

Koustuv Saha

Curriculum Vitae

E1556R, Coda Tech Square
765 W Peachtree Street NW, Atlanta, GA 30308, US
✉ koustuv.saha@gatech.edu
🌐 www.koustuv.com

I am a computer scientist interested in the interdisciplinary areas of computational social science and social computing. By complementing multimodal datastreams with social media, I adopt methods from machine learning, statistics, natural language, and causal inference analysis to sense, predict, and examine wellbeing and psychosocial dynamics of individuals and collectives, particularly in situated contexts, such as college campuses and workplaces.

Education

2016– **Doctor of Philosophy (Ph.D.) Candidate**, *Computer Science*.

School of Interactive Computing, Georgia Institute of Technology

Advisor: Dr. Munmun De Choudhury, **Specialization:** Social Computing

Dissertation Title: Computational and Causal Approaches on Social Media and Multimodal Data: Examining Wellbeing in Situated Contexts

Doctoral Minor: Computational Social Science and International Affairs

- April 2020, Ph.D. Proposal Defense. Committee: Dr. Munmun De Choudhury, Dr. Gregory D. Abowd, Dr. Thomas Ploetz, Dr. Emre Kiciman, Dr. Gloria Mark
- April 2018, Ph.D. Qualifying Examination. Committee: Dr. Munmun De Choudhury, Dr. Gregory D. Abowd, Dr. Jacob Eisenstein, Dr. Thomas Ploetz

2008–2012 **Bachelor of Technology (Hons.)**, *Computer Science and Engineering*.

Computer Science and Engineering, Indian Institute of Technology Kharagpur

Guide: Dr. Sujoy Ghose, **B.Tech Project:** Ontology based Semantic Information Retrieval

Publications

Refereed Conference and Journal Articles

- 2020 **Saha, K.**, & Sharma, A. (2020). Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities. In Proceedings of the 14th International AAAI Conference on Web and Social Media., Presented at ICWSM 2020, [PDF](#).
- 2020 Das Swain, V.*, **Saha, K.***, Reddy, M. D., Rajvanshy, H., Abowd, G. D., & De Choudhury, M. CHI (2020). Modeling Organizational Culture with Workplace Experiences Shared on Glassdoor. In Proceedings of the CHI Conference on Human Factors in Computing Systems, (* co-primary authors), Video Presented for CHI 2020, [PDF](#).
- 2020 Bin Morshed, M.*, **Saha, K.***, De Choudhury, M., Abowd, G.D., & Ploetz, T., (2020). Modeling Pervasive Self-Esteem with Passive Sensing. In Proceedings of the 14th EAI International Conference on Pervasive Health Computing Technologies for Healthcare, (* co-primary authors), Video Presented for PervasiveHealth 2020, [PDF](#).
- 2020 **Saha, K.**, Ernala, S. K., Dutta, S., Sharma, E., & De Choudhury, M. (2019). Understanding HCII Moderation in Online Mental Health Communities. In 22nd International Conference on Human-Computer Interaction, Springer LNCS, Presented at HCII 2020, [PDF](#).

