Indira Sen

Computer Science P.H.D Candidate interested in the application of NLP methods in Computational Social Science

✉ indira.sen@rwth-aachen.de | ⌂ https://indiiigo.github.io/ | Indiiigo | 🐦 @indiiigosky | Google Scholar

Research Interests

I'm interested in improving the measurement of psychosocial constructs, including political and prejudicial attitudes and behaviors, from digital trace data using mixed methods and natural language processing. My research aims to illuminate and characterize the potentials, pitfalls, and consequences of understanding such attitudes from unobtrusive digital traces.

Education

RWTH-Aachen and GESIS - Leibnitz Institute for Social Sciences
Ph.D in Computer Science (Magna cum Laude)
Advisors: Prof. Claudia Wagner, Prof. Wil van der Aalst

Cologne, Germany

Oct, 2018 - Apr, 2024

Indraprastha Institute of Information Technology (IIIT-D)
M.S in Computer Science
Advisor: Prof. Ponnurangam Kumaraguru (PK)
GPA: 8.67/10

New Delhi, India

Aug, 2015 - Dec, 2017

Vellore Institute of Technology (VIT)
B.S in Computer Science
GPA: 9.1/10

Vellore, India

Jul, 2010 - May, 2014

Experience

University of Konstanz
Postdoctoral Researcher, Department of Political Science and Public Administration. Advisor: Prof. David Garcia

Konstanz, Germany

Oct 2023 -

- Research: Designing and Evaluating Interventions against Harmful Online Communication
- Teaching: 1) The Social Informatics of Large Language Models, 2) Introduction to Python

Computational Social Science Department, GESIS
Doctoral Researcher. Advisors: Prof. Claudia Wagner, Dr. Katrin Weller

Cologne, Germany

Oct 2018 - Sept 2023

- Identifying, Characterizing and Quantifying errors that occur when measuring attitudes from Digital Traces
- Building robust and interpretable NLP methods for measuring social science phenomena
- Large scale Twitter data collection using Amazon Web Services (AWS) as a service for Social Scientists.

Snap Inc.
Research Intern, Computational Social Science. Advisor: Dr. Francesco Barbieri

Santa Monica, USA

Jul 2022 - Oct 2022

- Developing a Stakeholder-centred Framework for Multilingual Sensitive Content Detection

Nokia Bell Labs
Intern at Social Dynamics Group. Advisor: Dr. Daniele Quercia

Cambridge, UK (virtual)

Jun 2021 - Aug 2021

- Understanding and Characterizing the sustainability efforts of companies from employee reviews.
- Developed mixed-method techniques for detecting employees' views on their employer's sustainability efforts using a deep-learning framework (SBERT)
- Results from the internship were published at CSCW'22 and IEEE Vis'22.

SAGE Ocean
SAGE Ocean Fellow. Advisor: Dr. Daniela Duca

London, UK (virtual)

Oct 2020 - Dec 2020

- Contributing to a Text cleaning WebApp called Texti
- Working on NLP transformations in Python and documentation
- Results from the Fellowship are described here.

Precog Research Lab, IIIT Delhi
Research Associate. Advisors: Prof. Ponnurangam Kumaraguru and Prof. Arun Balaji Buduru

New Delhi, India

Dec 2017 - Aug 2018

- Landscaping the fake like marketplace on Instagram and characterizing fake likers, and their attack vectors.
- Detecting Multimodal Hatespeech on Social Media

CHILI Lab, École polytechnique fédérale de Lausanne *Lausanne, Switzerland*
SUMMER INTERN, ADVISOR: PROF. PIERRE DILLENBOURG *Jun 2017– Aug 2017*
- Worked on inferring dependencies between EPFL courses based on Concept Mining and grading patterns

Nanyang Technological University *Singapore*
RESEARCH ASSISTANT, ADVISOR: PROF. ANWITAMAN DUTTA *Jan 2017– May 2017*
- Master's Thesis: Worked on detecting fake engagement on Instagram.

Grants

2022	**Grant for Organizing the 'Conference on Harmful Online Communication'**, Co-Applicant, 20,000 EUR	*Fritz Thyssen Foundation, Germany*
2022	**Grant for the Project 'Not one bit sexist: A competitive game against sexist language'**, For creating a gamified crowdsourcing platform to collect perceptions of sexism from a diverse group of people. Co-Applicant, 10,000 EUR	*Meta (Facebook)*
2020	**Travel Grant**, For attending EMNLP, 200 EUR	*virtual*
2020	**Travel Grant**, For attending FAT* 2020, 1500 EUR	*Barcelona, Spain*
2019	**Travel Grant**, For attending Summer Institute in Computational Social Science (SICCS), 1000 EUR	*Bamberg, Germany*

