Swarnadeep Saha

CS PhD Candidate and Google PhD Fellow, UNC Chapel Hill, NC, USA

https://swarnahub.github.io
https://github.com/swarnaHub
swarna@cs.unc.edu 919-641-7845

RESEARCH INTERESTS

Natural Language Processing, Deep Learning, Interpretable AI, Reasoning.

EDUCATION

UNC Chapel Hill — **North Carolina, USA**
Ph.D. in Computer Science, Advisor: Prof. Mohit Bansal — *2019 - Present*
Google PhD Fellowship in NLP for 2023 and 2024
Munroe and Rebecca Cobey Fellowship

Indian Institute of Technology, Delhi — **Delhi, India**
M.Tech. in Computer Science, Advisor: Prof. Mausam, GPA: 9.01/10.0 — *2015 - 2017*
Thesis: Open Information Extraction from Numerical and Conjunctive Sentences
Best M.Tech Thesis Award in Computer Science

Jadavpur University — **Kolkata, India**
B.E. in Computer Science, GPA: 8.72/10.0 — *2010 - 2014*

PUBLICATIONS

1. **MURMUR: Modular Multi-Step Reasoning for Semi-Structured Data-to-Text Generation**
Swarnadeep Saha, Xinyan Velocity Yu, Mohit Bansal, Ramakanth Pasunuru, and Asli Celikyilmaz
Under Review at ACL 2023 [Long][pdf]

2. **Summarization Programs: Interpretable Abstractive Summarization with Neural Modular Trees**
Swarnadeep Saha, Shiyue Zhang, Peter Hase, and Mohit Bansal
Under Review at ICLR 2023 [Long][pdf]

3. **Are Hard Examples also Harder to Explain? A Study with Human and Model-Generated Explanations**
Swarnadeep Saha, Peter Hase, Nazneen Rajani, and Mohit Bansal
EMNLP 2022 [Short][Oral][Acceptance Rate: 12%][pdf]

4. **Explanation Graph Generation via Pre-trained Language Models: An Empirical Study with Contrastive Learning**
Swarnadeep Saha, Prateek Yadav, and Mohit Bansal
ACL 2022 [Long][Poster][Acceptance Rate: 21%][pdf]

5. **ExplaGraphs: An Explanation Graph Generation Task for Structured Commonsense Reasoning**
Swarnadeep Saha, Prateek Yadav, Lisa Bauer, and Mohit Bansal
EMNLP 2021 [Long][Oral][Acceptance Rate: 23%][pdf]

6. **multiPRover: Generating a Set of Proofs for Improved Interpretability in Rule Reasoning**
Swarnadeep Saha, Prateek Yadav, and Mohit Bansal
NAACL 2021 [Long][Oral][Acceptance Rate: 26%][pdf]

7. **PRover: Proof Generation for Interpretable Reasoning over Rules**
Swarnadeep Saha, Sayan Ghosh, Shashank Srivastava, and Mohit Bansal
EMNLP 2020 [Long][Oral][Acceptance Rate: 24%][pdf]

8. **ConjNLI: Natural Language Inference over Conjunctive Sentences**
Swarnadeep Saha, Yixin Nie, and Mohit Bansal
EMNLP 2020 [Long][Poster][Acceptance Rate: 24%][pdf]

9. **Pre-Training BERT on Domain Resources for Short Answer Grading**
Chul Sung, Tejas Dhamecha, **Swarnadeep Saha**, Tengfei Ma, Vinay Reddy, and Rishi Arora
EMNLP 2019 [Short][Poster][Acceptance Rate: 23%][pdf]

10. **Aligning Learning Objectives to Learning Resources: A Lexico-Semantic Spatial Approach**
Swarnadeep Saha, Malolan Chetlur, Tejas I. Dhamecha, Shantanu Godbole and others
IJCAI 2019 [Long][Oral+Poster][Acceptance Rate: 17%][pdf]

11. **Creating Scoring Rubric from Representative Student Answers for Improved Short Answer Grading**
Smit Marvaniya, **Swarnadeep Saha**, Tejas I. Dhamecha, Peter Foltz, Renuka Sindhgatta and Bikram Sengupta
CIKM 2018 [Long][Oral][Acceptance Rate: 17%][pdf]

12. **Joint Multi-Domain Learning for Automatic Short Answer Grading**
Swarnadeep Saha, Tejas I. Dhamecha, Smit Marvaniya, Peter Foltz, Renuka Sindhgatta and Bikram Sengupta
arXiv 1902.09183 [Long][pdf]

