Neeraj Varshney

Ph.D. Candidate (4th Year)
Computer Science (NLP/NLU)
Arizona State University

Email: nvarshn2@asu.edu
Website: nrjvarshney.github.io
Semantic Scholar: Neeraj-Varshney
LinkedIn: neerajvarshney97

RESEARCH STATEMENT

I work in Natural Language Processing /Understanding, primarily in the areas of Open-domain Question Answering, Large Language Models, Natural Language Inference, Retrieval augmented inference, and addressing different Efficiency & Reliability aspects of NLP systems.

On the efficiency topic, I have worked on improving the Inference Efficiency, Training Sample Efficiency, QA Reader Efficiency, Knowledge Indexing Efficiency, and Evaluation Efficiency.

On the reliability topic, I have worked on Selective Prediction, Post-Abstention, Calibration, and Mitigating Hallucinations of LLMs.

Publication Venues: ACL   EMNLP   EACL   NAACL   AAAI   AAMAS.

Research Interests: Efficiency and Reliability aspects of NLP Systems
Open-domain Question Answering
Mitigating Hallucinations of LLMs
Retriever-Reader Systems
Robustness and Generalization
Selective Prediction / Abstention
Multi-task Learning
Learning From Natural Language Instructions
Learning From Less Supervision

Thesis Committee: Dr. Chitta Baral (Chair)
Dr. Yezhou Yang
Dr. Nakul Gopalan
Dr. Pratyay Banerjee

TECHNICAL SKILLS

Languages	:	Python, Java, C, C++
Frameworks	:	PyTorch, PyTorch-lightning
Libraries	:	Huggingface Transformers, Spacy, Pyserini, NumPy, Matplotlib, Pandas, NLTK, Word2vec
Tools	:	Git, Linux, Amazon Mechanical Turk, PyCharm, Jupyter, Colab, MS Office

SELECTED PROJECTS

Detecting and Mitigating Hallucinations of LLMs by Validating Low-Confidence Generation   Preprint

N. Varshney, W. Yao, H. Zhang, J. Chen, D. Yu

- Addressing the critical problem of LLM hallucinations, we propose an approach that actively detects and mitigates hallucinations during the generation process.
- In our approach, we first identify the candidates of potential hallucination leveraging the model's logit output values, check their correctness through a validation procedure, mitigate the detected hallucinations, and then continue with the generation process.
- Through extensive experiments with the 'article generation task', we first demonstrate the individual efficacy of our detection and mitigation techniques. Specifically, the detection technique achieves a recall of 88% and the mitigation technique successfully mitigates 57.6% of the correctly detected hallucinations.
- Then, we show that the proposed active detection and mitigation approach successfully reduces the hallucinations of the GPT-3 model from 47.5% to 14.5% on average.

On Efficiently Indexing External Knowledge for Open-QA   Under Review

Post-Abstention: Towards Reliably Re-Attempting the Abstained Instances in QA   ACL, 2023

N. Varshney, C. Baral

- Post-Abstention aims at re-attempting the abstained instances of a given selective prediction system with the aim of increasing its 'coverage' without significantly sacrificing its 'accuracy'.

Can Open-Domain QA Reader Utilize External Knowledge Efficiently like Humans? AAAI @ KnowledgeNLP 2023

N. Varshney, M. Luo, C. Baral

- Studied an approach that utilizes both the 'closed-book' (parametric knowledge) and the 'open-book' (external knowledge) inferences in an efficient manner.
- Instead of using a large fixed number of passages for open-book inference, the system dynamically reads the external knowledge in multiple 'knowledge iterations'.
- Comparing with the Fusion-in-Decoder (FiD) reader, our approach matches FiD's accuracy by utilizing just 18.32% of its reader inference cost (FLOPs) and also outperformed it by achieving up to 55.10% and 77.32% accuracy on NQ Open and TriviaQA respectively.

Model Cascading: Towards Jointly Improving Inference Efficiency and Accuracy of NLP Systems EMNLP, 2022

N. Varshney, C. Baral

- Proposed a technique that utilizes a collection of models of varying capacities to accurately yet efficiently output predictions.

