

# Ishita Gopal

Tel: +1 (925) 394-8793

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

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

## SUMMARY

*Data scientist with 5 years of research experience. 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**

*August 2018 – Current*

- [Led a team of 6 and built models to analyze complex behavioral patterns of 4K policymakers'](#). Implemented network simulation models using high performance computing. Created advanced visualizations for data summarization. Produced working paper under review for publication.
- [Collaborated with C-SoDA researchers to analyze policymakers' discourse on public health crisis.](#) Built deep learning models to identify COVID-19 discussions in 1M+ tweets (F1 of 85%). Used hierarchical regression models for statistical analysis. Produced high-impact journal article.
- [Implemented and analyzed 2 online survey experiments \(9,000 participants\) to test limitations of MTurk.](#) Showed MTurk should not be used when treatment effect heterogeneity is expected in age/digital literacy. Produced journal article that provides guidance for academic/industry research.
- [Designed field experiment \(1000 subjects\) to estimate impact of peer effects on support for environmental policies.](#) Used backbone extraction methods to account for underlying network effects. Used zero-shot text classifiers on 90K bills to identify treatment policies. Results provide insights on policymakers' behavior.
- Worked in cross-functional team and built deep learning models to detect election integrity discourse. Tested various LLMs (BERT, RoBERTa, XLNet) performance and active learning approaches. Results support research at C-SoDA.
- Created instructional materials (book, [chapters](#), and [tutorials](#)). Conducted coding workshops in Python/R 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**

*August 2016 – August 2018*

- Developed time series (ARIMA) models for electricity demand forecasting.
- Conducted 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, USA**

Expected 2023

(Awards: Princeton University Dissertation Scholar, C-SoDA Predoctoral Fellow)

**M.Sc.** Economics, **University of Warwick, UK**

2015

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

2014