk of higher costs. We present a method to compute recourse policies that consider variability in cost and connect algorithmic recourse literature with risk-sensitive reinforcement learning. We also adopt measures “Value at Risk” and “Conditional Value at Risk” from the financial literature to summarize risk concisely. We apply our method to two real-world datasets and compare policies with different levels of risk-aversion using risk measures and recourse desiderata.

Speech & Language

Merging Generated and Retrieved Knowledge for Open-Domain QA

Yunxiang Zhang (University of Michigan), Muhammad Khalifa (University of Michigan), Lajanugen Logeswaran (LG AI Research), Moontae Lee (LG AI Research & University of Illinois at Chicago), Honglak Lee (University of Michigan & LG AI Research), Lu Wang (University of Michigan)

Open-domain question answering (QA) systems are often built with retrieval modules. However, retrieving passages froma given source is known to suff er from insufficient knowledge coverage. Alternatively, prompting large language models (LLMs) to generate contextual passages based on their parametric knowledge has been shown to improve QA performance. Yet, LLMs tend to "hallucinate" content that conflicts with the retrieved knowledge. Based on the intuitionthat answers supported by both sources are more likely to be correct, we propose COMBO, a Compatibility-Oriented knowledge Merging for Better Open-domain QA framework, to effectively leverage the two sources of information. Concretely, we match LLM-generated passages with retrieved counterparts into compatible pairs, based on discriminators trained with silver compatibility labels. Then a Fusion-in-Decoder-based reader model handles passage pairs to arrive at the final answer. Experiments on four benchmarks for single and multi-hop open-domain QA tasks show that COMBO uniformly outperforms competitive baselines on all testbeds.

MRAFE: Multimodal Retrieval Augmented Feature Extractor

Tien-Lan Sun (University of California, Berkeley), Hinson Chan (EduBeyond), Vincent Qi (EduBeyond), Alexander Ng (EduBeyond), Anaiy Somalwar (University of California, Berkeley)

Transformer-based Large Language Models (LLMs) are conventionally trained on diverse, large-scale, industry-agnostic data. However, recent research has demonstrated that significant improvements can be made for highly technical, domain-specific, downstream tasks. In addition to application-specialized fine-tuning, Retrieval Augmented Generation (RAG) with vector databases has been used to improve the accessible knowledge base for task-specific applications. In this paper, we introduce a novel multimodal text extractor and summarizer pipeline using a YOLO-based backbone and Bi-LSTM head to extract key information in education datasets before appending de-biased data to our vector database. We compare our methodology with a variety of state-of-the-art large-language models on CRISDE600: a proprietary cross-domain, education-focused dataset with over 600 human-annotated visual question-and-answer examples. We are able to improve upon state-of-the-art performance when testing accuracy, recall, and precision and are currently working towards implementing reinforcement learning with human feedback to improve domain-specific fine-tuning.

Guiding Chain-of-Thought Reasoning with a Correctness Discriminator

Muhammad Khalifa (University of Michigan), Lajanujen Logeswaran (LG AI), Moontae Lee (LG AI), Honglak Lee (University of Michigan), Lu Wang (University of Michigan)

In the context of multi-step reasoning, e.g., with chain-of-thoughts, language models (LMs) can easily assign a high likelihood to incorrect steps. As a result, decoding strategies that optimize for solution likelihood often yield incorrect solutions. To address this issue, we propose Guiding chain-of-thought ReAsoning with a CorrectnEss Discriminator (GRACE), a stepwise decoding approach that steers the decoding process towards producing correct reasoning steps. GRACE employs a discriminator model trained to differentiate correct from incorrect steps. The discriminator is used to adjust next-step likelihoods based on the correctness of each reasoning step. Importantly, GRACE only requires samples from the LM, and thus does not involve any extra training or fine-tuning. On six popular reasoning benchmarks (four math and two symbolic tasks), GRACE exhibits significant improvements in final answer accuracy compared to greedy decoding and self-consistency. In addition, human and LLM-based evaluations on GSM8K show that GRACE improves the correctness of the generated reasoning chains.

HI-TOM: A Benchmark for Evaluating Higher-Order Theory of Mind Reasoning in Large Language Models

Yinghui He,

Theory of Mind (ToM) is the ability to reasonabout one's own and others' mental states, and plays a critical role in the development of intelligence, language understanding, and cognitive processes. While previous work has primarily focused on first and second-order ToM, we explore higher-order ToM, which involves recursive reasoning on others' beliefs. We introduce HI-TOM, a Higher Order Theory of Mind benchmark. Our experimental evaluation using various Large Language Models (LLMs) indicates a decline in performance on higher- order ToM tasks, demonstrating the limitations of current LLMs. We conduct a thorough analysis of different failure cases of LLMs, and share our thoughts on the implications of our findings on the future of NLP.

Adaptive Endpointing with Deep Contextual Multi-Armed Bandits

Do June Min (University of Michigan), Andreas Stolcke (Amazon Alexa AI), Anirudh Raju (Amazon Alexa AI), Colin Vaz (Amazon Alexa AI), Di He (Amazon Alexa AI), Venkatesh Ravichandran (Amazon Alexa AI), Viet Anh Trinh (Amazon Alexa AI)

Current endpointing (EP) solutions learn in a supervised framework, which does not allow the model to incorporate feedback and improve in an online setting. Also, it is common practice to utilize costly grid-search to find the best configuration for an endpointing model. In this paper, we aim to provide a solution for adaptive endpointing by proposing an efficient method for choosing an optimal endpointing configuration given utterance-level audio features in an online setting, while avoiding hyperparameter grid-search. Our method does not require ground truth labels, and uses only online learning from reward signals. Specifically, we propose a deep contextual multi-armed bandit-based approach, combining the representational power of neural networks with the action exploration behavior of Thomp-son modeling algorithms. We compare our approach to several baselines, and show that our deep bandit models also succeed in reducing early cutoff errors while maintaining low latency.