Honors & Awards

2023	**Research Award for Best Paper by a Doctoral Researcher**, Chosen by the GESIS Scientific Board from 40+ doctoral candidates	*GESIS*
2022	**Research Award for Interdepartmental Collaboration**, for interdisciplinary research collaborations	*GESIS*
2021	**Summer Intern Award for Outstanding Innovation**, Nokia Bell Labs	*Cambridge, UK / virtual*
2021-2023	**Best Reviewer Award**, ICWSM (2021, 20222, 2023), The WebConf (2023)	
2020	**Selected for SAGE Text-Mining Fellowship**, SAGE, Ocean	*London, UK (virtual)*
2020	**Selected as Facilitator for the Alan Turing Institute Data Study Group**, 2 week hackathon where we worked with an Industry partner (CatsAI) to deploy explainable AI predictions	*virtual*
2017	**Nomination for Best Master's Thesis**, 5 among 20+ theses in the Computer Science Department	*IIIT-Delhi, India*
2017	**TA Excellence Award**, Foundations of Computer Security	*IIIT-Delhi, India*
2017	**Accepted to Summer@EPFL**, One of 60 interns selected out of 2000 applications	*EPFL, Switzerland*
2015-2016	**Dean's Merit List**, For Academic Performance	*IIIT-Delhi, India*
2015	**GATE Scholarship**, Scholarship for Master's Studies	
2011-2014	**Merit Scholarship**, For Outstanding Academic Achievement in 4 years of Bachelor's studies	*VIT, Vellore*

Publications

Journal Articles

Insider Stories: Analyzing Internal Sustainability Efforts of Major US Companies from Online Reviews. *2023*
INDIRA SEN, DANIELE QUERCIA, LICIA CAPRA, MATTEO MONTECCHI, AND SANJA ŠĆEPANOVIĆ.
HUMANITIES AND SOCIAL SCIENCES COMMUNICATIONS 10, NO. 1 (2023): 1-9.

Depression at Work: Exploring Depression in Major US Companies from Online Reviews. *2022*
INDIRA SEN, DANIELE QUERCIA, MARIOS CONSTANTINIDES, MATTEO MONTECCHI, LICIA CAPRA, SANJA SCEPANOVIC, AND RENZO BIANCHI.
PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION. TO BE PRESENTED AT THE 25TH ACM CONFERENCE ON COMPUTER-SUPPORTED COOPERATIVE WORK AND SOCIAL COMPUTING (CSCW)

Visualizing Internal Sustainability Efforts in Big Companies. *2022*
CHIARA CECCARINI, EDYTA PAULINA BOGUCKA, INDIRA SEN, MARIOS CONSTANTINIDES, CATIA PRANDI, AND DANIELE QUERCIA.
IEEE COMPUTER GRAPHICS AND APPLICATIONS

A Total Error Framework for Digital Traces of Human Behavior on Online Platforms *2021*

Indira Sen, Fabian Floeck, Katrin Weller, Bernd Weiss, and Claudia Wagner.
Public Opinion Quarterly

Conference Articles

People Make Better Edits: Measuring the Efficacy of LLM-generated Counterfactually Augmented Data for Harmful Language Detection *2023*

Indira Sen, Dennis Assenmacher, Mattia Samory, Isabelle Augenstein, Wil van der Aalst, and Claudia Wagner. *To appear*
Conference on Empirical Methods on Natural Language Processing (EMNLP)

Counterfactually Augmented Data and Unintended Bias: The Case of Sexism and Hate Speech Detection *2022*

Indira Sen, Mattia Samory, Claudia Wagner and Isabelle Augenstein. *short paper*
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)

How Does Counterfactually Augmented Data Impact Models for Social Computing Constructs? *2021*

Indira Sen, Mattia Samory, Fabian Floeck, Claudia Wagner and Isabelle Augenstein.
Conference on Empirical Methods on Natural Language Processing (EMNLP)

"Call me sexist, but…" : Revisiting Sexism Detection Using Psychological Scales and Adversarial Samples *2021*

Mattia Samory, **Indira Sen**, Julian Kohne, Fabian Floeck, and Claudia Wagner.
International Conference on Web and Social Media (ICWSM)

On the Reliability and Validity of Detecting Approval of Political Actors in Tweets *2020*

Indira Sen, Fabian Floeck, and Claudia Wagner.
Conference on Empirical Methods on Natural Language Processing (EMNLP)

Empowering First Responders through Automated Multimodal Content Moderation *2018*

Divam Gupta, **Indira Sen**, Niharika Sachdeva, Ponnurangam Kumaraguru, Arun Balaji Buduru. *Best Paper Award*
IEEE International Conference on Cognitive Computing

Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram *2018*

Indira Sen, Anupama Aggarwal, Shiven Mian, Siddharth Singh, Ponnurangam Kumaraguru, Anwitaman Datta. *short paper*
WebSci

Workshop Articles

The End of the Rehydration Era The Problem of Sharing Harmful Twitter Research Data *2023*

Dennis Assenmacher, **Indira Sen**, Leon Fröhling, Claudia Wagner *Best Paper Award*
NEATClass, In conjunction with ICWSM

What Human Rationales Tell us about Machine Explanations *2021*

Indira Sen, Mattia Samory, Fabian Floeck, Claudia Wagner
HCI+NLP Workshop, In conjunction with EACL

Measuring Public Opinion Online: Potentials and Pitfalls *2020*

Indira Sen, Fabian Floeck, Claudia Wagner *Poster*
Women in Machine Learning (WiML), In conjunction with Neurips

A Twitter Corpus for Hindi-English Code Mixed POS Tagging *2018*

Kushagra Singh, **Indira Sen**, Ponnurangam Kumaraguru. *Data Paper*
Workshop on Natural Language Processing for Social Media (SocialNLP), In conjunction with ACL

Language Identification and Named Entity Recognition in Hinglish Code Mixed Tweets *2018*

Kushagra Singh, **Indira Sen**, Ponnurangam Kumaraguru. ACL Student Research Workshop

Presentations and Invited Talks

Theory-driven NLP *Milan, Italy (virtual)*

Bocconi University *Nov 2023*

Social Media based Measurements of Psychosocial Phenomena: Potentials and Pitfalls *Düsseldorf, Germany*
Heinrich Heine University *May 2022*

Identifying and Mitigating Gender Bias in Algorithmic Systems *Cologne*
Cologne Business School *May 2022*

A Total Error Framework for Digital Traces of Humans *Virtual*
DGS Kongress *Sep 2020*

(Mis)Measuring Attitudes from Digital Traces *Virtual*
Third Annual Politics and Computational Social Sciences Conference (PaCSS) *Aug 2020*

Teaching

Introduction to Programming with Python *Konstanz, Germany*
University of Konstanz *Oct 2023*
- Introduction to Python Block Course at the University of Konstanz (16.10-20.10, Winter Semester 2023/2024)

The Social Informatics of Large Language Models *Konstanz, Germany*
University of Konstanz *Oct 2023 - Mar 2024*
- Co-teaching a seminar course on the impact of Large Language Models om Society with practical hands-on sessions.

Social Research with Social Media Data: Potentials and Pitfalls. *GESIS, Cologne, Germany*
GESIS *2020-2022*
- Tutorial on diagnosing errors and biases in observational studies from digital traces

Translation Tutorial on the Total Error Paradigm for Social Inferences *Barcelona, Spain*
FAT* 2020 *Jan 2020*
- Tutorial on diagnosing errors and biases in observational studies from digital traces

Introduction to Python and Data Science with Python *Cologne, Germany*
University of Cologne *Oct, 2019*
- Interactive, hands-on five-day lecture series for teaching introductory python concepts to social science doctoral candidates

Teaching Assistant *New Delhi, India*
IIIT Delhi *2015 - 2017*
- Linear Algebra
- Foundations of Computer Security

Academic Service and Mentorship

Tutorials Co-chair
International Conference on Web and Social Media (ICWSM'24)

Workshop Organizer
Documenting Web Data for Social Research (#DocuWeb22), co-located with WebSci'22

Workshop Organizer
Participatory Development of Quality Guidelines for Social Media Research, co-located with ICWSM'21

Reviewing
PC member
TheWebConf (2023, 2024), EMNLP (2023), ICWSM (2021, 2022, 2023), ACL Student Research Workshop (2020, 2021, 2022), WikiWorkshop (2022), HCI+NLP workshop (2022)
Reviewer
CSCW (2019, 2020), TheWebconf (2020, 2021), EMNLP (2020), ACL Rolling Review (2021, 2022, 2023), EPJ Data Science (2022)

Mentorship

- Esi Doda (RWTH Aachen). Master's Thesis. Racial Disparities in Social Media Commentary about Football
- Vani Agarwal (IIIT-D). Master's Thesis. Improving Content Moderation through Co-Training
- Kushagra Singh (IIIT-D). Bachelor's Thesis. Information Extraction Tools for Code-Mixed Hindi-English Social Media Content

Skills

Programming: Python, R, Java, C++, C, MongoDB, SQL, Bash

Programming Libraries/Tools: scikit-learn, pytorch, tensorflow, huggingface, transformers, keras, pandas, NLTK, NetworkX, Amazon Web Services

Surveys: Amazon Mechanical Turk, Prolific, Qualtrics, LimeSurvey

Natural Languages: English (fluent, C2), German (intermediate, B1), Hindi (fluent), Bengali (native)