Shubhra Kanti Karmaker "Santu"

Email: sks0086@auburn.edu
Mobile: +1-217-979-3244

Portfolio: https://karmake2.github.io/
Google Scholar: https://scholar.google.com/citations?user=y6pZKT4AAAAJ

CURRENT POSITION

- **Auburn University** — Auburn, Alabama, US
  Assistant Professor, Department of Computer Science and Software Engineering — Jan 2020 - Present

RESEARCH FOCUS AREA

- Natural Language Processing, Information Retrieval, Machine Learning

ACADEMIC PREPARATION

- **Massachusetts Institute of Technology (MIT)** — Boston, Massachusetts, US
  Postdoctoral Research Associate — Jan 2019 - Dec 2019
  Laboratory for Information & Decision Systems (LIDS)
  Host: *Dr. Kalyan Veeramachaneni*

- **University of Illinois at Urbana-Champaign (UIUC)** — Urbana, Illinois, US
  Ph.D. in Computer Science — Aug 2014 - Dec 2018
  Advisor: *Prof. ChengXiang Zhai*
  Thesis Committee Members: Prof. Jiawei Han, Prof. Hari Sundaram, Dr. Hao Ma
  Thesis Title: *"Influence mining from unstructured big data"*

- **Bangladesh University of Engineering and Technology (BUET)** — Dhaka, Bangladesh
  *M.S. in Computer Science & Engineering, CGPA: **4.00** out of 4.00.* — Apr 2012 - Apr 2014

- **Bangladesh University of Engineering and Technology (BUET)** — Dhaka, Bangladesh
  *B.S. in Computer Science & Engineering, CGPA: **3.98** out of 4.00. [**Rank: 1**]* — Jun 2007 - Feb 2012

GRANTS (≈1.4 Million)

- [**AFOSR Awarded $542,485.00 as Sole PI**] **(2023-26)** Basic Research Grant. Program - Information & Networks, Topic - Trust & Influence. Sponsor: Air Force Office of Scientific Research (**AFOSR**). Title: "A Novel Human-AI Collaborative Framework For Multi-Perspective Narrative Analytics & Braiding At Scale". Award #: FA95502310426. Award Link

- [**NSF Awarded $700,854.00 as Lead PI (2023-25)**] Program: Research on Emerging Technologies for Teaching and Learning (RETTL). Sponsor: National Science Foundation (**NSF**). Title: "An Intelligent Assistant to Support Teachers and Students in Simulation-Based Science Learning". Award #: 2302974. Award Link

- [**ARO Awarded $60,000 as Sole PI (2022-23)**] Short-Term Innovative Research (STIR) Grant. Sponsor: Army Research Office (**ARO**). Title: "Semantic Machine For Robust Interpretation Of Noisy Intelligence". Award #: FA95502310426. Award Link

- [**USDA Awarded $50,000 as Lead PI (2022)**] Champion in "Food for Thought" NLP challenge hosted by Coleridge Initiative in collaboration with **USDA**. News 1 News 2.

- [**ASU Awarded $23,135 as Sole PI (2021)**] Deep Time Series Forecasting Techniques for modeling Molecular Dynamics. Sponsor: Arizona State University. Title: "What Does 'Self' Look Like?"

INDUSTRY EXPERIENCE

- **Microsoft Research (Mentor: Riham Mansour)** — Bellevue, Seattle, US
  Summer Research Intern @FUSE labs — Summer, 2018
  Project: Developed semantic representations for understanding utterances for LUIS

- **Microsoft Research (Mentor: Hao Ma)** — Redmond, Seattle, US
  Summer Research Intern @Internet Service Research Center (ISRC) — Summer, 2017
  Project: Automatic Self-Evolving Text Generation

- **Yahoo Research (Mentor: Yi Chang)** — Sunnyvale, California, US
  Summer Research Intern @Search Science Group — Summer, 2016
  Project: Influence Modeling for User Search Behavior

- **@WalmartLabs (Mentor: Parikshit Sondhi)** — Sunnyvale, California, US
  Summer Research Intern @Search Relevance Team — *Summer, 2015*
  Project: Learning-to-Rank for E-Commerce Search
- **Jump Trading LLC (Mentor: Qiaozhu Mei)** — Chicago, Illinois, US
  Text Mining Intern @Search Relevance Team — *Fall 2016 - Spring 2017*
  Project: *Confidential and Restricted*

