Siddhartha Mishra

📞 (413) 522-8842
✉ im.siddhartha.mishra@gmail.com
🖥 sidmishra.com
⦿ mishra-sid
in sidmishra97
⁜ Siddhartha Mishra

Education

2021 - 2022	**MS, Computer Science**, *4.0/4.0 GPA*. University of Massachusetts Amherst
2015 - 2019	**BTech, Computer Science and Engineering**, *8.78/10 GPA*. Indian Institute of Technology Hyderabad

Publications

Manuscript In Prep
Vision Reward Bench: Towards Scalable Benchmarking Without Supervision.
Siddhartha Mishra, Alexi Gladstone, Riashat Islam

EMNLP 2022
Super-NaturalInstructions: Generalization via Declarative Instructions on 1600+ NLP Tasks.
Yizhong Wang, Swaroop Mishra, ..., **Siddhartha Mishra**, and 37 others

ACL 2022
Word2Box: Capturing Set-Theoretic Semantics of Words using Box Embeddings.
Shib Dasgupta, Michael Boratko, **Siddhartha Mishra**, Shriya Atmakuri, Dhruvesh Patel, Xiang Li, Andrew McCallum

AAAI 2022
An Evaluative Measure of Clustering Methods Incorporating Hyperparameter Sensitivity.
Siddhartha Mishra, Nicholas Monath, Michael Boratko, Ari Kobren, Andrew McCallum

Academic Service

Reviewer AISTATS 2024, ICLR 2024, NeurIPS 2024, EMNLP 2024, ACL 2023, AAAI 2023, EMNLP 2022.

Teaching Assistant Machine Learning (Fall 2021), Statistics (Spring 2019), Computer Networks (Fall 2018), Probability (Fall 2017).

Experience

Jan 2023 - Current
Machine Learning Researcher, *Palantir Technologies*, Seattle, WA.
- **Foundry AIP Assist:** Conducted research and developed customized workflows for domain-specific agentic retrieval-augmented generation (RAG), tool-calling, and structured prediction components in the Foundry AIP Assist product. The fine-tuned models allowed greater control over the end-to-end pipeline, resulting in 7-10% improvement in auto-eval scores in offline and confidential environments at lower costs and latency.
- **USG Use Cases:** Advised US Government customers on deploying and integrating foundation models in critical use cases effectively and ethically with the analyst *in-the-loop*.
- **Model Infrastructure:** Upgraded the model deployment tooling and infrastructure (vLLM, TensorRT, Kubernetes) in Palantir to efficiently host large models (LLMs, VLMs) and enable model-agnostic integration of AI-native features across disparate environments in the Foundry platform.

Aug 2022 - Dec 2022
Applied Scientist Intern, *Amazon Alexa NLU*, New York, NY.
- Proposed a novel method to generate synthetic data for the zero-shot semantic parsing task by utilizing domain grammars for constrained sampling with large language models.
- Obtained 5 - 9% improvement in denotation accuracy on Geo and Scholar benchmarks over previous naive sampling/filtering baselines.
- Notable improvements in the SMCalflow dataset and internal benchmarks in domains with complex grammar in which brute force sampling is computationally infeasible.

Jan 2022 - Aug 2022	**Student Researcher**, *Amazon Alexa AI*, New York, NY. ◦ Proposed a parameter-efficient model that improves the domain adaptation of large language models in Natural Language Understanding tasks using Continuous Prompt tuning methods. ◦ Obtained 8 - 17% improved performance over prompting baselines on different tasks in GLUE/SuperGLUE benchmarks. ◦ Improved performance up to 21% in low resource domains of intent classification datasets.
May 2021 - Sep 2021	**Summer Research Intern**, *Information Extraction and Synthesis Lab, UMass Amherst*. ◦ Formulated geometric region-based embeddings for representing words using n-dimensional hyper-rectangles trained with a CBOW objective on the Wikipedia corpus. ◦ Obtained 3 - 15% improved performance on Word similarity benchmarks over vector-based baselines and qualitatively evaluated the set-theoretic expressivity. ◦ Work published in ACL 2022.
May 2019 - Jan 2021	**Machine Learning Engineer**, *Goldman Sachs*, Bengaluru, India. ◦ Worked in the Enterprise Machine Learning platform team on metric analysis, models for alert prediction and automatic resolution using serverless frameworks. ◦ Maintained dashboards for monitoring real-time alerts and managed model deployment pipeline. ◦ Improved performance of search queries by 25% in Big data log analysis platform by migrating to Elastic Stack.
May 2018 - Aug 2018	**Summer Analyst**, *Goldman Sachs*, Bengaluru, India. ◦ Built a plugin for a graphical pipeline design tool for constructing workflows to automatically resolve alerts. ◦ Designed and implemented a compiler for validating the pipeline and transform one payload format to another.

Awards / Scholarships

2017	Awarded Academic Excellence Award for the Academic year 2016 - 2017.
2017	Represented my university for ACM ICPC Amritapuri Asia regionals, ranked 30th (out of 250 candidates).
2015	Granted KVPY Fellowship by Indian Institute of Science, Bangalore, ranked 210th (out of 50000 candidates).
2015	Qualified JEE Advanced with 97.66 percentile (out of 150,000 candidates).
2014	Qualified Indian National Mathematics Olympiad (INMO), ranked 7th (out of 20000 candidates).

Technical skills

Languages	`Python, C/C++, Java, Go`.
ML	`PyTorch, Tensorflow, JAX, Huggingface, NumPy, Pandas, vLLM, TensorRT`.
Web	`Angular, React, NodeJS, FastAPI, MongoDB`.
Tools	`Docker, Kubernetes, Prometheus, Grafana, Kafka, Elasticsearch`.