Shubhra Kanti Karmaker "Santu"

Email: sks0086@auburn.edu
Portfolio: https://karmake2.github.io/
Mobile: +1-217-979-3244
Google Scholar: https://scholar.google.com/citations?user=y6pZKT4AAAAJ

Current Position

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

Research Area

- Natural Language Processing, Information Retrieval, Machine Learning, Big Data

Academic Preparation

- **Massachusetts Institute of Technology (MIT)** — Boston, Massachusetts, US
  Postdoctoral Research Associate — *Jan 2019 - Dec 2019*
  Laboratory for Information & Decision Systems (LIDS)
- **University of Illinois at Urbana-Champaign (UIUC)** — Urbana, Illinois, US
  Ph.D. in Computer Science — *Aug 2014 - Dec 2018*
  CGPA: **4.00** out of 4.00
- **Bangladesh University of Engineering and Technology (BUET)** — Dhaka, Bangladesh
  M.S. in Computer Science & Engineering — *Apr 2012 - Apr 2014*
  CGPA: **4.00** out of 4.00
- **Bangladesh University of Engineering and Technology (BUET)** — Dhaka, Bangladesh
  B.S. in Computer Science & Engineering — *Jun 2007 - Feb 2012*
  CGPA: **3.98** out of 4.00. [**Class Rank: 1** (out of 126 in the entire department)]

Grants and Awards

- [**AFOSR Awarded $542,485.00 as 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 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 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 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 PI (2021)**] Deep Time Series Forecasting Techniques for modeling Molecular Dynamics. Sponsor: Arizona State University. Title: "What Does 'Self' Look Like?"

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]

Selected External Services

- **Communication Chair**: ACL Rolling Reviews (ARR) Initiative, 2023
- **Tutorial Track Chair**: CIKM, 2022
- **Action Editor**: ACL Rolling Reviews, [2021 - Current]
- **Panelist**: NSF/CISE Grant Proposal Review Panel, [2020, 2021]
- **PC Member**: EMNLP, ACL, SIGIR, CIKM, WSDM, IUI (Multiple Years: 2017 - Current)
- **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]

Previous Industrial Experience

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

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

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

- **@WalmartLabs** — Sunnyvale, California, US
  Summer Research Intern @Search Relevance Team — *Summer, 2015*
  Project: Learning-to-Rank for E-Commerce Search
  Mentor: Parikshit Sondhi

- **Jump Trading LLC** — Chicago, Illinois, US
  Text Mining Consultant @Search Relevance Team — *Fall 2016 - Spring 2017*
  Project: *Confidential and Restricted*

- **Stochastic Logic, ACI Limited** — Dhaka, Bangladesh
  Quantitative Software developer — *May 2012 - August 2014*
  Worked on stock market forecasting using time series analysis

Teaching Experience

- **Auburn University** — Dhaka, Bangladesh
  Assistant Professor, Department of CSSE — *Jan 2020 - Present*
  Courses Taught: Information Retrieval (Senior Undergrad Level), NLP (Graduate Level).

- **University of Illinois at Urbana-Champaign (UIUC)** — Urbana, Illinois, US
  Instructor, Department of Computer Science — *Fall 2018*
  Courses Taught: Text Mining Seminar (Graduate Level).

- **University of Illinois at Urbana-Champaign (UIUC)** — Urbana, Illinois, US
  Teaching Assistant, Department of Computer Science — *Fall 2017*
  Courses Taught: Advanced Information Retrieval (Graduate Level).

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

Published Research (My students are underlined)

- [**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"*. Accepted for publication at the Information Processing and Management Journal, 2023.
- [**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"*. In ACM Transactions on Intelligent Systems and Technology, 2023.
- [**ACL 2022**]: M. Akter, N. Bansal, **Shubhra Kanti Karmaker Santu**. *"Revisiting Automatic Evaluation of Extractive Summarization Task: Can We Do Better than ROUGE?"* In ACL, 2022.
- [**EMNLP 2022**]: N. Bansal, M. Akter, **Shubhra Kanti Karmaker Santu**. *"Learning to Generate Overlap Summaries through Noisy Synthetic Data"*. In EMNLP, 2022.
- [**EMNLP 2022**]: N. Bansal, M. Akter, **Shubhra Kanti Karmaker Santu**. *"SEM-F1: an Automated Metric for Evaluating Multi-Narrative Overlap Summaries"*. In EMNLP, 2022.
- [**COLING 2022**]: N. Bansal, M. Akter, **Shubhra Kanti Karmaker Santu**. *"Semantic Overlap Summarization among Multiple Alternative Narratives: an Exploratory Study"*. In COLING, 2022.
- [**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*". In ACM Computing Surveys, 2022.
- [**AACL 2022**]: R. Knipper, Md. Mahadi Hassan, Mehdi Sadi and **Shubhra Kanti Karmaker**. *"Analogy-Guided Evolutionary Pretraining of Binary Word Embeddings"*. In AACL/IJCNLP, 2022.
- [**AACL 2022**]: Souvika Sarkar, Dongji Feng and **Shubhra Kanti Karmaker**. *"Exploring Universal Sentence Encoders for Zero-shot Text Classification"*. In AACL/IJCNLP, 2022.
- [**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 IEEE BigData, 2022.
- [**UserNLP@WWW 2022**]: Souvika Sarkar, **Shubhra Kanti Karmaker Santu**. *"Concept Annotation from Users Perspective: A New Challenge"*. In UserNLP: User-centered Natural Language Processing Workshop @The ACM Web Conference, 2022.
- [**IUI 2021**] : Biddut Sarker Bijoy, Syeda Jannatus Saba, Souvika Sarkar, Md Saiful Islam, Sheikh Rabiul Islam, Md. Ruhul Amin, **Shubhra K Karmaker Santu**. "*COVID19α: Interactive Spatio-Temporal Visualization of COVID-19 Symptoms through Tweet Analysis*". In ACM IUI, 2021.
- [**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 ACM CIKM, 2020.
- [**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 ICWSM 2020.
- [**CoNLL 2019**]: **Shubhra Kanti Karmaker Santu**, Kalyan Veeramachaneni, ChengXiang Zhai. "Neural Language Models with Evolving Topical Influence". In ACL SIGNLL CoNLL 2019.
- [**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 ACM CIKM 2019.
- [**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". In ACM SIGKDD Explorations [Position Paper] 2018.
- [**CIKM 2018**]: **Shubhra Kanti Karmaker Santu**, Liangda Li, Yi Chang, ChengXiang Zhai. "JIM: Joint Influence Modeling for Collective Search Behavior". In ACM CIKM 2018.
- [**WPES 2018**]: **Shubhra Kanti Karmaker Santu**, Vincent Bindschaedler, ChengXiang Zhai, Carl A. Gunter. "NRF: A Naive Re-identification Framework". In WPES@ACM CCS 2018.