enhancing long-form text generation in mental health with task-adaptive tokenization

Siyang Liu (University of Michigan), Naihao Deng (University of Michigan), Sahand Sabour (Tsinghua University), Minlie Huang (Tsinghua University), Rada Mihalcea (University of Michigan)

We propose task-adaptive tokenization as a way to adapt the generation pipeline to the specifics of a downstream task and enhance long-form generation in mental health. Inspired by insights from cognitive science, our task-adaptive tokenizer samples variable segmentations from multiple outcomes, with sampling probabilities optimized based on task-specific data. We introduce a strategy for building a specialized vocabulary and introduce a vocabulary merging protocol that allows for the integration of task-specific tokens into the pre-trained model's tokenization step. Through extensive experiments on psychological question-answering tasks in both Chinese and English, we find that our task-adaptive tokenization approach brings a significant improvement in generation performance while using up to 60\% fewer tokens. Preliminary experiments point to promising results when using our tokenization approach with very large language models.

Improving User Experience in Speech Recognition with Large Language Model

Liang-Yuan Wu (CSE, University of Michigan), Dhruv Jain (CSE, University of Michigan)

The development of Automatic Speech Recognition (ASR) makes it possible for machines to transcribe human speech into text and provides assistance in many aspects, especially when state-of-the-art ASR systems may outperform humans in ideal lab settings. However, in a real world scenario, we can find a huge space of improvement for ASR systems. Environment noise, speakers' accent, speaking styles, uncommon words are some of the reasons why ASR systems fail as a real-world application. Recently, large language models (LLMs) have shown unprecedented ability of language understanding, and we see the potential of LLMs to improve the quality of ASR system's output and to make useful real-world applications. With different prompting strategies, we demonstrate LLM's ability to improve the quality of speech recognition transcripts, and LLM's potential to be used in developing human-centered speech recognition systems.

Syntax-free Meaning Representation for Natural Language

Nasim Tohidi (K.N. Toosi University of Technology), Chitra Dadkhah (K.N. Toosi University of Technology)

Semantic representation in Natural Language Processing (NLP) is undergoing rapid changes. Recently there has been an increase in research focusing on models that introduce meaning structures independently from any syntactic or lexical criteria. One of the most used meaning representation models is Abstract Meaning Representation (AMR), which has its own weaknesses. Accordingly, in this paper, we propose Integrated Semantic Representation (ISR) model to annotate the semantic content of a text apart from the syntactic level. Unlike AMR, ISR is not dependent on any natural languages and can represent several language phenomena which have not been tackled in related semantic representation models, including AMR, such as co-reference, tense, quantification, aspect, and certainty. Plus, it can be applied in various NLP tasks which require semantic representation of text at sentence-level or document-level. For evaluating and checking the feasibility of applying the proposed model, a small dataset is produced, and a parser is implemented and applied on it.

Human-Computer Interaction

NLP Reproducibility For All: Understanding Experiences of Beginners

Shane Storks (University of Michigan), Keunwoo Peter Yu (University of Michigan), Ziqiao Ma (University of Michigan), Joyce Chai (University of Michigan)

As natural language processing (NLP) has recently seen an unprecedented level of excitement, and more people are eager to enter the field, it is unclear whether current research reproducibility efforts are sufficient for this group of beginners to apply the latest developments. To understand their needs, we conducted a study with 93 students in an introductory NLP course, where students reproduced the results of recent NLP papers. Surprisingly, we find that their programming skill and comprehension of research papers have a limited impact on their effort spent completing the exercise. Instead, we find accessibility efforts by research authors to be the key to success, including complete documentation, better coding practice, and easier access to data files. Going forward, we recommend that NLP researchers pay close attention to these simple aspects of open-sourcing their work, and use insights from beginners' feedback to provide actionable ideas on how to better support them.

BrushLens: Hardware Interaction Proxies for Accessible Touchscreen Interface Actuation

Chen Liang (University of Michigan), Yasha Iravantchi (University of Michigan), Thomas Krolikowski (University of Michigan), Ruijie Geng (University of Michigan), Alanson Sample (University of Michigan), Anhong Guo (University of Michigan)

Touchscreen devices, designed with an assumed range of user abilities and interaction patterns, often present challenges for individuals with diverse abilities to operate independently. Prior efforts to improve accessibility through tools or algorithms necessitated alterations to touchscreen hardware or software, making them inapplicable for the large number of existing legacy devices. In this paper, we introduce BrushLens, a hardware interaction proxy that performs physical interactions on behalf of users while allowing them to continue utilizing accessible interfaces, such as screenreaders and assistive touch on smartphones, for interface exploration and command input. BrushLens maintains an interface model for accurate target localization and utilizes exchangeable actuators for physical actuation across a variety of device types, effectively reducing user workload and minimizing the risk of mistouch. Our evaluations reveal that BrushLens lowers the mistouch rate and empowers visually and motor impaired users to interact with otherwise inaccessible physical touchscreens more effectively.

Deploying VizLens: Characterizing User Needs, Preferences, and Challenges of Physical Interfaces Usage in the Wild

Andi Xu (University of Michigan, Ann Arbor, Michigan, United States) Mahdi Qazwini (University of Michigan, Ann Arbor, Michigan, United States) Chen Liang (University of Michigan, Ann Arbor, Michigan, United States) Anhong Guo (University of Michigan, Ann Arbor, Michigan, United States)

Blind or Visually Impaired (BVI) people often encounter flat, inaccessible interfaces. Current solutions lack cost-effectiveness, portability, and robustness in real-world settings. We introduce VizLens, a fully-automated, full-stack mobile application powered by computer vision algorithms. The system is deployed and publicly available through the Apple App Store (https://vizlens.org/). From May to August 2023, we had 665 users, who uploaded 1,320 interface images. We aim to use it to study usage patterns and possible challenges BVI users may encounter with flat interfaces through a large-scale study in real-world settings. With in-depth analysis of user data and activity logs, our study will provide insights into BVI users' interface interests, preferred assistance modes, and potential challenges due to system limitations or users' diverse abilities. Our goal is to enhance the understanding of how BVI users interact with inaccessible, flat interfaces, and inform future assistive technology design.