- 2019 **Saha, K.**, Reddy, M. D., Mattingly, S. M., Moskal, E., Sirigiri, A., & De Choudhury, M., (2019). CSCW LibRA: On LinkedIn based Role Ambiguity and Its Relationship with Wellbeing and Job Performance. In Proceedings of the ACM on Human-Computer Interaction (PACM HCI), 3, (CSCW), 137., Presented at CSCW 2019, [PDF](#).
- 2019 **Saha, K.**, Kim, S.C., Reddy, M. D., Carter, A.J., Sharma, E., Haimson, O.L., & De Choudhury M., CSCW (2019). The Language of LGBTQ+ Minority Stress Experiences on Social Media. In Proceedings of the ACM on Human-Computer Interaction (PACM HCI), 3, (CSCW), 89., Presented at CSCW 2019, [PDF](#).
- 2019 **Saha, K.**, Sugar, B., Torous, J., Abrahao, B., Kiciman, E., & De Choudhury, M. (2019). A Social ICWSM Media Study on The Effects of Psychiatric Medication Use. In Proceedings of the 13th International AAAI Conference on Web and Social Media., Presented at ICWSM 2019, [Outstanding Study Design Paper](#), [PDF](#).
- 2019 **Saha, K.**, Chandrasekharan, E., & De Choudhury, M. (2019). Prevalence and Psychological Effects WebSci of Hateful Speech in Online College Communities. In Proceedings of the 11th ACM Conference on Web Science, Presented at WebSci 2019, [PDF](#).
- 2019 **Saha, K.**, Reddy, M.D., Das Swain, V., Gregg, J.M., Grover, T., Lin, S., Martinez, G.J., Mattingly, ACII S.M., Mirjafari, S., Mulukutla, R., Nies, K., Robles-Granda, P., Sirigiri, A., Yoo, D.W., Audia, P., Campbell, A.T., Chawla, N.V., D'Mello, S.K., Dey, A.K., Jiang, K., Liu, Q., Mark, G., Moskal, E., Striegel, A., & De Choudhury, M. (2019). Imputing Missing Social Media Data Streams in Multisensor Studies of Human Behavior. In Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction (ACII), (Non *first three* authors are listed alphabetically), Presented at ACII 2019, [PDF](#).
- 2019 Das Swain, V., **Saha, K.**, Rajvanshy, H., Sirigiri, A., Gregg, J.M., Lin, S., Martinez, G.J, Mattingly, IMWUT S.M., Mirjafari, S., Mulukutla, R., Nies, K.A., Reddy, M.D., Robles-Granda, P., Campbell, A.T., Chawla, N.V., D'Mello, S.K., Dey, A. K., Jiang, K., Liu, Q., Mark, G., Moskal, E., Striegel, A., Tay, L., Abowd, G.D., & De Choudhury, M (2019). A Multisensor Person-Centered Approach to Understand the Role of Daily Activities in Job Performance with Organizational Personas. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3,(4),130., (Non *first four* authors are listed alphabetically), [PDF](#).
- 2019 Bin Morshed, M., **Saha, K.**, Li, R., D'Mello, S.K., De Choudhury, M., Abowd, G.D., & Ploetz, T., IMWUT (2019). Prediction of Mood Instability with Passive Sensing. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3,(3),75., [PDF](#).
- 2019 **Saha, K.***, Torous, J.*, Ernala, S.K., Rizuto, C., Stafford, A., & De Choudhury, M. (2019). TBM A Computational Study of Mental Health Awareness Campaigns on Social Media. Translational Behavioral Medicine, 9(6), (* co-primary authors), [Link](#).
- 2019 **Saha, K.**, Bayraktaraglu, A.E., Campbell, A.T., Chawla, N.V., De Choudhury, M., D'Mello, S.K., CHI Case Dey, A.K., Gao, G., Gregg, J.M., Jagannath, K., Mark, G., Martinez, G.J, Mattingly, S.M., Moskal, Study E., Sirigiri, A., Striegel, A., & Yoo, D. W., (2019). Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems Extended Abstracts, (Non *first* authors are listed alphabetically), Presented at CHI 2019, [PDF](#).