- [**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 ACM CIKM [Short Paper] 2017.
- [**SIGIR 2017**]: **Shubhra Kanti Karmaker Santu**, Parikshit Sondhi, ChengXiang Zhai . "On Application of Learning to Rank for E-Commerce Search". In ACM SIGIR 2017.
- [**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 WWW 2017.
- [**CIKM 2016**]: **Shubhra Kanti Karmaker Santu**, Parikshit Sondhi, ChengXiang Zhai . "Generative Feature Language Models for Mining Implicit Features from Customer Reviews". In ACM CIKM 2016.
- [**IJCNN 2014**]: Md. Mustafizur Rahman, **Shubhra Kanti Karmaker Santu**, Md. Monirul Islam, Kazuyuki Murase. "Forecasting time series - A layered ensemble architecture". In IJCNN 2014.
- [**CEC 2014**]: **Shubhra Kanti Karmaker Santu**, Md. Mustafizur Rahman, Md. Monirul Islam, Kazuyuki Murase . "Towards better generalization in Pittsburgh learning classifier systems". In IEEE CEC (Congress on Evolutionary Computation) 2014.

Ongoing Major Research Projects

- **Multi-Perspective Narrative (MPN) Understanding/Braiding**: In this project, I am developing a novel human-AI collaborative framework called CAMPeN ("Collaborative Analytics of Multi-Perspective Narratives"), where the AI, given multiple alternative narratives as input, first extracts a set of candidate clauses w.r.t. Overlap-Unique-Conflict criteria, separately, in a zero-shot fashion. Next, the human actively verifies clauses that were labeled with low confidence by the AI. Finally, the machine braids the high-confidence/verified clauses to construct the ultimate Overlap-Unique-Conflict style summary, which will be presented to the user with a higher degree of *Information Assurance*.
- **Quantifying Author Trust and Influence**: This project aims to develop (work-in-progress) a novel metric to quantify an author's influence in the braiding process by comparing the final braided narrative against the individual author's contribution in terms of the Overlap-Unique-Conflict clauses extracted by the CAMPeN framework (described above). Over time, influential authors become more trustworthy and serve as trusted sources for future narratives.
- **Ad-Hoc Predictive Skill Acquisition**: In this project, I am currently developing conversational AI techniques which are capable of acquiring new predictive skills on the go through intuitive, natural conversations using self-directed learning. The core technical challenges we are addressing in this project are the following: 1) accurately understand and define the goal skill to acquire (as defined by a teacher) and, consequently, 2) formulate a precise and relevant Machine Learning (ML) task, 3) curate data sets and assign model hyper-parameters accordingly, 4) train AutoML models to learn the skill, and 5) apply the skill effectively. More precisely, we are utilizing Large Language Models (LLMs) to build a natural language interface between the human teachers and the AutoML tools (like Scikit-Learn), which, in turn, facilitates acquiring new prediction skills via self-directed learning.
- **iLab**: This is an ongoing project where we are developing an artificial intelligence-based conversational framework (iLab) to create dialog-based interactive laboratory experiences for middle school science students and teachers in the context of simulation-based science experiments. A key component of the framework is an intelligent conversational agent (SimPal) that engages with teachers in a dialog to solicit their instructional goals associated with simulation experiments and store them using a computational representation. The agent then uses this representation to facilitate and mediate an interactive dialog (powered by state-of-the-art large language models) with students as they run experiments to enhance their learning experience. The agent proactively asks students reflection questions, provide them with real-time customized feedback, tracks students' progress, and then analyzes their responses and reports back to the teacher. Unlike existing intelligent tutoring systems and pedagogical conversational agents, SimPal works with any off-the-shelf third-party simulations, a unique feature of this project.
- **Utility Centric NLP/IR Evaluation**: Previous studies have shown that popular Natural Language Generation (NLG) and Information Retrieval (IR) and evaluation metrics, e.g., $nDCG$, $ROUGE$, MAP, are not robust and often do not correlate with the utility perceived by the humans. In this project, our main goal is to investigate how to make NLP/IR evaluation metrics more utility centric.

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)

Other Honors and Awards

- Best Poster Award Nomination at CIKM 2020
- SIGIR Travel Grant for attending CIKM 2018
- Dean's list award (For brilliant results in each year during the undergraduate program at BUET)
- Crest of honor (Highest CGPA in the department, presented by BUET alumni association)