ImageExplorer Deployment: Understanding Text-Based and Touch-Based Image Exploration in the Wild

Andi Xu (University of Michigan, Ann Arbor, Ann Arbor, Michigan, United States), Minyu Cai (University of Michigan, Ann Arbor, Michigan, United States), Dier Hou (University of Michigan, Ann Arbor, Ann Arbor, Michigan, United States), Ruei-Che Chang (University of Michigan, Ann Arbor, Michigan, United States), Anhong Guo (University of Michigan, Ann Arbor, Michigan, United States)

Blind and visually-impaired (BVI) users often rely on alt-texts to understand digital images. AI-generated alt-texts can be scalable and efficient but may lack details and are prone to errors. Multi-layered touch interfaces, on the other hand, can provide rich details and spatial information, but may take longer to explore and cause higher mental load. To understand how BVI users leverage these two methods, we deployed ImageExplorer, an iOS app on the Apple App Store that provides multi-layered image information via both text-based and touchable interfaces with customizable levels of granularity. From March to August 2023, 283 users uploaded 518 images and explored 654 times. We collected user activity logs to understand their image reading patterns and preferences in the wild. This work informs a holistic understanding of BVI users' image exploration behavior and influential factors. We provide design implications for future image caption models and visual access tools.

Discovering the Right Things to Design with Artificial Intelligence

Nur Yildirim (Carnegie Mellon University), James McCann (Carnegie Mellon University), John Zimmerman (Carnegie Mellon University)

Advances in artificial intelligence have enabled unprecedented technical capabilities, yet making these advances useful in the real world remains challenging. Today, most AI projects fail, often because product teams select suboptimal places to apply AI. I argue that the current process for designing AI products and services is broken. In this talk, I present a new innovation process that helps teams identify low risk, high value use cases where moderate AI performance can create value for people. Through case studies, I highlight the vital role human-computer interaction research plays in finding applications where humans can benefit and thrive with AI.

Show, Not Tell: A Pattern-Based, Deaf-Centric Classification Approach for Everyday Sounds

Jeremy Zhengqi Huang (University of Michigan), Reyna Wood (University of Michigan), Hriday Chhabria (University of Michigan), Dhruv Jain (University of Michigan)

Current sound recognition systems for deaf or hard of hearing (DHH) people identify sound sources (e.g., dishwasher) or discrete events (e.g., door knocks). However, since different sources can produce similar sounds (e.g., both washing machine and dishwasher produce a “hum”), the categories often overlap. We introduce a novel approach to categorizing sounds based on their underlying sound patterns. To ensure our classification approach aligns with the Deaf-centered practices of describing sounds, we interviewed eight ASL interpreters on how they sign different sounds. Through cluster analysis of the interpreter responses, we arrived at an 18-category taxonomy that distinguishes sound patterns based on their ASL descriptions. We evaluated our taxonomy with nine DHH people, finding the initial promise of inferring sound events and other sound cues from the signing patterns. We also used our taxonomy to train a sound recognition model, revealing a near-perfect classification accuracy on a small dataset.

Human-AI Collaboration in Medical Diagnosis: Enhancing Acute Respiratory Distress Syndrome Detection

Negar Farzaneh (Weil Institute for Critical Care Research & Innovation, Department of Emergency Medicine); Sardar Ansari (Weil Institute for Critical Care Research & Innovation, Department of Emergency Medicine); Elizabeth Lee (Department of Radiology); Kevin R. Ward (Weil Institute for Critical Care Research & Innovation, Department of Emergency Medicine, Department of Biomedical Engineering); Michael W. Sjoding (Weil Institute for Critical Care Research & Innovation, Department of Internal Medicine, Division of Pulmonary and Critical Care)

There is a growing gap between studies describing the capabilities of artificial intelligence (AI) diagnostic systems using deep learning versus efforts to investigate how or when to integrate AI systems into a real-world clinical practice to support physicians and improve diagnosis. To address this gap, we investigate potential strategies for AI model deployment and physician collaboration to determine their potential impact on diagnostic accuracy. As a case study, we examine an AI model trained to identify findings of the acute respiratory distress syndrome (ARDS) on chest X-ray images. While this model outperforms physicians at identifying findings of ARDS, there are several reasons why fully automated ARDS detection may not be optimal nor feasible in practice. Among several collaboration strategies tested, we find that if the AI model first reviews the chest X-ray and defers to a physician if it is uncertain, this strategy achieves a higher diagnostic accuracy (0.869, 95% CI 0.835–0.903) compared to a strategy where a physician reviews a chest X-ray first and defers to an AI model if uncertain (0.824, 95% CI 0.781–0.862), or strategies where the physician reviews the chest X-ray alone (0.808, 95% CI 0.767–0.85) or the AI model reviews the chest X-ray alone (0.847, 95% CI 0.806–0.887). If the AI model reviews a chest X-ray first, this allows the AI system to make decisions for up to 79% of cases, letting physicians focus on the most challenging subsets of chest X-rays.

Personalized Preference-Bandits: Learning to make customized user predictions through preference elicitation

Aadirupa Saha (Apple)

Customer statistics collected in several real-world systems have reflected that users often prefer eliciting their liking for a given pair of items, say (A,B), in terms of relative queries like: "Do you prefer Item A over B?", rather than their absolute counterparts: "How much do you score items A and B on a scale of [0-10]?". Drawing inspirations, in the search for a more effective feedback collection mechanism, led to the famous formulation of Dueling Bandits (DB), which is a widely studied online learning framework for efficient information aggregation from relative/comparative feedback. However despite the novel objective, unfortunately, most of the existing DB techniques were limited only to simpler settings of finite decision spaces, and stochastic environments, which are unrealistic in practice. We will see the basic problem formulations for DB and familiarize ourselves with some of the breakthrough results. Following this, will dive deep into a more practical framework of contextual dueling bandits (C-DB) where the goal of the learner is to make customized predictions based on the user contexts: We will see a new algorithmic approach that can efficiently achieve the optimal O(sqrt T) regret performance for this problem, resolving an open problem from Dudík et al. [COLT, 2015]. We will conclude with some interesting open problems.