- 2019 IMWUT Mirjafari, S., Masaba, K., Grover, T., Wang, W., Audia, P., Campbell, A.T., Chawla, N.V., Das Swain, V., De Choudhury, M., D'Mello, S.K., Dey, A.K., Gao, G., Gregg, J.M., Grover, T., Jagannath, K., Jiang, K., Lin, S., Liu, Q., Mark, G., Martinez, G.J., Mattingly, S.M., Moskal, E., Mulukutla, R., Nepal, S., Nies, K.A., Reddy, M.D., Robles-Granda, P., **Saha, K.**, Sirigiri, A., & Striegel, A., (2019). Differentiating Higher and Lower Job Performers in the Workplace using Mobile Sensing. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3, (2), 37, (Non *first four* authors are listed alphabetically), [PDF](#).
- 2019 CHI Case Study Mattingly, S.M., Gregg, J.M., Audia, P., Bayraktaraglu, A.E., Campbell, A. T., Chawla, N.V, Das Swain, V., De Choudhury, M., D'Mello, S.K., Dey, A.K., Gao, G., Jagannath, K., Jiang, K., Lin, S., Liu, Q., Mark, G., Martinez, G.J, Masaba, K., Mirjafari, S., Moskal, E., Mulukutla, R., Nies, K., Reddy, M.D., Robles-Granda, P., **Saha, K.**, Sirigiri, A., & Striegel, A., (2019). The Tesseract Project: Large-Scale, Longitudinal, *In Situ*, Multimodal Sensing of Information Workers. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems Extended Abstracts, (Non *first two* authors are listed alphabetically), [PDF](#).
- 2019 FAT* Ribeiro, F.N.*, **Saha, K.***, Babaei, M., Henrique, L., Messias, J., Benevenuto, F., Goga, O., Gummadi, K.P., & Redmiles, E.M. (2019). On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook. In Proceedings of the 2nd ACM Conference on Fairness, Accountability, and Transparency (FAT*), (* co-primary authors), Presented at FAT* 2019, [PDF](#).
- 2019 SocInfo **Saha, K.**, Reddy, M.D., & De Choudhury M., (2019). JobLex: A Lexico-Semantic Knowledgebase of Occupational Information Descriptors. 11th International Conference on Social Informatics (SocInfo), [PDF](#).
- 2018 ICWSM **Saha, K.**, Weber, I, & De Choudhury, M. (2018). A Social Media Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses. In Proceedings of the 12th International AAAI Conference on Web and Social Media., Presented at ICWSM 2018, [PDF](#).
- 2018 CSCW **Saha, K.**, & De Choudhury, M. (2017). Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. In Proceedings of the ACM on Human-Computer Interaction, 1, (CSCW), 92., Presented at CSCW 2018, [PDF](#).
- 2017 IMWUT **Saha, K.**, Chan, L., de Barbaro, K., Abowd, G. D., & De Choudhury, M. (2017). Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1, (3), 95., *Presented at UbiComp 2017*, [PDF](#).
- 2017 JMIR **Saha, K.**, Weber, I., Birnbaum, M.L., & De Choudhury, M. (2017). Characterizing Awareness of Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates. Journal of Medical Internet Research, 19(5)., Presented at Mental Health Symposium, CHI 2017, [Link](#).
- 2017 CSS Sharma, E.*, **Saha, K.***, Ernala, S.K.*, Ghoshal, S.*, & De Choudhury, M. (2017). Analyzing Ideological Discourse on Social Media: A Case Study on the Abortion Discourse. In Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas (CSS), (* co-primary authors), [PDF](#).

Refereed Workshop Articles and Presentations

- 2020 CSCW DC **Saha, K.** (2020). Computational and Causal Examinations of Wellbeing in Situated Contexts by Leveraging Social Media and Multimodal Data. In CSCW 2020 Doctoral Consortium, [PDF](#).