AWARDS

- **Auburn University Undergraduate Research Mentor Award, 2023**
- **Best Poster Award Nomination at CIKM 2020**
- SIGIR Student Travel Grant for attending CIKM 2018
- Crest of honor (Highest CGPA in the department, presented by BUET alumni association, 2012)
- Dean's List Fellowship (Received fellowship as a distinguished student within the College, 2007-2011)

PUBLISHED RESEARCH (MY STUDENTS ARE UNDERLINED)

- [**EMNLP 2023**]: Souvika Sarkar, Dongji Feng and **Shubhra Kanti Karmaker Santu**. *"Zero-Shot Multi-Label Topic Inference with Sentence Encoders"*. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (to appear soon).

- [**EMNLP 2023**]: Mousumi Akter, Souvika Sarkar and **Shubhra Kanti Karmaker Santu**. *"On Evaluation of Bangla Word Analogies"*. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (to appear soon).

- [**EMNLP 2023**]: **Shubhra Kanti Karmaker Santu** and Dongji Feng. *"TELeR: A General Taxonomy of LLM Prompts for Benchmarking Complex Tasks"*. In Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (to appear soon).

- [**IP&M 2023**]: D. Feng and **Shubhra Kanti Karmaker Santu**. *"Joint Upper & Expected value Normalization for Evaluation of Retrieval Systems: A Case Study with Learning-to-Rank methods"*. Information Processing & Management 60.4 (2023): 103404.

- [**TIST 2023**]: S. Sarkar, B. S. Bijoy, S. J. Saba, D. Feng, Y. Mahajan, S. R. Islam, Md. R. Amin and **Shubhra Kanti Karmaker Santu**. *"Ad-Hoc Monitoring of COVID-19 Global Research Trends for Well-Informed Policy Making"*. ACM Transactions on Intelligent Systems and Technology, 14(2), pp.1-28.

- [**ACL 2022**]: M. Akter, N. Bansal, **Shubhra Kanti Karmaker Santu**. *"Revisiting Automatic Evaluation of Extractive Summarization Task: Can We Do Better than ROUGE?"* In Findings of the Association for Computational Linguistics: ACL 2022 (pp. 1547-1560).

- [**EMNLP 2022**]: N. Bansal, M. Akter, **Shubhra Kanti Karmaker Santu**. *"Learning to Generate Overlap Summaries through Noisy Synthetic Data"*. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 11765-11777).

- [**EMNLP 2022**]: N. Bansal, M. Akter, **Shubhra Kanti Karmaker Santu**. *"SEM-F1: an Automated Metric for Evaluating Multi-Narrative Overlap Summaries"*. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 780-792).

- [**COLING 2022**]: N. Bansal, M. Akter, **Shubhra Kanti Karmaker Santu**. *"Semantic Overlap Summarization among Multiple Alternative Narratives: an Exploratory Study"*. In Proceedings of the 29th International Conference on Computational Linguistics (pp. 6195-6207).

- [**CSUR 2022**]: **Shubhra Kanti Karmaker Santu**, Md. Mahadi Hassan, Micah J. Smith, Lei Xu, ChengXiang Zhai, Kalyan Veeramachaneni. *"AutoML to Date and Beyond: Challenges and Opportunities"*. ACM Computing Surveys (CSUR), 54(8), pp.1-36.

- [**AACL 2022**]: R. Knipper, Md. Mahadi Hassan, Mehdi Sadi and **Shubhra Kanti Karmaker**. *"Analogy-Guided Evolutionary Pretraining of Binary Word Embeddings"*. In Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 683-693).

- [**AACL 2022**]: Souvika Sarkar, Dongji Feng and **Shubhra Kanti Karmaker**. *"Exploring Universal Sentence Encoders for Zero-shot Text Classification"*. In Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (pp. 135-147).

- [**BigData 2022**] : Y. Mahajan, S. R. Islam, Md. R. Amin, **Shubhra K Karmaker Santu**. *"Data-Driven Estimation of Effectiveness of COVID-19 Non-pharmaceutical Intervention Policies"*. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 5312-5321). IEEE.

- [**UserNLP@WWW 2022**]: <u>Souvika Sarkar</u>, **Shubhra Kanti Karmaker Santu**. *"Concept Annotation from Users Perspective: A New Challenge"*. In UserNLP: User-centered Natural Language Processing Workshop @Companion Proceedings of the Web Conference 2022 (pp. 1180-1188).