"DIY Assistive Software: Creating Personalized AI Assistive Technology Through End-User Programming"

Jaylin Herskovitz (University of Michigan, CSE), Andi Xu (University of Michigan, CSE), Rahaf Alharbi (University of Michigan, SI), Anhong Guo (University of Michigan, CSE)

Existing AI-based assistive technologies often fail to support the unique and personalized needs of blind and visually impaired (BVI) people. Despite BVI people often being early adopters of AI technologies, AI models are trained for limited types of scenarios. Consequently, AI-based assistive technologies tend to assume 'universal' needs of BVI people and are thus one-size-fits-all, rather than accounting for unique differences and desires. In my research, I aim to support Do-It-Yourself (DIY) assistive software design and development, so that BVI people can design, create, and customize assistive software for themselves and their communities. My work investigates end-user programming as an approach to support people in DIY-ing assistive technologies, enabling anyone to participate in AI creation, overcoming both skill and accessibility barriers to do so. By enabling detailed customization, control, and understanding of AI technologies, my research aims to improve how AI technologies are leveraged and interacted with in order to improve assistive applications.

Customizable Interactive Systems for Accessible Immersion

Ruei-Che Chang (UMich CSE), Anhong Guo (UMich CSE)

Users may have varied needs and goals when using a system. Therefore, creating customizable interactive systems capable of grounding end users' needs and adapting to their context is important. This is especially important when designing systems for people with disabilities who have different abilities and usage contexts. In this poster, I will present the works I did for designing customizable interactive system for people who are blind or have low vision, which includes OmniScribe to author spatial audio descriptions for 360-degree videos, Sound Blending to customize audio feedback to enhance perceptions on sounds, and OmniCap to customize captions on understand omnidirectional visual information. In the future, we aim to explore ways to enable customizations on different sensory information (e.g., auditory, visual, etc) to reduce users' cognitive load and improve their perceptions on tasks.

Making Authoring Tools for Immersive Content More Expressive, Collaborative, and Intelligent

Lei Zhang (University of Michigan)

The recent advances of augmented, virtual, and mixed reality (XR) necessitate the development of authoring tools to effectively create XR content. Immersive authoring is a paradigm that aims to lower the barrier of entry by allowing end-users to create XR content directly within the immersive environment. I seek to make immersive authoring tools more powerful for creating XR content by answering the following research questions: How can we raise the ceiling of expressiveness of immersive authoring tools to create complex XR interactions? How can we design immersive authoring tools to foster meaningful social interactions? How can we make immersive authoring tools more intelligent to facilitate the collaboration between users and the diversity of content? In my work, I build and assess innovative immersive authoring systems, providing insights into the capabilities and limitations of these tools. By unlocking the potential of immersive authoring, my research endeavors to democratize and empower the creation practices surrounding XR content, fostering greater expressiveness, collaboration, and intelligence.

Computer Vision & Multimodal Perception

Automating the detection and classification of marine wildlife in aerial imagery

Kyle Landolt (USGS), Timothy White (BOEM), Mark Koneff (USFWS), Brad Pickens (USFWS), Aaron Murphy (USGS), Matthew Walker (USGS), Jennifer Dieck (USGS), Luke Fara (USGS), Dave Fronczak (USFWS), Tsung-Wei Ke (CMU), Stella Yu (UMich)

Avian and wildlife population surveys can help inform environmental assessments, and impact analyses of offshore energy development projects. Low-flying ocular aerial surveys have historically been used to estimate waterfowl populations, but place personnel at risk of injury and survey results are prone to bias and misclassification. The U.S. Geological Survey (USGS), in collaboration with the Bureau of Ocean Energy Management (BOEM), the U.S. Fish and Wildlife Service Division of Migratory Bird Management (USFWS-DMBM), and the University of Michigan, Ann Arbor, is advancing the development of deep learning algorithms and tools to automate the detection, enumeration, and classification of seabirds, waterfowl, and other marine wildlife. Aerial imagery collected from the Atlantic Outer Continental Shelf and the Great Lakes provide data for algorithm development. OpenCV's Computer Vision Annotation Tool (CVAT) is providing the framework for an interactive GUI, allowing wildlife experts to efficiently create annotations and support database development. Our research expands on the utilization of object detection and classification algorithms to generate population estimates of wildlife.

Fully Automated Pipeline for Measurement of the Thoracic Aorta Using Joint Segmentation and Localization Neural Network

Sudeep Katakola (University of Michigan EECS), Timothy J. Baker (University of Michigan Radiology), Zhangxing Bianc (Johns Hopkins ECE), Yanglong Lub (University of Michigan Radiology), Greg Spahlinger (University of Michigan Radiology), Charles R. Hatt (Imbio Inc.), Nicholas S. Burris (University of Michigan Radiology)

Diagnosis and surveillance of thoracic aortic aneurysm (TAA) involves measuring the aortic diameter at various locations along the length of the aorta, often using computed tomography angiography (CTA). Currently, measurements are performed by human raters using specialized software for 3D analysis, a process that requires 15 to 45 minutes of focused effort. In this work, we develop a convolutional neural network (CNN)-based algorithm that fully automates this time-consuming process. We use multi-task learning to train a CNN to perform joint segmentation and localization of key aortic landmarks. The segmentation mask and landmarks are subsequently used to obtain the centerline and cross-sectional diameters. Compared to single task learning, joint training yielded higher accuracy for segmentation, especially at the aortic boundary which is critical to accurate diameter measurement. The automated diameter measurements were compared with measurements performed manually by experts and we find that mean absolute error was less than 1 mm at most measurement locations. We investigated cases of high error and implemented corrections to improve the reliability of the automated pipeline. Overall, we find that fully automated aortic diameter measurements in TAA are feasible using a CNN-based algorithm that can save significant amounts of time compared to manual measurement.

