

Ishita Gopal, PhD

Tel: +1 (925) 394-8793

Mail: ishitagopal@gmail.com

Web: ishitagopal.github.io

Summary: Data Scientist with research experience currently developing and applying computational methods to analyze advertisement campaign data, model communication dynamics, and inform decision-making.

Tools and Languages: Python (Pandas, scikit-learn, PyTorch, transformers), R, Git, SQL, AWS

Skills: Machine Learning, Natural Language Processing, Causal Inference, Statistics

DATA SCIENCE EXPERIENCE

Data Science Researcher, Transdisciplinary Institute in Applied Data Sciences
Washington University in St Louis

Jan 2024 – Current

- **Deep Learning for Attack Ad Detection:** Built a pipeline to automate search, retrieval, transcription and translation of political ads from YouTube. Developed deep learning models using text and audio features to identify negative advertising (90% F1 score). Currently improving model efficiency by focusing on audio-based classification instead of text based analysis to streamline processing, reduce inference time, and generate more scalable solutions.
- **Data Science Workshops & Tutorials:** Led workshops on machine learning, NLP (e.g. classification, topic models), and deep learning for text data (e.g. LLMs, transfer learning, zero-shot and few-shot learning). [[Tutorials](#)]

Doctoral Researcher, Center for Social Data Analytics
Pennsylvania State University

Aug 2018 – Dec 2023

- **Discourse Detection & Attention:** Developed ML models to classify 1M+ tweets and identify factors predicting attention to COVID-19 discourse. Built data processing pipelines using Twitter's API. Fine-tuned BERT-based large language model, achieving 85% F1 score and 10% improvement over Random Forest and XGBoost. Used Bayesian shrinkage statistic to reveal group specific topic focus and variations in engagement. [[Published Paper](#)]
- **Text & Network models for Behavior Prediction:** Led a team of 6 in developing network regression models to identify predictors of online engagement among U.S. politicians by analyzing 300K Twitter observations. Applied lasso-assisted cosine similarity on 50K policy documents to measure peer similarities. Created [visualizations](#) to depict network structures and highlight individual interaction clusters. [[Paper Accepted at PSRM](#)]
- **Experimental Test of Peer Influence Diffusion:** Designed an email experiment (1K subjects) that controls for network effects to analyze the impact of peer information on policy related decision-making among U.S. politicians. Tracked 6K+ individuals, applied backbone extraction to identify peers, used zero-shot models on 90K bills to parse treatment policies, and ran regressions to analyze email response rate and click behavior. [[Paper Accepted at ANS](#)]
- **Experimental Evaluation of MTurk and Misinformation Recall on Facebook:** Ran 2 online experiments (9K participants) to examine limitations of MTurk's recruitment population and impact of Facebook comments on misinformation recall. Findings revealed biases in MTurk samples related to age and digital literacy and provides usage guidance for deriving insights from online samples. [[Published Paper](#)]

Data Science Intern

May - Aug 2022

Aware HQ, Columbus

- **Convolutional Neural Networks for Sensitive Data Detection:** Developed and deployed a credit card detection model for digital workspaces. Used CNN (EfficientNets) for transfer learning on hand-labeled data, utilized data augmentation techniques to reduce overfitting, improved model performance and achieved a 90% accuracy rate.

Economist

Aug 2016 – Aug 2018

The Energy & Resources Institute, Delhi

- Worked with government stakeholders to develop time series (ARIMA) models for electricity demand forecasting.
- Conducted scenario modeling to forecast impact of renewable uptake on coal capacity growth in India.

EDUCATION

Ph.D. Social Data Analytics, Pennsylvania State University, USA

2023

M.Sc. Economics, University of Warwick, UK

2015

B.A. (Hons) Economics, Miranda House, India

2014