- [**IUI 2021**] : Biddut S. Bijoy, Syeda J. Saba, <u>Souvika Sarkar</u>, Md Saiful Islam, Sheikh R. Islam, Md. Ruhul Amin, **Shubhra K Karmaker Santu**. *"COVID19α: Interactive Spatio-Temporal Visualization of COVID-19 Symptoms through Tweet Analysis"*. In 26th International Conference on Intelligent User Interfaces-Companion (pp. 28-30).

- [**CIKM 2020 (**Best Poster Nomination**)**] : **Shubhra Kanti Karmaker Santu**, Parikshit Sondhi and ChengXiang Zhai. *"Empirical Analysis of Impact of Query-Specific Customization of nDCG: A Case-Study with Learning-to-Rank Methods"*. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (pp. 3281-3284).

- [**ICWSM 2020**]: Naeemul Hassan, Amrit Poudel, Jason Hale, Claire Hubacek, Khandakar Tasnim Huq, **Shubhra Kanti Karmaker Santu**, Syed Ishtiaque Ahmed. *"Towards Automated Sexual Violence Report Tracking"*. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 14, pp. 250-259).

- [**CoNLL 2019**]: **Shubhra Kanti Karmaker Santu**, Kalyan Veeramachaneni, ChengXiang Zhai. *"Neural Language Models with Evolving Topical Influence"*. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL) (pp. 778-788).

- [**CIKM 2019**]: Saar Kuzi, Sahiti Labhishetty, **Shubhra Kanti Karmaker Santu**, Prasad Pradip Joshi and ChengXiang Zhai. *"Analysis of Adaptive Training for Learning to Rank in Information Retrieval"*. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 2325-2328).

- [**SIGKDD Explorations 2018**]: **Shubhra Kanti Karmaker Santu**, C. Geigle, D. C. Ferguson, W. Cope, M. Kalantzis, D. Searsmith, Chengxiang Zhai. *"SOFSAT: Towards a Setlike Operator based Framework for Semantic Analysis of Text"*. ACM SIGKDD Explorations Newsletter, 20(2), pp.21-30.

- [**CIKM 2018**]: **Shubhra Kanti Karmaker Santu**, Liangda Li, Yi Chang, ChengXiang Zhai. *"JIM: Joint Influence Modeling for Collective Search Behavior"*. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (pp. 637-646).

- [**WPES 2018**]: **Shubhra Kanti Karmaker Santu**, Vincent Bindschaedler, ChengXiang Zhai, Carl A. Gunter. *"NRF: A Naive Re-identification Framework"*. In Proceedings of the 2018 Workshop on Privacy in the Electronic Society (pp. 121-132).

- [**CIKM 2017**]: Yiren Wang, Dominic Seyler, **Shubhra Kanti Karmaker Santu**, ChengXiang Zhai. *"A Study of Feature Construction for Text-based Forecasting of Time Series Variables"*. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 2347-2350).

- [**SIGIR 2017**]: **Shubhra Kanti Karmaker Santu**, Parikshit Sondhi, ChengXiang Zhai . *"On Application of Learning to Rank for E-Commerce Search"*. In Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval (pp. 475-484).

- [**WWW 2017**]: **Shubhra Kanti Karmaker Santu**, Liangda Li, Dae Hoon Park, Yi Chang, ChengXiang Zhai . *"Modeling the Influence of Popular Trending Events on User Search Behavior"*. In Proceedings of the 26th International Conference on World Wide Web Companion (pp. 535-544).

- [**CIKM 2016**]: **Shubhra Kanti Karmaker Santu**, Parikshit Sondhi, ChengXiang Zhai . *"Generative Feature Language Models for Mining Implicit Features from Customer Reviews"*. In Proceedings of the 25th ACM International on conference on Information and knowledge management (pp. 929-938).

- [**IJCNN 2014**]: Md. Mustafizur Rahman, **Shubhra Kanti Karmaker Santu**, Md. Monirul Islam, Kazuyuki Murase. *"Forecasting time series - A layered ensemble architecture"*. In 2014 International Joint Conference on Neural Networks (IJCNN) (pp. 210-217). IEEE.

- [**CEC 2014**]: **Shubhra Kanti Karmaker Santu**, Md. Mustafizur Rahman, Md. Monirul Islam, Kazuyuki Murase . *"Towards better generalization in Pittsburgh learning classifier systems"*. In 2014 IEEE Congress on Evolutionary Computation (CEC) (pp. 1666-1673). IEEE.