Detection-Based State Estimation for Instructional Video Understanding

Wuao Liu (University of Michigan, Ann Arbor, Robotics Department), Jason Corso(University of Michigan, Ann Arbor, Robotics Department)

Virtual reality (VR) and artificial intelligence (AI) are redefining the way users approach complex physical tasks, exponentially expanding their skillsets and reducing errors. In this study, we dive into a unique application of these technologies: the state estimation in instructional videos. Focusing on diverse kitchen environments, our approach harnesses the power of an object detector-based method. We utilized Faster-RCNN, a well-established object detector, and achieved a promising 93.696 mAP in identifying generic kitchen tools and ingredients in our uniquely curated video collection. To translate this raw data into actionable insights, a heuristic-based encoding method was employed. This method effectively grounds the detected information, allowing for the extraction of informative feature maps. In essence, our research presents a pioneering effort to marry VR and AI with video analysis, opening new avenues for training and skill acquisition.

Post-training Quantization on Diffusion Models

Illinois Institute of Technology

Denoising diffusion (score-based) generative models have recently achieved significant accomplishments in generating realistic and diverse data. These approaches define a forward diffusion process for transforming data into noise and a backward denoising process for sampling data from noise. Unfortunately, the generation process of current denoising diffusion models is notoriously slow due to the lengthy iterative noise estimations, which rely on cumbersome neural networks. It prevents the diffusion models from being widely deployed, especially on edge devices. Previous works accelerate the generation process of diffusion model (DM) via finding shorter yet effective sampling trajectories. However, they overlook the cost of noise estimation with a heavy network in every iteration. In this work, we accelerate generation from the perspective of compressing the noise estimation network. Due to the difficulty of retraining DMs, we exclude mainstream training-aware compression paradigms and introduce post-training quantization (PTQ) into DM acceleration. However, the output distributions of noise estimation networks change with time-step, making previous PTQ methods fail in DMs since they are designed for single-time step scenarios. To devise a DM-specific PTQ method, we explore PTQ on DM in three aspects: quantized operations, calibration dataset, and calibration metric. We summarize and use several observations derived from all-inclusive investigations to formulate our method, which especially targets the unique multi-time-step structure of DMs. Experimentally, our method can directly quantize full-precision DMs into 8-bit models while maintaining or even improving their performance in a training-free manner. Importantly, our method can serve as a plug-and-play module on other fast-sampling methods, e.g., DDIM.

Language-Guided Pose Forecasting

Yayuan Li (ECE, University of Michigan)

We firstly propose to forecast poses from visual input instructed by language description. This task aims at generating reasonable motion of the target in a controllable way with natural language as the interface. It is interesting and useful for real-world interactive applications like AR/VR and robotics. To tackle this problem, we design a transformer encoder-decoder architecture. The encoder captures visual information and language guidance jointly while the decoder is trained to forecast future poses based on the encoded vision and language conditions. To evaluate the performance in diverse use cases, we use three datasets across two targets (human and hand), two viewpoints (egocentric and exocentric), and multiple domains (cooking, sports and etc.). Our extensive experiments show that our method successfully forecasts poses that aligns with language description in various use cases. We also demonstrate the ability of our method to deal with open-vocabulary guidance which enables much more flexible usage in real-world applications. Besides, we locate this work in pre-recognition research community and compare the state-of-the-art pose/trajectory forecasting method that only takes visual information as input. Our method outperforms the SOTAs by a large margin which proves the effectiveness of using language. We do ablation studies that show the effectiveness of each module.

An improved perception model for traffic light handling through the integration of deep learning and V2X infrastructure.

Daphne Tsai, University of Michigan Department of Computer Science and Engineering

An accurate and robust traffic light handling model is crucial for the safe and efficient operation of autonomous vehicles. Current autonomous driving stacks leverage Vehicle-to-Everything (V2X) infrastructure in the form of Signal Phase and Timing (SPaT) messages to receive crucial information for path planning. However, thus has several limitations, such as missing or incorrect SPaT messages. At the same time, camera-based perception systems suffer from computing power limitations, detection from far distances, and operation in bad weather. To create an improved perception model that can handle traffic light predictions more accurately while remaining less susceptible to external factors, input from V2X and cameras was fused together, and a deep learning model was developed that achieved high accuracy on proprietary testing data. Furthermore, a control flow logic was created to evaluate the perception model under both ideal and realistic conditions for accuracy and efficiency. As a result, a more robust perception model was created for traffic light handling by integrating deep learning with existing V2X infrastructure, which exploits the strengths of both types of technologies while mutually compensating for their weaknesses.

Towards A Richer 2D Understanding of Hands at Scale

Tianyi Cheng (University of Michigan), Dandan Shan(University of Michigan), Ayda Sultan Hassen(Addis Ababa Institute of Technology), Richard Ely Locke Higgins(University of Michigan),, David Fouhey (New York University)

As humans, we learn a lot about how to interact with the world by observing others interacting with their hands. To help AI systems obtain a better understanding of hand interactions, we introduce a new model that produces a rich understanding of hand interaction. Our system produces a richer output than past systems at a larger scale. Our outputs include boxes and segments for hands, in-contact objects, and second objects touched by tools as well as contact and grasp type. Supporting this method are annotations of 257K images, 401K hands, 288K objects, and 19K second objects spanning four datasets. We show that our method provides rich information and performs and generalizes well.

