Koustuv Saha

Curriculum Vitae

341A, Technology Square Research Building
85 Fifth Street NW, Atlanta, GA 30308, US
☎ *(+1) 404 692 9496*
✉ *koustuv.saha@gatech.edu*
www.koustuv.com

A computer scientist interested in the interdisciplinary area of computational social science and its applications in mental wellbeing and precision medicine, working with real-world problems from natural language processing, machine learning, causal inference analysis, and statistical modeling. Current research involves sensing mental health using social media with a focus on situated communities, such as college campuses and workplaces.

Education

2016– **Doctor of Philosophy Student**, *Computer Science*.
School of Interactive Computing, Georgia Institute of Technology
Advisor: Dr. Munmun De Choudhury, Specialization: Social Computing
- April 2018, Passed Ph.D. Qualifying Examination. Committee: Dr. Gregory D. Abowd (Chair), Dr. Jacob Eisenstein, Dr. Thomas Ploetz, and Dr. Munmun De Choudhury (Advisor)

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

2019 WebSci **Saha, K.**, Chandrasekharan, E., & De Choudhury, M. (2019). Prevalence and Psychological Effects of Hateful Speech in Online College Communities. *To Appear* In Proceedings of the 11th ACM Conference on Web Science, PDF.

2019 ICWSM 🏆 **Saha, K.**, Sugar, B., Torous, J., Abrahao, B., Kıcıman, E., & De Choudhury, M. (2019). A Social Media Study on The Effects of Psychiatric Medication Use. *To Appear* In Proceedings of the 13th International AAAI Conference on Web and Social Media, **Outstanding Study Design Paper**, PDF.

2019 TBM **Saha, K.***, Torous, J.*, Ernala, S. K., Rizuto, C., Stafford, A., & De Choudhury, M. (2019). A Computational Study of Mental Health Awareness Campaigns on Social Media. Translational Behavioral Medicine, *(* co-primary authors)*, Link.

2019 CHI Case Study **Saha, K.**, Bayraktaraglu, A. E., Campbell, A. T., Chawla, N. V., De Choudhury, M., D'Mello, S. K., Dey, A. K., Gao, G., Gregg, J. M., Jagannath, K., Mark, G., Martinez, G. J, Mattingly, S. M., Moskal, 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), PDF.

2019 UbiComp 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, (Non *first four* authors are listed alphabetically).

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 Tesserae 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.

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 UbiComp — **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

2019 Norms, ICWSM — **Saha, K.**, Chandrasekharan, E., & De Choudhury, M. (2019). Hateful Speech in Online College Communities. In Managing and Designing for Norms in Online Communities Workshop, colocated with ICWSM 2019.

2019 ML, SIOP — Hickman, L., **Saha, K.**, De Choudhury, M., & Tay, L. (2019). Automated Tracking of Components 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.

2018 SMHealth, ICWSM — **Saha, K.**, Ernala, S., Rizuto, C., Torous, J., & De Choudhury, M. (2018). A Study of Mental Health Awareness Campaigns on Social Media. In Social Media and Health (SMHealth) Workshop, colocated with ICWSM 2018.

2017 MH, CHI — **Saha, K.**, Weber, I., Birnbaum, M. L., and De Choudhury, M. (2017). Characterizing Awareness of 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 2019 – **Research Intern**, *Microsoft Research*.
Advised by Dr. Amit Sharma, I intend to develop computational methodologies to identify and understand causal markers of intervention that lead to improving mental health in online mental health communities. In addition, I intend to work on the effects of psychiatric drugs in antisocial and violent behavior using large-scale and longitudinal web search logs.

May 2018 – July 2018 **Visiting Scholar**, *Max Planck Institute for Software Systems (MPI-SWS)*.
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 – July 2017 **Research Intern**, *Fred Hutch*, Hutch Data Commonwealth.
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. Also, worked on quantifying the 'impact' of tweets to analyze the outreach of Cancer Research information and awareness.

Dec 2012 – Aug 2016 **Senior Researcher**, *Abzooba Inc.*, X, NLP & Data Sciences R&D.
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 – Jul 2014 **Visiting Faculty Member**, *Praxis Business School*, Analytics.
Conducted courses on Big Data Analytics and Retail Analytics. Subjects included Hadoop, Map Reduce, Unix and Programming fundamentals, and Natural Language Processing.

Aug 2012 – Nov 2012 **Junior Research Associate**, *Infosys Ltd.*, Infosys Labs, R&D.
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 – July 2011 **Research Intern**, *Siemens Information Systems Ltd.*, Corporate Technologies R&D.
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.

Academic Distinctions

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.

Miscellaneous

Skills and Interest	Computational Social Science, Crisis Informatics, Social Computing, Mental Health, Campus Wellbeing, Workplace Wellbeing, Natural Language, Causal Inference, Time Series, Statistical Modeling, Machine Learning, Data Sciences
Reviewing	JMIR, ACM CSCW (2017, 2018, 2019), ACM CHI (2017, 2018, 2019), IEEE TAFF, WebConf (2019), PLOS ONE (2019), ACM TOCHI, AAAI ICWSM (2019), JOPR, Nature Scientific Reports (2019)
PC Member	AAAI ICWSM (2019), ASONAM (2019), ACII ML for Mental Health Workshop (2019)
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