Investigating Selective Prediction Approaches Across Several Tasks in IID, OOD, and Adversarial Settings ACL, 2022

N. Varshney, S. Mishra, C. Baral

- Systematically studied 'selective prediction' in a large-scale setup of 17 datasets across NLI, QA, and Duplicate Detection tasks under in-domain, out-of-domain, and adversarial settings.
- Showed that despite leveraging additional resources (held-out data/computation), none of the existing approaches consistently and considerably outperforms a simple *MaxProb* baseline.
- Evaluated several approaches on their task-transfer ability.

ILDAE: Instance-Level Difficulty Analysis of Evaluation Data ACL, 2022

N. Varshney, S. Mishra, C. Baral

- Computed instance-level difficulty scores for evaluation instances and demonstrated their five novel applications:
  - Conducting efficient-yet-accurate evaluations with fewer instances saving computational cost and time,
  - Improving the quality of existing evaluation datasets by repairing erroneous and trivial instances,
  - Selecting the best model based on application requirements,
  - Analyzing dataset characteristics for guiding future data creation,
  - Estimating Out-of-Domain performance reliably.

Unsupervised Natural Language Inference Using PHL Triplet Generation ACL, 2022

N. Varshney, P. Banerjee, T. Gokhale, C. Baral

- Designed three novel unsupervised settings for NLI and proposed a procedural data generation approach that outperforms existing approaches by $\sim 13\%$ and raises the SOTA unsupervised performance to 66.75%.
- Also proposed a model-in-the-loop adversarial data collection strategy to efficiently collect high-quality non-trivial data instances that help achieve 12.2% higher accuracy with as little as $\sim 0.1\%$ of the training dataset.

NumGLUE: A Suite of Mathematical Reasoning Tasks ACL, 2022

S. Mishra, A. Mitra, N. Varshney, B. Sachdeva, P. Clark, C. Baral, A. Kalyan

- Proposed a multi-task benchmark that evaluates AI systems on eight different numerical understanding tasks and evaluated neural models including large language models.
- Proposed a knowledge-retrieval based multi-task learning method that outperforms existing models.

Towards Improving Selective Prediction Ability of NLP Systems ACL @ RepL4NLP, 2022

N. Varshney, S. Mishra, C. Baral

- Proposed a method that improves over MaxProb by calibrating the model outputs using prediction confidence and difficulty level of the instances.
- Instantiated the proposed method in NLI and Duplicate Detection tasks and showed that it outperforms existing approaches and achieves up to 15% improvement over MaxProb.

Benchmarking Generalization via In-Context Instructions on 1,600+ Language Tasks EMNLP, 2022

Y. Wang, ..., N. Varshney, ..., C. Baral, Y. Choi, N. Smith, H. Hajishirzi1, D. Khashabi

- Introduced Super-NaturalInstructions, a benchmark of 1,616 diverse NLP tasks and their expert-written instructions.
- Built Tk-Instruct, a transformer model trained to follow a variety of in-context instructions (plain language task definitions or k-shot examples).

A Unified Evaluation Framework for Novelty Detection and Accommodation in NLP   ACL, 2023

N. Varshney, H. Gupta, E. Robertson, B. Liu, C. Baral

On Dealing with Questions that Don't have Definitive Answers   ACL @ TrustNLP, 2023

N. Varshney*, A. Agarwal*, N. Patel*, M. Parmar, ..., and C. Baral

On Evaluating NLP Models' Understanding of Feasibility   EACL, 2023

H. Gupta, N. Varshney, S. Mishra, K. Pal, S. Sawant, K. Scaria, S. Goyal, C. Baral

- Introduced FeasibilityQA, a question-answering dataset involving binary classification and multi-choice multi-correct questions that test understanding of feasibility.
- Highlighted that even state-of-the-art models such as GPT-3 struggle to answer the feasibility questions correctly.

Let the Model Decide its Curriculum for Multitask Learning   NAACL @ DeepLo, 2022

N. Varshney, S. Mishra, C. Baral

- Proposed dataset and instance-level techniques to arrange training instances into a learning curriculum based on the model's own interpretation of difficulty.
- Achieved 4% accuracy improvement over other methods on experiments conducted for 12 datasets covering a variety of language understanding tasks.