Machine Learning

Exploiting Second-order Information in Conjunction with Variance Reduction Promotes Robustness

Sachin Garg(PhD Student CSE University of Michigan), Michal Derezinski (Assistant Professor CSE University of Michigan), Albert S. Berahas (Assistant Professor IOE University of Michigan)

Incorporating variance-reduced techniques accelerates the convergence of stochastic first-order methods, providing global linear convergence for smooth and strongly convex functions. However, the convergence rate of first-order variance-reduced stochastic methods deteriorates with an increase in the mini-batch sample size used for the stochastic gradient. Moreover, the literature on the benefits of variance reduction techniques to accelerate the popular second-order methods is limited and requires further investigation. In this work, we propose a stochastic second-order method with variance-reduced gradients which achieves the same fast convergence regardless of the mini-batch size, and takes advantage of variance reduction to provide improved local convergence rates (per data pass) compared to both first-order stochastic variance-reduced methods and stochastic second-order methods. Specifically, we prove that in the big data regime where the number of components n is much larger than the condition number ϰ, our algorithm achieves local linear convergence rate (Cϰn log(n/ϰ))^t with high probability after t iterations, independent of the choice of the mini-batch size. We provide empirical evidence that the proposed method achieves a faster convergence rate, and is more robust to the noise in the stochastic Hessian and step-size.

Survival Analysis with Multiple Noisy Labels: Formalization of a Novel Problem Setting

Donna Tjandra (University of Michigan), Jenna Wiens (University of Michigan)

Labeling disease onset in electronic health record (EHR) data often requires manual chart review, which can be infeasible in large datasets. Instead, pragmatic labeling tools based on heuristics are often used (e.g., patients billed for the disease are positive), at the potential cost of introducing noise (e.g., incorrectly identifying time-to-events [TTEs] for survival analysis). When multiple, different labeling tools are used to identify the TTE for a condition of interest, we find ourselves in the setting of multiple noisy labels. Here, the term `multiple' refers to the different labeling tools, and the term `noisy' refers to the potential inaccuracies each tool introduces. Previous work studying multiple noisy labels focuses on classification and proposes different strategies to aggregate the labels. However, survival analysis presents novel challenges to the multiple noisy labels setting since data may be censored. That is, individuals may be lost to followup such that not all labelers have the opportunity to assign an outcome. As a result, aggregation from past work does not apply. Here, we introduce the problem of multiple noisy labels in survival analysis and explore the shortcomings of approaches from classification in the 0-mean noise setting (i.e., the expected average of the complete [i.e., uncensored] set of observed TTEs is the ground truth TTE). When the rate of censorship is high or when the label noise is instance-dependent, we show that existing approaches perform poorly, highlighting the need for new approaches for survival analysis with multiple noisy labels.

Counterfactual-Augmented Importance Sampling for Semi-Offline Policy Evaluation

Shengpu Tang (UM-CSE), Jenna Wiens (UM-CSE)

In applying reinforcement learning (RL) to high-stakes domains, quantitative and qualitative evaluation using observational data can help practitioners understand the generalization performance of new policies. However, this type of off-policy evaluation (OPE) is inherently limited since offline data may not reflect the distribution shifts resulting from the application of new policies. On the other hand, online evaluation by collecting rollouts according to the new policy is often infeasible, as deploying new policies in these domains can be unsafe. In this work, we propose a semi-offline evaluation framework as an intermediate step between offline and online evaluation, where human users provide annotations of unobserved counterfactual trajectories. While tempting to simply augment existing data with these annotations during evaluation, as we show, doing so can lead to biased results. Thus, we design a new family of OPE estimators based on a novel weighting scheme and importance sampling (IS) to incorporate counterfactual annotations without introducing additional bias. We analyze the theoretical properties of our approach, showing its potential to reduce both bias and variance compared to standard IS estimators. In a series of proof-of-concept experiments involving bandits and a healthcare-inspired simulator, we demonstrate that our approach outperforms purely offline IS estimators and is robust to noise and missingness of the annotations. Our framework, combined with principled human-centered design of annotation solicitation, can enable practical RL in high-stakes domains.

Neural Caches for Monte Carlo Partial Differential Equation Solver

Zilu Li (Cornell), Guandao Yang (Cornell), Xi Deng (Cornell), Christopher De Sa (Cornell), Bharath Hariharan (Cornell), Steve Marschner (Cornell)

This paper presents a method that uses neural networks as a caching mechanism to reduce the variance of Monte Carlo Partial Differential Equation solvers, such as the Walk-on-Spheres algorithm [Sawhney and Crane 2020]. While these Monte Carlo PDE solvers have the merits of being unbiased and discretization-free, their high variance often hinders real-time applications. On the other hand, neural networks can approximate the PDE solution, and evaluating these networks at inference time can be very fast. However, neural-network-based solutions may suffer from convergence difficulties and high bias. Our hybrid system aims to combine these two potentially complementary solutions by training a neural field to approximate the PDE solution using supervision from a WoS solver. This neural field is then used as a cache in the WoS solver to reduce variance during inference. We demonstrate that our neural field training procedure is better than the commonly used self-supervised objectives in the literature. We also show that our hybrid solver exhibits lower variance than WoS with the same computational budget: it is significantly better for small compute budgets and provides smaller improvements for larger budgets, reaching the same performance as WoS in the limit.

Federated Adversarial Learning: A Framework with Convergence Analysis

Jiaming Yang (UMich), Xiaoxiao Li (UBC), Zhao Song (Adobe)

Federated learning (FL) is a trending training paradigm to utilize decentralized training data. FL allows clients to update model parameters locally for several epochs, then share them to a global model for aggregation. This training paradigm with multi-local step updating before aggregation exposes unique vulnerabilities to adversarial attacks. Adversarial training is a popular and effective method to improve the robustness of networks against adversaries. In this work, we formulate a general form of federated adversarial learning (FAL) that is adapted from adversarial learning in the centralized setting. On the client side of FL training, FAL has an inner loop to generate adversarial samples for adversarial training and an outer loop to update local model parameters. On the server side, FAL aggregates local model updates and broadcast the aggregated model. We design a global robust training loss and formulate FAL training as a min-max optimization problem. Unlike the convergence analysis in classical centralized training that relies on the gradient direction, it is significantly harder to analyze the convergence in FAL for three reasons: 1) the complexity of min-max optimization, 2) model not updating in the gradient direction due to the multi-local updates on the client-side before aggregation and 3) inter-client heterogeneity. We address these challenges by using appropriate gradient approximation and coupling techniques and present the convergence analysis in the over-parameterized regime. Our main result theoretically shows that the minimum loss under our algorithm can converge to ϵ small with chosen learning rate and communication rounds. It is noteworthy that our analysis is feasible for non-IID clients.