- 2020 Das Swain, V.*, **Saha, K.***, Abowd, G. D., & De Choudhury, M. (2020). Social and Ubiquitous New Future Technologies for Future of Worker Wellbeing. In Microsoft Research Symposium on the New Future of Work, (* co-primary authors), [In Highlighted Talks, PDF](#).
- 2020 Martinez, G. J.*, Mattingly, S. M.*, Young, J., Faust, L., Dey, A.K., Campbell, A.T., De Choudhury, WristSense, M., Mirjafari, S., Nepal, S., Robles-Granda, P., **Saha, K.**, & Striegel, A. (2019). Improved Sleep IEEE PerCom Detection Through the Fusion of Phone Agent and Wearable Data Streams. In WristSense 2020:  Workshop on Sensing Systems and Applications Using Wrist Worn Smart Devices, colocated with IEEE PerCom, (* co-primary authors), [Best Paper Honorable Mention, PDF](#).
- 2019 **Saha, K.**, Das Swain, V., Yoo, D. W., Abowd, G. D., & De Choudhury, M. (2019). Tesserae: ITU Multimodal Sensing To Model Individual Differences and Job Performance at Workplaces. In ITU Kaleidoscope Kaleidoscope Demos 2019.
- 2019 **Saha, K.***, Moshed, M. B.*, Moon, S., O'Connor, J., Das, S., Abowd, G. D., Ploetz, T., & De ITU Choudhury, M. (2019). Understanding Health-Oriented Privacy Perceptions Regarding Ubiquitous Kaleidoscope Technology Use. In ITU Kaleidoscope Demos 2019, (* co-primary authors).
- 2019 **Saha, K.**, Chandrasekharan, E., & De Choudhury, M. (2019). Hateful Speech in Online College Norms, Communities. In Managing and Designing for Norms in Online Communities Workshop, colocated with ICWSM ICWSM 2019, [PDF](#).
- 2019 Hickman, L., **Saha, K.**, De Choudhury, M., & Tay, L. (2019). Automated Tracking of Components ML, SIOP of Job Satisfaction via Text Mining of Twitter Data. In Machine Learning Symposium, colocated with Annual Society for Industrial and Organizational Psychology Conference (SIOP 2019), [PDF](#).
- 2019 Lin, S., Mattingly, S.M., Han, X., Audia, P., Campbell, A. T., Das Swain, V., De Choudhury, Future of M., D'Mello, S.K., Dey, A.K., Gao, G., Gregg, J.M., Jagannath, K., Jiang, K., Liu, Q., Mark, G., Work, Martinez, G.J, Masaba, K., Moskal, E., Mulukutla, R., Nies, K.A., Reddy, M.D., Robles-Granda, CHI P., **Saha, K.**, Sirigiri, A., & Striegel, A., Chawla, N.V., (2019). Sensing Personality to Predict Job Performance. In the Future of Work Workshop, colocated with ACM CHI 2019, [PDF](#).
- 2018 **Saha, K.**, Ernala, S., Rizuto, C., Torous, J., & De Choudhury, M. (2018). A Study of Mental Health SMHealth, Awareness Campaigns on Social Media. In Social Media and Health (SMHealth) Workshop, colocated ICWSM with ICWSM 2018.
- 2017 **Saha, K.**, Weber, I., Birnbaum, M.L., and De Choudhury, M. (2017). Characterizing Awareness of MH, CHI Schizophrenia Among Facebook Users by Leveraging Facebook Advertisement Estimates. In Mental Health Symposium, colocated with CHI 2017.

Patents

- 2015 Patel, L., **Saha, K.**, Das, S., & Biswas, A., A System and Method for Sentiment Analysis of Text., *US Provisional Patent Filed, Aug'14, US Patent Filed, Aug'15*.
- 2015 Patel, L., **Saha, K.**, Das, S., & Biswas, A., A System and Method for scalable and continuous learning and annotation of data., *US Provisional Patent Filed, Mar'15*.
- 2014 Bose, A., **Saha, K.**, Das, S., & Biswas, N., A System and Method for Predicting Compliance Issues., *US Provisional Patent Filed, Feb'14*.

