

# Ishita Gopal, PhD

Tel: +1 (925) 394-8793

Mail: [ishitagopal@gmail.com](mailto:ishitagopal@gmail.com)

Web: [ishitagopal.github.io](https://ishitagopal.github.io)

## SUMMARY

Data scientist skilled in analyzing large structured and unstructured data (text, audio, networks), developing hypothesis-driven solutions, and using advanced computational methods to tackle real-world challenges.

**Skills:** Machine learning, deep learning, natural language processing (NLP), statistics, causal inference, network analysis

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

## DATA SCIENCE EXPERIENCE

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

*Jan 2024 – Current*

- **Developing language agnostic machine learning and deep learning models to detect negative advertising on YouTube (achieved 90% accuracy).** Investigating if audio features can exclusively be used for classifying ads and improve efficiency of analysis by eliminating the need for complex (multi-lingual) text analysis.
- **Conducted multi-session data science workshops in Python for an interdisciplinary group of faculty and graduate students.** Developed tutorials on web scraping, machine and deep learning (classification, regression, clustering), NLP and LLMs. [[Tutorials](#)]

**Doctoral Researcher, Center for Social Data Analytics**  
*Pennsylvania State University*

*Aug 2018 – Dec 2023*

- **Built machine learning classifiers and statistical models to measure impact of health and policy indicators on policymakers' discussions of COVID-19.** Collected 1M+ tweets using Twitter's API. Trained large language model (BERT) to identify COVID-19 discourse (F1 score of 85%). Improved F1 by ~10% over random forest and XGBoost algorithms. Integrated health and policy databases and used panel regressions to test hypotheses. [[Published paper](#)]
- **Led a team of 6 and developed predictive models to analyze group dynamics and communication patterns of 4K+ policymakers on Twitter.** Collected ~300K observations of legislator interactions from Twitter. Created [advanced visualizations](#) that identify clustering by party and state. Used permutation models on high-performance computing (HPC) and showed gender and race as significant predictors of interactions. [[Accepted paper at PSRM](#)]
- **Conducted 2 online experiments with 9K participants to evaluate MTurk's recruitment limitations and analyze impact of Facebook comments on recalling misinformation.** Demonstrated MTurk samples exhibit bias when treatment effects vary by age and digital literacy, providing guidance for academia and industry. [[Published Paper](#)]
- **Designed and implemented an email experiment with 1K subjects to test impact of peer effects on policy support diffusion, while accounting for network effects.** Tracked cosigning behavior of 6K+ US policymakers, identified peers using backbone extraction methods. Used zero-shot text classifiers (90K bills) to identify relevant treatment policies and built regression models to analyze email response and click behavior. [[Working Paper](#)]

**Data Science Intern**  
*Aware HQ, Columbus*

*May - Aug 2022*

- Developed and deployed a credit card detection model to flag sensitive data sharing in digital workspaces.
- Used deep learning 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**  
*The Energy & Resources Institute, Delhi*

*Aug 2016 – Aug 2018*

- 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