PRE-PRINTS (MY STUDENTS ARE UNDERLINED)

- [**Arxiv 2023**]: <u>Mousumi Akter</u> and **Shubhra Kanti Karmaker Santu**. *"FaNS: a Facet-based Narrative Similarity Metric"*. CoRR abs/2309.04823 (2023).

- [**Arxiv 2023**]: <u>Md. Mahadi Hassan</u>, <u>Alex Knipper</u> and **Shubhra Kanti Karmaker Santu**. *"Introducing "Forecast Utterance" for Conversational Data Science"*. CoRR abs/2309.03877 (2023).

- [**Arxiv 2023**]: <u>Yash Mahajan</u>, <u>Naman Bansal</u> and **Shubhra Kanti Karmaker Santu**. *"The Daunting Dilemma with Sentence Encoders: Success on Standard Benchmarks, Failure in Capturing Basic Semantic Properties"*. CoRR abs/2309.03747 (2023)

- [**Arxiv 2023**]: <u>Mousumi Akter</u> and **Shubhra Kanti Karmaker Santu**. *"Redundancy Aware Multi-Reference Based Gainwise Evaluation of Extractive Summarization"*. CoRR abs/2308.02270 (2023)

- [**Arxiv 2023**]: <u>Md. Mahadi Hassan</u>, <u>Alex Knipper</u> and **Shubhra Kanti Karmaker Santu**. *"ChatGPT as your Personal Data Scientist"*. CoRR abs/2305.13657 (2023).

- [**Arxiv 2023**]: <u>Souvika Sarkar</u>, Mohammad Fakhruddin Babar,<u>Md. Mahadi Hassan</u>, Monowar Hasan and **Shubhra Kanti Karmaker Santu**. *"Processing Natural Language on Embedded Devices: How Well Do Modern Models Perform?"*. CoRR abs/2304.11520 (2023)
- [**RWDS 2023**]: <u>Alex Knipper</u>, <u>Naman Bansal</u>, Jingyi Zheng, Wenying Li, and **Shubhra Kanti Karmaker Santu**. *"Food for Thought: First place winners – Auburn Big Data"*. Real World Data Science, August 21, 2023.

Teaching Experience

- **Auburn University**   Auburn, Alabama, US
  Assistant Professor, Department of CSSE   *Jan 2020 - Present*
  <u>Courses Taught</u>: Information Retrieval (Senior Undergrad), Natural Language Processing (Graduate).

- **University of Illinois at Urbana-Champaign (UIUC)**   Urbana, Illinois, US
  Teaching Assistant, Department of Computer Science   *Fall 2017 and 2018*
  <u>Courses Taught</u>: Advanced Information Retrieval (Graduate), Text Mining Seminar (Graduate).

- **Bangladesh University of Engineering and Technology (BUET)**   Dhaka, Bangladesh
  Lecturer, Department of Computer Science & Engineering   *May 2012 - August 2014*
  <u>Courses Taught</u>: Computer Graphics, Computer Architecture, Operating Systems, etc (Undergrad).

Selected External Services

- **Communication Chair**: ACL Rolling Reviews (ARR) Initiative, 2023
- **Tutorial Track Chair**: CIKM, 2022
- **Action Editor**: ACL Rolling Reviews, [2021 - 2023]
- **Panelist**: NSF/CISE Grant Proposal Review Panel, [2020, 2021]
- **PC Member**: EMNLP, ACL, SIGIR, CIKM, WSDM, IUI (Multiple Years: 2017 - 2023)
- **Reviewer**: IEEE Transactions on Knowledge and Data Engineering (TKDE), Knowledge and Information Systems Journal, Neuro-Computing Journal
- **PC Member**: WSDM Demo Track [2019, 2018], IUI Demo Track [2018, 2019]