Experience

- Aug 2016 – **Graduate Research Assistant**, *Georgia Tech*, School of Interactive Computing.
Advised by Dr. Munmun De Choudhury, as a Ph.D. (Computer Science) student, I specialize in social computing. Broadly motivated in topics and methods related to computational social science and human-centered machine learning. In particular, my research interest involves social media sensing of mental wellbeing in situated communities, such as college campuses and workplaces.
- May 2020 – **Research Intern**, *Snap Research*.
Aug 2020 Advised by Dr. Maarten Bos, I conducted causal studies on understanding the effect of timing advertisements, and how people's momentary user states are associated with ad consumption on social media. This research is motivated to contribute towards better and privacy-preserving understanding of human behavior and improving content recommendation.
- May 2019 – **Research Intern**, *Microsoft Research*.
Aug 2019 Advised by Dr. Amit Sharma, I conducted retrospective case-control based causal study in observational setting. In particular, we investigated the question of *what* factors lead to effective psychosocial outcomes in online mental health communities. We quantified individuals' psychosocial outcomes in an online mental health community, and then compared the social support based interventions received by individuals who improved the most (Case) against those who improved the least (Control).
- May 2018 – **Visiting Scholar**, *Max Planck Institute for Software Systems (MPI-SWS)*.
July 2018 Advised by Dr. Krishna P. Gummadi, I worked on the problem of how social media based advertising platforms enable advertisers to micro-target their audience in a malicious way. We experimentally studied how Facebook ads from the Russian IRA run prior to 2016 U.S. elections exploited Facebook's targeted advertising platform to target ads on divisive or polarizing topics (eg, immigration, race-based policing) at vulnerable sub-populations.
- May 2017 – **Research Intern**, *Fred Hutch*, Hutch Data Commonwealth.
July 2017 Advised by Dr. Naveen Ashish, I worked in the domain of biomedical data sciences and computational social science. My primary research problem included high dimensional machine learning using genomics data to predict the risk of Colorectal Cancer. I also worked on quantifying the 'impact' of tweets to analyze the outreach of Cancer Research information and awareness.
- Dec 2012 – **Senior Researcher**, *Abzooba*, X, NLP & Data Sciences R&D.
Aug 2016 As the product development lead of XPresso, Abzooba's flagship product, improvised techniques, and implemented modules for extracting information in an end-to-end automated, scalable, and efficient fashion. Worked with predictive analytics based business problems from finance, retail, and healthcare sector. Instrumental in conceptualization and implementation of Continuous Learning System (CLS). Ranked 1st at a US based retail giant's Hackathon for Compliance Risk Index (CRIX). Role involved dealing with algorithms and techniques in NLP and Machine Learning, as well as technological advances in software development at production level.
 - NLP problems such as aspect based sentiment analysis, emotion categorization, document classification, semantic search, trending topics detection, ontology building, and so on.
 - NLP techniques like Semantic Role Labeling, Dependency Parsing, Word-Vectors, Topic Modeling, Latent Semantic Hashing.
 - Data Sciences problems like predicting compliance risk, member expenses, distributor costs, etc.
- Sep 2013 – **Visiting Faculty Member**, *Praxis Business School*.
Jul 2014 Conducted courses on Big Data Analytics and Retail Analytics. Subjects included Hadoop, Map Reduce, Unix and Programming fundamentals, and Natural Language Processing.
- Aug 2012 – **Junior Research Associate**, *Infosys*, Infosys Labs, R&D.
Nov 2012 Conceptualized & implemented an information extraction tool from unstructured financial reports. Contributed in ideation of a semantic similarity based recommendation system of research projects.
- May 2011 – **Research Intern**, *Siemens*, Corporate Technologies R&D.
July 2011 Set up a Hadoop cluster of 30 computers, installed Mahout and Katta for running jobs. Developed a tool for accessing Lucene indexes to search and analyze their content based on parameters.

Highlights

- 2020 Nominated as a Finalist for GVU Foley Scholarship Award.
- 2020 Selected for CSCW 2020 Doctoral Consortium.
- 2019 Outstanding Study Design Paper Award at ICWSM 2019.
- 2019 Finalist in Symantec Graduate Research Fellowship Award.
- 2014 Won Data Science Hackathon on Compliance Risk Prediction, organized by a U.S. Retail Giant.
- 2008 Qualified in Top 0.5% at IITJEE 2008 and Top 0.3% at AIEEE 2008.
- 2006 Awarded National Talent Search Examination (NTSE) Scholarship by NCERT, Govt. of India.

