Kevin Joseph Scaria

Linkedin | kscaria@asu.edu | Ph. No.: +1-602-773-8226 | Github | Google Scholar

Experience

- **Amazon Web Services** — Atlanta, USA
  Software Dev Engineering Intern — *May 2023 - August 2023*
  - **AWS DynamoDB**: Developed a module to intelligently add lognodes for subpar partitions based on write activity to the DB which improved the durability and availability of active traffic partitions during rack-down events.
- **ASU Cognition and Intelligence Labs** — Tempe, USA
  Research Assistant under Dr. Chitta Baral — *Jan 2022 - Present*
  - **Automatic Synthetic Data Generation using LLMs**: Developed LLMDataGen, a low bias high diversity synthetic data generation strategy to generate corpora using ChatGPT via multi-step prompting without human intervention. Our approach also employed *self-correction* to generate noise-free labels as well. [Preprint, Under Review]
  - **Commonsense Reasoning**: Demonstrated the limitations of the SOTA LLMs like GPT-3 & T5 for numerical feasibility tasks, by developing the FeasibleQA dataset. [Paper 1 - EACL 23']
- **ASU Decision Theater Network** — Tempe, USA
  Applied Data Scientist — *Apr 2022 - Present*
  - **InstructABSA**: Achieved SOTA on various aspect-based sentiment analysis (ABSA) subtasks by instruction tuning the T5 model, surpassing 7x larger models' performance. Demonstrated sample efficiency of the approach, effect of various instruction prompts, and cross-domain generalizability. [Paper 2 Preprint, Under Review EMNLP 23']
  - **Epidemic Modelling**: Developing a neural network-based climate niche model to estimate valley fever incidences.
- **ASU Lab V2** — Tempe, USA
  Research Assistant under Dr. Paulo Shakarian — *Nov 2022 - May 2023*
  - **Metacognitive Error Correction**: Developed a meta-cognitive error correction methodology that improves the performance of neural classification models using first-order logic rules automatically extracted data to operationalize movement trajectory classification tasks. [Paper 3 Preprint, Under Review AAAI 24']
- **Tiger Analytics** — Chennai, India
  Data Scientist II - ML Engineering — *Jan 2021 - Dec 2021*
  - **Risk modeling for healthcare insurance provider**: Developed a framework to estimate the risk of hospitalization, high-cost claimants and ER visit given the patients claim journey for a leading health care insurance provider.
  - **Image segmentation for satellite images** : Developed a ML model using U-Net architecture to segment low resolution satellite images into various classes which was utilised to track temporal land cover changes for an AgTech startup.
  - **COVID War Room Dashboard**: Developed a scalable web application using Flask web development framework and Plotly Dash that reports near real-time COVID-19 information & metrics. Additionally, developed an approval based preferential access service to auto-generated reports, proprietary forecasting models and the epidemic calculator.

  Data Scientist I - ML Engineering — *Jan 2020 - Dec 2020*
  - **Cognitive Insights Dashboard - Text to code**: Developed a natural language processing (NLP) solution that converted language prompts to code for rapid generation of market insights in the form of interactive visualizations.
  - **Customer Lifetime Value (CLTV)Estimation for insurance provider**: Developed an ML solution to recommend the most optimal premium rates to maximize the CLTV for a Medicaid insurance provider.
  - **Information extraction from financial statements**: Developed a rule based parser using NLP to extract 16 attributes with high coverage and high precision from financial statements for a global financial analytics company.

Additional Research Work

- **Unsupervised Aspect Term Sentiment Classification**: Designed a multi-step unsupervised framework for aspect term sentiment classification. This involved utilizing semantic attention of encoder models like BERT, DeBERTA, and RoBERTa to re-weight candidate dependency relations among tokens to extract opinion words and sentiment polarity.
- **Attention Matrix Segmentation for Sequence Labelling Tasks**: Developing a new hybrid architecture with a language model encoder and vision transformer decoder to perform attention matrix segmentation for sequence labeling tasks such as named entity recognition (NER) and ABSA.

Education

- **Arizona State University** — Tempe, Arizona
  MS - Information Technology — CGPA: 4/4 — *Jan 2022 - December 2023*
- **SRM Institute of Science & Technology** — Chennai, India
  Bachelor of Technology — CGPA: 8.43/10 — *June 2014 - May 2018*

Skills, Achievements & Relevant Coursework

- **Programming Languages**: Java, Python, R, Spark, Scala, C++, JavaScript, SQL
- **Frameworks & Libraries**: Unix, Scikit, NLTK, SpaCy, TensorFlow, Keras, PyTorch, PySpark, Transformers, Accelerate, Flask, Plotly Dash, Streamlit, LangChain, Vector DB
- **Reviewer**: EMNLP 23'
- **Tools & Platforms**: AWS, GIT, Tableau, Microsoft Azure, GCP, Google DialogFlow, Tableau, Power BI, JIRA
- **Graduate Courses**: Statistical Machine Learning, Advanced DBMS, Analyzing Big Data, Natural Language Processing
- **Online Courses**: Machine Learning, Deep Learning, Computer Vision, NLP [Link], Data structures and algorithms [Link]