13. **Open Information Extraction from Conjunctive Sentences**
Swarnadeep Saha and Mausam
COLING 2018 [Long][Oral][Acceptance Rate: 37%][pdf]

14. **Balancing Human Efforts and Performance of Student Response Analyzer in Dialog-based Tutors**
Tejas I. Dhamecha, Smit Marvaniya, **Swarnadeep Saha**, Renuka Sindhgatta and Bikram Sengupta
AIED 2018 [Long][Oral][Acceptance Rate: 25%][pdf]

15. **Sentence Level or Token Level Features for Automatic Short Answer Grading?: Use Both Swarnadeep Saha**, Tejas I. Dhamecha, Smit Marvaniya, Renuka Sindhgatta and Bikram Sengupta
AIED 2018 [Long][Oral][Acceptance Rate: 25%][pdf]

16. **Bootstrapping for Numerical Open IE**
Swarnadeep Saha, Harinder Pal and Mausam
ACL 2017 [Short][Poster][Acceptance Rate: 18%][pdf]

INDUSTRY EXPERIENCE

FAIR Labs, Meta AI Research — **Seattle, USA**
Research Intern, Mentor: Dr. Asli Celikyilmaz — *May 2022-Dec 2022*
- **Multi-step Reasoning** for **Text generation from semi-structured data**.
- Paper under review at ACL 2023.

Salesforce AI Research — **Palo Alto, USA**
Research Intern, Mentors: Dr. Nazneen Rajani and Dr. Jesse Vig — *June 2021-August 2021*

- Analyzing **Explanation hardness** by connecting **Interpretability** to **Sample hardness**.
- Paper accepted as oral to **EMNLP 2022**.

IBM Research **Bangalore, India**
Research Engineer, Manager: Dr. Shantanu Godbole *July 2017 - June 2019*

- Designed and implemented large scale **Machine Learning** and **NLP** solutions for **Intelligent Tutoring Systems (Watson Tutor)**, notably in the areas of **Automatic Short Answer Grading** and **Text Segmentation**.
- *Lab-wide Research Appreciation* award and twice *Manager's Choice* award.

Adobe Systems **Noida, India**
Member of Technical Staff, Manager: Rajeev Sharma *June 2014 - July 2015*

- Worked as a full-stack software developer in the **Acrobat Reader Team** of Adobe.

ACHIEVEMENTS AND AWARDS

- **Google PhD Fellowship** (one of eight students worldwide) in NLP with full funding for 2 years.
- **Munroe and Rebecca Cobey Fellowship** at UNC Chapel Hill.
- **Best M.Tech. Thesis** in Computer Science at IIT Delhi.
- Lab-wide **Research Appreciation Award** at IBM Research.
- Twice **Manager's Choice Award** at IBM Research.
- **All India Rank of 142** in Graduate Aptitude Test in Engineering (GATE), 2014 among 155190 applicants.
- **State Rank of 96** in West Bengal Joint Entrance Examination (WBJEE), 2010 among 127196 applicants.

PROFESSIONAL SERVICE

- Conference Reviewer: ARR 2022, EMNLP 2022, ARR 2021, EMNLP 2021, NAACL 2021, AAAI 2020, AIED 2019, NAACL 2019, EMNLP 2018.
- Journal Reviewer: AI Journal (AIJ), Computational Linguistics (CL).

SOFTWARE SKILLS

- Programming Languages: C, C++, Java, Scala, Python, Perl, Assembly Languages.
- Databases: MySQL, PostgreSQL.
- Frameworks and Tools: PyTorch, Keras, Hadoop, Git, Perforce, Maven, SBT.

RELEVANT GRADUATE LEVEL COURSES

- Machine Learning, Advanced Machine Learning, Graphical Models, Generative Models, Advanced NLP, Grounding in NLP, Structured Prediction, Machine Learning and Graphics.

REFERENCES

- **Dr. Mohit Bansal**, John R. & Louise S. Parker Associate Professor of CS, UNC Chapel Hill.
- **Dr. Mausam**, Professor, Jai Gupta Chair of CSE and Founding Head of School of AI, IIT Delhi and Affiliate Professor of CS, University of Washington, Seattle.
- **Dr. Asli Celikyilmaz**, Senior Research Manager, Fundamental AI Research (FAIR), Meta.
- **Dr. Shashank Srivastava**, Assistant Professor of CS, UNC Chapel Hill.
- **Dr. Nazneen Rajani**, Robustness Research Lead, Hugging Face.