An Architecture for Novelty Handling in a Multi-Agent Stochastic Environment   AAAI Sym. 2022

T. Thai, M. Shen, N. Varshney, S. Gopalakrishnan, U. Soni, M.Scheutz, C. Baral, J. Sinapov

- Introduced an architecture that allows agents to detect novelties, characterize those novelties, and build an appropriate adaptive model to accommodate them.

Experience

Tencent AI   May 2023 – Present
NLP Research Intern   *Bellevue, Washington*

- Addressing Hallucinations of Large Language Models.

Amazon Science   May 2022 – Aug 2022
Applied Scientist Intern   *Seattle, Washington*

- Web Question-Answering System for Alexa AI.

Microsoft   July 2018 – Aug 2019
Software Developer   *Bangalore, India*

- Worked towards developing a machine learning driven chat recommendation system aimed at augmenting user engagement with the product Microsoft 'Teams'.

Samsung R&D Institute   Summer 2017
Research Intern   *Bangalore, India*

- Developed a 'context prediction' application incorporating features based on device events (e.g app usage, location) and sensor data (proximity sensor).

Education

Arizona State University   Tempe, AZ, USA
Ph.D. in Computer Science   *Aug 2019 – Spring 2024 (Expected)*

- **Advisor**: Dr. Chitta Baral
- **CPGA**: 4/4
- **Awards**: SCAI doctoral fellowship, GPSA awards (3 times), SCAI conference award (2 times), Graduate College awards (5 times), AAAI student scholarship, ACL registration award.
- **Internships**: Amazon Science (Summer 2022), Tencent AI (Summer 2023)

BITS Pilani, Pilani Campus, India   Pilani, India
B.E (Hons) Computer Science   *2014-2018*

- **CGPA**: 9.11/10 (with Distinction)
- **Experience**: 'Web Intelligence & Social Computing' research lab under Prof. Poonam Goyal, CEERI research lab under Dr. J.L. Raheja.
- **Internships**: Microsoft, Samsung R&D Institute, Valuefirst Digital Media.

Honors and Awards

- SCAI Doctoral Fellowship, School of Computing and AI at ASU, 2023.
- AAAI Student Scholarship, 2023
- Graduate College Travel Award, ASU for AAAI 2023, ACL 2022, NAACL 2022, EMNLP 2022, and ACL 2023.
- GPSA Travel Award, ASU for EMNLP 2022 and ACL 2022.
- SCAI Conference Award, ASU for EMNLP 2022 and ACL 2022.
- Registration award from Repl4NLP for ACL, 2022.
- GPSA Internship Travel Award, ASU 2023.

Service

- Reviewer for ACL'23 (Question Answering track).
- Reviewer for EMNLP'23.
- Reviewer for EACL'23 (Question Answering track).
- Reviewer for CVPR workshop (Open-Domain Retrieval Under a Multi-Modal Setting) 2022 and 2023.
- Mentored B.S and M.S students for course projects and co-authored multiple research papers with them.
- Served as Maths and Science teacher for underprivileged kids in National Service Scheme (NSS), India.
- Participated in blood donation camps and health awareness drives.

Courses

- Natural Language Processing
- Statistical Machine Learning
- Artificial Intelligence
- NLP Methods in BioMedical
- Knowledge Representation
- Data Mining
- Social Media Mining
- Mobile Computing

Collaborators

- Swaroop Mishra (Research Scientist at Google Brain)
- Hongming Zhang (Senior Research Scientist at Tencent AI)
- Wenlin Yao (Senior Research Scientist at Tencent AI)
- Jianshu Chen (Principal Researcher at Tencent AI)
- Dong Yu (Distinguished Scientist at Tencent AI)
- Tejas Gokhale (Assistant Professor at University of Maryland, Baltimore County)
- Arindam Mitra (Data and Applied Scientist at Microsoft Research)
- Daniel Khashabi (Allen AI, Assistant Professor at Johns Hopkins University)
- Bing Liu (Professor at University of Illinois at Chicago)
- Pratyay Banerjee (Applied Scientist at Alexa AI, Amazon Science)
- Ashwin Kalyan (Allen AI)
- Yizhong Wang (Allen AI, University of Washington)
- Rik Koncel-Kedziorski (Alexa AI)
- Eric Robertson (PAR Government)
- Kuntal Pal (Applied AI ML Senior Associate at JPMorgan Chase & Co.)
- Man Luo (ASU)
- Mihir Parmar (ASU)