Invited Talks

- **Future Faculty Workshop Panelist, October 2023**: "Why I chose Academia?"
- **AI@AU Forum, September 2023**: "Democratizing AI with Information Assurance and Knowledge Grounding"
- **Apple Knowledge Platform Tech Talk, February 2023**: "Data Science for 'All' "
- **EMNLP 2022 Oral Presentation**: "SEM-F1: an Automatic Way for Semantic Evaluation of Multi-Narrative Overlap Summaries at Scale"
- **ACL 2022 Oral Presentation**: "Revisiting Automatic Evaluation of Extractive Summarization Task: Can We Do Better than ROUGE?"
- **COLING 2022 Oral Presentation**: "Semantic Overlap Summarization among Multiple Alternative Narratives: an Exploratory Study"
- **NSF EPSCoR Workshop: Artificial Intelligence (AI) with No-Boundary Thinking (NBT), April 2022**: "Data Science for 'All' "
- **Illinois Computer Science Speaker Series, Fall 2021**: "Data Science for 'All' "
- **CIKM 2020 Oral Presentation**: "Empirical Analysis of Impact of Query-Specific Customization of nDCG: A Case-Study with Learning-to-Rank Methods"
- **CIKM 2018 Oral Presentation**: "JIM: Joint Influence Modeling for Collective Search Behavior"
- **SIGIR 2017 Oral Presentation**: "On Application of Learning to Rank for E-Commerce Search"
- **CIKM 2016 Oral Presentation**: "Generative Feature Language Models for Mining Implicit Features from Customer Reviews"

Media Coverage and Expert Opinion

- **Auburn College of Engineering**: Assistant professor in CSSE earns two grants that combine for $1.25 million
- **AL.com**: Rise of the Chatbot: Alabama lawmakers confront questions about artificial intelligence
- **Auburn College of Engineering**: Artificial intelligence expert in CSSE weighs in on the rise of chatbots
- **Real Word Data Science**: FOOD FOR THOUGHT: FIRST PLACE WINNERS – AUBURN BIG DATA

Awards Received by my Students

- **Auburn Outstanding Doctoral Student Award**: Naman Bansal [2023], Mousumi Akter [2023]
- **Auburn University Undergraduate Research Fellowship**: Hugh Williams [2023]
- **100+ Women Strong Outstanding Ph.D. Student**: Souvika Sarkar [2023], Mousumi Akter [2023]
- **Best poster at AU Engineering Research Showcase**: Naman Bansal [2021], Mousumi Akter [2022]

Current Student Advisees

- 10 Ph.D. Students
    - Alex "Ralph" Knipper [2020 - Current]
    - Md. Mahadi Hasan "Sibat" [2020 - Current]
    - Souvika Sarkar [2020 - Current] (Under-Represented in computing)
    - Naman Bansal [2020 - Current]
    - Mousumi Akter [2020 - Current] (Under-Represented in computing)
    - Yash Mahajan [2021 - Current]
    - John Salvador [2022 - Current]
    - Sanjeev Sinha [2022 - Current]
    - Sri Ram Pavan Kumar Guttikonda [2023 - Current]
    - Samariya Nawrin [Incoming] (Under-Represented in computing)
- 2 Undergraduate Students
    - Matthew Freestone (Fall 2023 - Current)
    - Hugh Williams (Spring 2023 - Current)

Alumni

- Ph.D. Alumni
    - Dongji Feng, Summer 2023, assistant professor at Gustavus Adolphus College, Minnesota.
- B.S. Alumni
    - Saksham Goel, Fall 2021, Software Engineer, Microsoft.

Internal Services at Auburn

- Ph.D. Dissertation Committee Member.

  1. Gabrielle Taylor (Graduated Summer, 2023)
  2. Zijie Zhang (Graduated Summer, 2023)
  3. Bo Hui (Graduated Spring, 2023)
  4. Hung Nguyen (Graduated Spring, 2023)
  5. Chengfei Wang (Graduated Fall, 2022)
  6. Chao Jiang (Graduated Summer, 2022)
  7. Kenan Xiao (Graduated Summer, 2022)
  8. Ting Cao (Graduated Summer, 2021)
  9. Wenyu Zhu (Graduated Summer, 2021)

- Ph.D. Dissertation University Reader.

  1. Yuqiao Zhang (Graduated Summer, 2022)
  2. Ziqi Zhou (Graduated Summer, 2021)

- Departmental Service.

  1. CSSE Faculty Search Committee, 2022-23.
  2. CSSE Faculty Search Committee, 2020-21.

- College Level Service.

  1. College of Engineering Graduate Recruiting Committee Member (2023-24)

- University Level Service.

  1. AI@AU Initiative Education Committee.
  2. AI@AU Initiative Research and Equipment Committee Member (2022-24)
  3. AI@AU Initiative Presentation Schedule Organizing Committee Member (2022-24)