Talks and Demos

- 2020 AdverTiming Matters: Examining User Ad Consumption for Effective Ad Reallocation on Social Media, *Snap Research*.
- 2020 Computational and Causal Approaches on Social Media and Multimodal Data For Wellbeing, *TCS Innovation Research Labs*.
- 2019 A Social Media Study on The Effects of Psychiatric Medication Use, *Microsoft Research*.
- 2019 Tesserae: Measuring Role Ambiguity Using LinkedIn Data, *GVU Research Showcase*.
- 2019 Tesserae: Modeling Organizational Culture Using Glassdoor Data, *GVU Research Showcase*.
- 2019 A Social Media Study on The Effects of Psychiatric Medication Use, *GVU Research Showcase*.
- 2018 Tesserae: Multimodal Sensing To Model Individual Differences and Job Performance at Workplaces, *GVU Research Showcase*.
- 2017 Examining Colorectal Cancer Risk using SNPs, *Fred Hutch*.
- 2017 Using Social Media to Model Stress in College Campuses Around Violent Incidents, *GVU Research Showcase*.
- 2016 CampusLife: Understanding Student Mental Wellbeing with Multimodal Sensing, *GVU Research Showcase*.
- 2016 Using Facebook Ad Audience Estimates to Study Psychosis Awareness in the US, *GVU Research Showcase*.
- 2016 An Hybrid Approach for Knowledgebase Buildup and Lookup, *Georgia Tech, COC*.

Miscellaneous

- Skills and Interest Computational Social Science, Causal Inference, Crisis Informatics, Social Computing, Mental Health, Wellbeing, Natural Language, Time Series, Statistical Modeling, Machine Learning, Data Sciences
- Reviewing (Journals) J. Medical Internet Research (JMIR), IEEE Trans. Affective Computing (TAFF), PLOS One, ACM Trans. Computer-Human Interaction (TOCHI), J. Psycholinguistic Research (JOPR), Nature Scientific Reports, ACM Trans. Social Computing (TSC), ACM Trans. Web (TWEB), J. Homosexuality, J. Family Violence, EPJ Data Science, Information Processing & Management (IPM), PACM IMWUT

Reviewing (Conferences) ACM CSCW (2017, 2018, 2019, 2020), ACM CHI (2017, 2018, 2019, 2020*), WebConf (2019, 2020), AAAI ICWSM (2019, 2020), ACII (2019), AACL-IJCNLP (2020) ([* Special Recognition for Outstanding Reviews](#))

SPC/PC Member AAAI ICWSM (2019, 2020, 2021), WebConf (2020), ACM WebSci (2020), ASONAM (2019), EMNLP W-NUT (2019, 2020), ACII ML for Mental Health Workshop (2019)

Org. Comm. Web Chair (AAAI ICWSM 2020)

Mentoring (at Georgia Tech) *Graduate students:* Shrija Mishra, Pranshu Gupta, Asra Yousuf, Manikanta D. Reddy, Benjamin Sugar, Jayant Jain, Sheryl Ratnam, Kunal Agarwal, Amelia Glaese, Andrea Hu, Dong Whi Yoo, Akshay Agarwal, Sherry Yoo; *Undergraduate students:* Jordyn Seybolt, Sarah Yoo, Chaitanya Konjeti, Tushna Eduljee, Yujia Gao, Abhirup Ghosh, Sang-Chan Kim, Tahirah Ahmad, Hemang Rajvanshy, Soo Aah Moon, James O'Connor

Programming Java, Python, R, C, Bash, HTML, CSS, Javascript

Libraries Stanford CoreNLP, LIWC, NLTK, WordNet, Pandas, Scikit-Learn, Vowpal Wabbit

Framework MySQL, MongoDB, BigQuery, Hadoop, Neo4J

Languages English, Bengali, Hindi