Applications of AI

OptoGPT: A Foundation Model for Inverse Design in Multilayer Thin Film

Taigao Ma (Physics, Umich), Haozhu Wang (EECS, Umich), L. Jay Guo (EECS, Umich)

Existing inverse design methods for multilayer thin films either fail to explore the global design space (material & thickness) or suffer from low computational efficiency, making them unsuitable for universal design applications. To bridge this gap, we propose the Opto Generative Pretrained Transformer (OptoGPT), which is a decoder-only transformer that auto-regressively generates designs based on specific spectrum target inputs. Trained on a large dataset, our model demonstrates: 1) autonomous global design by determining the number of layers (up to 20) while selecting the material (up to 18 distinct types) and thickness at each layer, 2) efficient designs for structural color, absorbers, filters, distributed brag reflectors, and Fabry–Pérot resonators within 0.1 seconds (comparable to simulation speeds), 3) the ability to output diverse designs, and 4) seamless integration of user-defined constraints. By overcoming design barriers regarding optical targets, material selections, and design constraints, OptoGPT can serve as a foundation model for universal optical multilayer thin film structure inverse design.

Predicting Neuron Morphology from Single-Cell Gene Expression using Deep Generative Models

Hojae Lee (ECE), Joshua Welch (DCMB, CSE)

Gene expression and morphology both play a key role in determining the types and functions of cells, but the relationship between molecular and morphological features is largely uncharacterized. We present MorphNet, a computational approach that can draw pictures of a cell's morphology from its gene expression profile. Our approach leverages paired morphology and molecular data to train a neural network that can predict nuclear or whole-cell morphology from gene expression. We employ state-of-the-art data augmentation techniques that allow training using as few as 1000 images. We find that MorphNet can generate novel, realistic morphological images that retain the complex relationship between gene expression and cell appearance. We then train MorphNet to generate nuclear morphology from gene expression using brain-wide MERFISH data. In addition, we show that MorphNet can generate neuron morphologies with realistic axonal and dendritic structures. MorphNet generalizes to unseen brain regions, allowing prediction of neuron morphologies across the entire mouse isocortex and even non-cortical regions. Using MorphNet, we predicted morphologies for striatal interneurons from gene expression alone and validated our predictions using fMOST reconstructions. Additionally, MorphNet performs meaningful latent space interpolation, allowing prediction of the effects of gene expression variation on morphology, which are confirmed by gradients in real neuron morphologies. Finally, we provide a web server that allows users to predict morphologies for their own scRNA-seq data. MorphNet represents a powerful new approach for linking gene expression and morphology at cellular resolution.

Analyzing High School (9-12th Grade) AP Scores Using Machine Learning

Saatvik Barla (Frisco High School)

Advanced Placement (AP) exams are tests students can take to earn college credit while still in high school. A multitude of factors influence student pass rates regarding these exams. This study analyzes the influence of socioeconomic status, race, and English language abilities – among other factors – on AP exam scores by implementing machine learning technology. This study used five machine learning models, namely, Linear Regression, Support Vector Machine (SVM), Random Forest, Adaptive Boosting (AdaBoost), and XG Boost, to predict AP exam scores based on the aforementioned factors. This study then used SHapley Additive exPlanations (SHAP) to analyze the models and decipher what factors the machine learning models value most when predicting AP exam scores. This study concluded that being economically disadvantaged or a non-native English language speaker (Bil/ESL) negatively impacted AP exam scores; all five models valued these two factors highest when making predictions on student AP exam performance. These findings corroborate similar studies, like Clark et al. (2018), that suggest negative socioeconomic factors have a strong negative influence on exam scores. Further research should be conducted to analyze whether providing economically disadvantaged and/or Bil/ESL students with additional resources and attention positively influences their AP exam scores. Through these findings, therefore, school administrations might be able to take preemptive action to improve AP exam scores amongst economically disadvantaged and non-native English speaking students.

A sociotechnical approach to algorithmic audit of targeted advertising

Lu Xian (School of Information), Matt Bui (School of Information), Abigail Jacobs (School of Information)

Targeted ads are important sources of information that shape how individuals get access to various opportunities and resources. Targeted advertising also introduces bias and discrimination against groups of users along the lines of race, gender, and so on. Audits of targeted advertising algorithms in various domains like employment have provided evidence for gender-based and race-based discrimination in ad delivery by focusing on the association between user profile and ad type and content. Building on prior work, our work provides an account for how biases reflected in targeted advertising reinforce historical inequalities in the housing and mortgage sectors in metropolitan cities. To demonstrate, we collected web traffic data within Google's search engine results page at the zip code level in New York City in 2020 from a third party vendor. We combined the data with recently released appraisal record data at the census tract level, which serves as a proxy for home values. By triangulating those datasets with census data, we analyze how communities are differentially targeted using spatial analysis, clustering, and regression models. This work reveals how biased distribution of information and resources impact individuals' access to economic opportunities.

e-HAIL: Making the University of Michigan a Premier Hub for e-Health and AI to Improve Health Using Technology

Henrike Florusbosch, PhD (e-HAIL, MM and CoE); Rada Mihalcea, PhD (CSE, CoE); Akbar Waljee, MD (DLHS, MM)

e-HAIL is a joint Michigan Medicine and College of Engineering initiative that aims to make U-M a premier hub for research that innovates in health through AI. Our focus is on collaboration, grant development, and infrastructure to support a multi-disciplinary approach to innovations in healthcare and AI/ML methodologies. The initiative provides several opportunities for faculty researchers from MM and CoE to connect and learn from others operating at the intersection of AI and health. We provide concrete support for preparing grant applications, including finding collaborators, grant writing assistance, software development, data set creation, and hiring student support.

