Shilpa Suresh

✉ shilpa15397@gmail.com | 🔵 shilpa-suresh | 🐦@shilpa15397 | +1 (413) 512 - 3604

PROFESSIONAL EXPERIENCE

NLP & Data Scientist | Norstella   *Nov'23 – Present*

- Primary NLP specialist in the Real-World Data team; built NLP-enabled offerings by working closely with clinical analysts and technical teams from Electronic Health Records.
- Created domain-adapted production-level LLM Pipelines to perform information extraction using finetuning and Retrieval Augmented Generation (RAG)
- Developed generalizable biomarker extraction and disease diagnosis pipelines for oncological conditions to aid drug discovery pipelines and clinical trial matching.
- Building LLM-driven prompt-generation pipelines to reduce human-in-the-loop prompt engineering for rare oncological conditions

NLP Research Associate | Harvard Medical School & Boston Children's Hospital   *Jul'22 – Oct'23*

- Built a privacy-aware Large Language Model (LLM) pipeline for Clinical Information Extraction.
- Trained Domain adaptable models for Named Entity Recognition (NER) using BERT and a novel generalizable task-oriented learning architecture (published at BioNLP@ACL 2023)

Systems Analyst | Visa Inc.   *Jul '19 –Aug'20*

- Performed Data-driven Project Management and Business Process Improvement on a key company-wide cloud migration project with stakeholders that included upper management

Natural Language Processing Research Engineer | KeyReply   *Dec '17 – Jan '19*

- Designed a Task-driven, Domain Adaptable Model to Generate Questions from Text for Conversational AI

EDUCATION

M.S. Computer Science | *University of Massachusetts – Amherst*   *2020 – 2022*
Research in Natural Language Processing and Machine Learning for Domain Adaptation and Language Generation

B.Sc. Information Systems & Advanced Analytics | *Singapore Management University*   *2015 – 2019*
Focus on Program Management and Operational Excellence for AI/ML

KEY TECHNICAL SKILLS

Languages: Python, SQL, Java | **Libraries:** PyTorch | **Models:** BERT, LLaMA, Mixtral | **Prod:** AWS, Spark

PUBLICATIONS

- *Intermediate Domain Finetuning for Weakly Supervised Domain-adaptive Clinical NER* | BioNLP at **ACL** 2023
- *Improved Latent Tree Induction with Distant Supervision via Span Constraints* | **EMNLP** 2021
- *Can a chatbot read your mind?: Domain and Answer Adaptive Task-oriented Question Generation through Meta-Learning* | WiML Workshop at **NeurIPS** 2018

RECENT PROJECTS

- **Multilingual Multi-task Learning using DIORA**
  Research Study to improve the accuracy of the unsupervised parsing DIORA model by cross-training on multiple languages using multi-task learning. *Faculty Mentor: Prof. Andrew McCallum*
- **Continual Learning for NLP**
  Research study aimed at tackling catastrophic forgetting applied to self-supervised, meta-learning-based models for few-shot classification. *Faculty Mentor: Prof. Andrew McCallum*
- **Zero-shot Domain-adaptive TLDR Generation for Scientific Papers using BART**
  Generated short summaries of Medical Research papers for better biomedical science communication.

RESEARCH & PROFESSIONAL ACTIVITIES

- **Invited Program Committee Member** & Reviewer | NLP-OSS Workshop **EMNLP** 2020, **EMNLP** 2023