

# Ishita Gopal

Tel: +1 (925) 394-8793

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

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

## SUMMARY

*PhD candidate with 5 years of experience in data science. Adept at transforming abstract questions into tangible data problems. Leverage diverse data types (text, networks, images) and use cutting-edge techniques in machine learning, statistics, and experimentation to generate actionable insights.*

## SKILLS

**Tools and Languages:** Python, R, SQL, Git, AWS, Azure

**Data Science & Statistics:** Machine learning, natural language processing (NLP), deep learning, network analysis, inferential statistics, hypothesis testing, causal inference, time-series, panel-data models

## DATA SCIENCE EXPERIENCE

### Doctoral Research Scientist, Center for Social Data Analytics (C-SoDA), University Park

Aug 2018 – Current

- [Led a team of 6 to model complex behavioral patterns of 4K policymakers.](#) Built network simulation models using high performance computing. Created advanced visualizations for data summarization and produced a working paper currently under review for publication.
- [Delivered a high-impact journal article that analyzes policymakers' response to public health crisis.](#) Built text classifiers using large language models (BERT) to identify COVID-19 discussions in 1M+ tweets (F1 of 85%). Validated performance against Random Forest and XGBoost. Used hierarchical models for statistical analysis.
- [Produced a journal article that tests the limits of experiments conducted on MTurk \(9000 participants\).](#) Conducted 2 online experiments and showed MTurk should not be used when treatment effect heterogeneity is suspected in age or digital literacy. Results provide guidance for academic and industry research.
- [Designed an experiment for estimating impact of peer effects on support for environmental policies \(1000 subjects\).](#) Used backbone extraction methods to account for underlying network effects. Used zero-shot text classifiers on a dataset of 90K bills to identify treatment policies. Results provide insights about behavior of policymakers.
- [Worked in a cross-functional team and built deep learning models for election integrity detection.](#) Tested the performance of various LLMs (BERT, RoBERTa, XLNet) and active learning approaches to support C-SoDA's research initiatives.
- [Created instructional materials for practitioners \(book, chapters, and tutorials\).](#) Conducted coding workshops in Python to teach how to collect and analyze unstructured text and network data using data science methods.

### Data Science Intern, Aware HQ, Columbus

May - August 2022

- Developed and deployed a credit card detection model with AWS SageMaker to flag sensitive data sharing in digital workspaces to be used in Aware's product.
- Used deep learning (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

- Developed time series (ARIMA) models for electricity demand forecasting.
- Used scenario modeling to forecast the impact of renewable uptake on coal capacity growth in India. Results provided policy assessment support to the Indian Government.

## EDUCATION

Ph.D. Social Data Analytics & Political Science, *Pennsylvania State University*  
(Awards: *Princeton University Dissertation Scholar, C-SoDA Predoctoral Fellow*)

Expected 2023

M.Sc. Economics, *University of Warwick*

2015

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

2014