Targeting C. difficile Infection Prevention Efforts with AI: A Pre-Post Study at Michigan Medicine

Shengpu Tang (UMich), Erkin Ötleş (UMich), Stephanie Shepard (UMich), Rebekah Clark (UMich), Anastasia Wasylyshyn (UMich), Ji Hoon Baang (UMich), Paul Grant (UMich), Krishna Rao (UMich), Jenna Wiens (UMich)

We developed an institution-specific daily risk prediction model which leverages electronic health record data to identify patients at high risk for developing C. difficile infection (CDI). At Michigan Medicine, the model achieved an AUROC of 0.778 (95% CI, 0.744 - 0.814) and 0.767 (95% CI, 0.737 - 0.801) in retrospective and prospective cohorts, respectively. Working closely with various stakeholders we designed and launched an intervention bundle that incorporates scores from our daily risk prediction model and aims to reduce patients' exposure to C. difficile spores and susceptibility to infection. The intervention bundle consists of two Best Practice Advisory (BPA) alerts that will fire when a patient is flagged by our model as being at high risk for developing CDI, which based on simulation results we defined as above the 91st percentile of model scores. We are currently running a year-long pre-post study to understand the effect of this AI-driven intervention strategy on overall CDI incidence as well as prescribing practices of antimicrobial agents that contribute to CDI risk.

Deep Learning Tools for Large-Scale Multi-Modal Neural Data

Lu Mi (Allen Institute for Brain Science, University of Washington)

Recent years have marked notable advancements in neuroscience, largely attributable to the development of advanced tools enabling more detailed studies of the brain. However, despite such advancements, our accessibility and comprehension of the brain's intricacies are still in nascent stages. My research primarily aims to expand our abilities to collect and interpret large-scale, multi-modal neural data - including anatomical structure, functional activity, transcriptomics, behaviors - by leveraging cutting-edge computer vision and machine learning methods. By enhancing the accuracy, efficiency, and scalability of acquisition and analysis workflows for neuroscience studies, we aspire to expedite scientific discoveries in unraveling of the mysteries of coding, computation, and learning processes in the brain. Furthermore, this research paves the way for developing innovative brain-inspired AI frameworks, potentially closing the loop between artificial and natural intelligence.

DeepLig: A de-novo Computational Drug Design Approach to Generate Multi-Targeted Drugs

Anika Chebrolu (Department of Biomedical Engineering, University of North Texas, Denton, Texas, USA)

Mono-targeted drugs can be of limited efficacy against complex diseases. Recently, multi-target drug design has been approached as a promising tool to fight against these challenging diseases. However, the scope of current computational approaches for multi-target drug design is limited. DeepLig presents a de-novo drug discovery platform that uses reinforcement learning to generate and optimize novel, potent, and multitargeted drug candidates against protein targets. DeepLig’s model consists of two networks in interplay: a generative network and a predictive network. The generative network, a Stack-Augmented Recurrent Neural Network, utilizes a stack memory unit to remember and recognize molecular patterns when generating novel ligands from scratch. The generative network passes each newly created ligand to the predictive network, which then uses multiple Graph Attention Networks simultaneously to forecast the average binding affinity of the generated ligand towards multiple target proteins. With each iteration, given feedback from the predictive network, the generative network learns to optimize itself to create molecules with a higher average binding affinity towards multiple proteins. DeepLig was evaluated based on its ability to generate multi-target ligands against two distinct proteins, multi-target ligands against three distinct proteins, and multi-target ligands against two distinct binding pockets on the same protein. With each test case, DeepLig was able to create a library of valid, synthetically accessible, and novel molecules with optimal binding energies. We propose that DeepLig provides an effective approach to design multi-targeted drug therapies that can potentially show higher success rates during in-vitro trials.

Can Large Language Models Reason About Goal-Oriented Tasks?

Filippos Bellos (University of Michigan), Yayuan Li (University of Michigan), Wuao Liu (University of Michigan), Jason Corso (University of Michigan)

Most adults can complete a sequence of steps to achieve a certain goal, such as making a sandwich or repairing a bicycle tire. In completing these goal-oriented tasks, or simply tasks in this paper, one must use sequential reasoning to understand the relationship between the sequence of steps and the goal. LLMs have shown impressive capabilities across various natu- ral language understanding tasks. However, how well LLMs can perform this sequential reasoning is not clear. In this paper, we address this gap and conduct a comprehensive evaluation of how well LLMs are able to conduct this reasoning for tasks. Using adapted subsets of the YouCook2 and CrossTask datasets, with varying levels of sequence permutations (33.3%, 50%), we aimed to understand how LLMs responded to disrup- tions in order. Considering the comprehensiveness of evalua- tions, we included three representative LLMs (GPT 3.5-turbo, GPT-4, and Llama 2-13b), and evaluated them using different prompt configurations. A key aspect of our approach was to assess the capability of LLMs to identify transitions between steps in these permuted goal-oriented tasks but also reason about the permuted tasks as a whole determining their viability. Our results reveal that GPT- 4 emerged as the most proficient in sequential reasoning tasks, demonstrating a consistent ability to grasp task objectives and logical progression of steps. Our analysis also indicates that the performance of LLMs, is significantly influenced by the structural integrity of the input they receive. Finally, central to our findings is the pivotal role that prompt design plays in unlocking the reasoning capabilities of these models, underlin- ing a critical connection between prompt structure and model performance in sequential reasoning tasks.