Indira Sen

COMPUTER SCIENCE RESEARCHER INTERESTED IN THE APPLICATION OF NLP METHODS IN COMPUTATIONAL SOCIAL SCIENCE

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

Education

RWTH-Aachen and GESIS - Leibnitz Institute for Social Sciences   *Cologne, Germany*

PH.D IN COMPUTER SCIENCE (MAGNA CUM LAUDE)

ADVISORS: PROF. CLAUDIA WAGNER, PROF. WIL VAN DER AALST

DISSERTATION: IDENTIFYING, CHARACTERIZING, AND MITIGATING ERRORS IN THE AUTOMATED MEASUREMENT   *Oct, 2018 - Apr, 2024*
OF SOCIAL CONSTRUCTS FROM TEXT DATA

Indraprastha Institute of Information Technology (IIIT-D)   *New Delhi, India*

M.S IN COMPUTER SCIENCE   *Aug, 2015 - Dec, 2017*

ADVISOR: PROF. PONNURANGAM KUMARAGURU (PK)

GPA: 8.67/10

Vellore Institute of Technology (VIT)   *Vellore, India*

B.S IN COMPUTER SCIENCE   *Jul, 2010 - May, 2014*

GPA: 9.1/10

Experience

University of Mannheim, Business School, Chair of Data Science for the Social and Economic Sciences   *Mannheim, Germany*

JUNIOR FACULTY MEMBER, ADVISOR: PROF. MARKUS STROHMAIER   *Sep 2024 - present*

- Research: Designing and Evaluating Representative and Equitable Language Technology
- Supervising Master's Theses for the Master's in Data Science.

University of Konstanz, Department of Politics and Public Administration   *Konstanz, Germany*

POSTDOCTORAL RESEARCHER AND LECTURER, ADVISOR: PROF. DAVID GARCIA   *Oct 2023 - Aug 2024*

- Research: Designing and Evaluating Interventions against Harmful Online Communication
- Teaching courses for Social and Economic Data Science (SEDS) Master's Program.

GESIS, Computational Social Science Department   *Cologne, Germany*

DOCTORAL RESEARCHER. ADVISORS: PROF. CLAUDIA WAGNER, DR. KATRIN WELLER   *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.   *Santa Monica, USA*

RESEARCH INTERN, COMPUTATIONAL SOCIAL SCIENCE. ADVISOR: DR. FRANCESCO BARBIERI   *Jul 2022 - Oct 2022*

- Developing a Stakeholder-centred Framework for Holistic Content Moderation

Nokia Bell Labs   *Cambridge, UK (virtual)*

INTERN AT SOCIAL DYNAMICS GROUP. ADVISOR: DR. DANIELE QUERCIA   *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
- Results from the internship were published at CSCW'22 and IEEE Vis'22.

Publications

Journal Articles

The Unseen Targets of Hate – A Systematic Review of Hateful Communication Datasets.   *2024*

ZEHUI YU, **INDIRA SEN**, DENNIS ASSENMACHER, MATTIA SAMORY, LEON FRÖHLING, CHRISTINA DAHN, DEBORA NOZZA AND
CLAUDIA WAGNER.   *To Appear*
SOCIAL SCIENCE COMPUTER REVIEW.

Assessing Data Quality in the Age of Digital Social Research: A Systematic Review. *2024*

Jessica Daikeler, Leon Fröhling, **Indira Sen**, Lukas Birkenmaier, Tobias Gummer, Jan Schwalbach, Henning Silber, Bernd Weiß, Katrin Weller, and Clemens Lechner.
To Appear

Social Science Computer Review

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

The Geography of US Companies That Care About Their Employees. *2024*

Indira Sen, Sanja Scepanovic, Marios Constantinides, and Daniele Quercia.
In Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work

An Open Multilingual System for Scoring Readability of Wikipedia. *2024*

Mykola Trokhymovych, **Indira Sen**, and Martin Gerlach.
In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics

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.
Conference on Empirical Methods on Natural Language Processing (EMNLP)

A Multidisciplinary Lens of Bias in Hate Speech. *2023*

Paula Reyero Lobo, Joseph Kwarteng, Mayra Russo, Miriam Fahimi, Kristen Scott, Antonio Ferrara, **Indira Sen**, and Miriam Fernandez.
In Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM). *short paper*

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

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

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

Indira Sen, Mattia Samory, Fabian Flöck, 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

Peer-Reviewed 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 Flöck, Claudia Wagner
HCI+NLP Workshop, In conjunction with EACL

Measuring Public Opinion Online: Potentials and Pitfalls   *2020*
 Indira Sen, Fabian Flöck, 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

Grants and Third-Party Funding

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, 30,000 EUR	*Meta (Facebook)*
2020	**Travel Grant,** For attending EMNLP, 200 EUR	*virtual*
2020	**Travel Grant,** For attending FAT* 2020 (now called FAccT), 1500 EUR	*Barcelona, Spain*
2019	**Travel Grant,** For attending Summer Institute in Computational Social Science (SICCS), 1000 EUR	*Bamberg, Germany*

Teaching

Measurement and Representation Biases in Digital Trace Data-based Studies   *Konstanz, Germany*
University of Konstanz   *Apr 2024-Aug 2024*
• Teaching a reading seminar course on using digital trace data from web and social media platforms like Facebook, Instagram, Wikipedia and others to measure social phenomena such as political attitudes and health behaviours.

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.

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)

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

Conference Tutorials Co-chair

International Conference on Web and Social Media (ICWSM'24)

Workshop Organizer

4th Workshop on Computational Linguistics for the Political and Social Sciences (CPSS'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

Supervised or Co-supervised theses

- 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

Selected Invited Talks

Multilingual and Multicultural Misrepresentation in LLM Simulations of People   *Vienna, Austria*

Lecture Series on Network Inequality, Complexity Science Hub   *August 2024*

Incorporating Theory into Computational Measures of Social Phenomenon   *Bielefeld, Germany*

Workshop on "Computational Social Science: Bridging Data and Methods with Theory and Applications",

University of Bielefeld   *May 2024*

Building Robust Social NLP Models   *Berlin, Germany*

Speech and Language Technology Department, German Research Center for Artificial Intelligence (DFKI)   *Feb 2024*

Incorporating Theory into NLP Models for Measuring Social Constructs   *Bremen, Germany (hybrid)*

Social Science Lab, University of Bremen   *Nov 2023*

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

Bocconi University   *Nov 2022*

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, Germany*
COLOGNE BUSINESS SCHOOL *May 2022*

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	**Best Master's Thesis Finalist,** 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*

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)