TUHIN CHAKRABARTY

tuhinjubcse.github.io

219 W 13th Street, NY, NY 10011

(+1)347-630-6405 ⋄ tc2896@columbia.edu

EDUCATION

Columbia University, New York *Present*
PhD in Computer Science (Major in Natural Language Processing) GPA : 4.29
Amazon PhD Fellow
Advisor : **Smaranda Muresan**

Columbia University, New York *August 2017 - 2019*
Master of Science (Major in Natural Language Processing) GPA : 3.91
Advisor : **Kathy Mckeown**
Department of Computer Science

Jadavpur University, Kolkata, India *July 2010 - June 2014*
Bachelor of Engineering in Computer Science GPA : 3.85

AWARDS AND FELLOWSHIPS

- Receipent of Amazon PhD Fellowship for 2022,2023 [1]

- NYC Media Lab x New York Times R&D Researcher Fellowship, 2021

PUBLICATIONS

ARCHIVAL JOURNAL PAPERS

1. **Its not Rocket Science: Interpreting Figurative Language in Narratives**
Tuhin Chakrabarty, Yejin Choi and Vered Shwartz
Transactions of the Association for Computational Linguistics (TACL), 2022 (Journal)

ARCHIVAL CONFERENCE PAPERS

1. **Help me write a Poem - Instruction Tuning as a Vehicle for Collaborative Poetry Writing**
Tuhin Chakrabarty*, Visakh Padmakumar* and He He
Proceedings of The 2022 Conference on Empirical Methods in Natural Language Processing (Long)

2. **FLUTE: Figurative Language Understanding through Textual Explanations**
Tuhin Chakrabarty, Arkadiy Saakyan, Debanjan Ghosh and Smaranda Muresan
Proceedings of The 2022 Conference on Empirical Methods in Natural Language Processing (Long)

3. **Fine-tuned Language Models are Continual Learners**
Thomas Scialom*,Tuhin Chakrabarty* and Smaranda Muresan
Proceedings of The 2022 Conference on Empirical Methods in Natural Language Processing (Long)

---

[1]https://www.amazon.science/academic-engagements/cait-announces-two-new-phd-student-fellowships-and-five-new-faculty-research-awards

4. **Multitask Instruction-based Prompting for Fallacy Detection**
Tariq Alihindi, **Tuhin Chakrabarty**, Elena Musi and Smaranda Muresan
Proceedings of The 2022 Conference on Empirical Methods in Natural Language Processing (Long)

5. **CONSISTENT: Open-Ended Question Generation From News Articles**
Tuhin Chakrabarty, Justin Lewis and Smaranda Muresan
Findings of The 2022 Conference on Empirical Methods in Natural Language Processing (Long)

6. **Don't Go Far Off: An Empirical Study on Neural Poetry Translation**
Tuhin Chakrabarty, Arkadiy Saakyan and Smaranda Muresan
Proceedings of The 2021 Conference on Empirical Methods in Natural Language Processing (Long)

7. **Implicit Premise Generation with Discourse-aware CommonsenseKnowledge Models**
Tuhin Chakrabarty, Aadit Trivedi and Smaranda Muresan
Proceedings of The 2021 Conference on Empirical Methods in Natural Language Processing (Short)

8. **Metaphor Generation with Conceptual Mappings**
Kevin Stowe, **Tuhin Chakrabarty**, Nanyun Peng, Smaranda Muresan and Iryna Gurevych
Proceedings of The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021) (Long)

9. **COVID-Fact: Fact Extraction and Verification of Real-World Claims concerning the COVID-19 pandemic**
Arkadiy Saakyan, **Tuhin Chakrabarty** and Smaranda Muresan
Proceedings of The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021) (Long)

10. **Figurative Language in Recognizing Textual Entailment**
Tuhin Chakrabarty, Debanjan Ghosh, Adam Poliak and Smaranda Muresan
Findings of The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021) (Short)

11. **MERMAID: Metaphor Generation with Symbolism and Discriminative Decoding**
Tuhin Chakrabarty, Xurui Zhang, Smaranda Muresan and Nanyun Peng
Proceedings of The 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Long)

12. **ENTRUST: Argument Reframing with Language Models and Entailment**
Tuhin Chakrabarty, Christopher Hidey and Smaranda Muresan
Proceedings of The 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Long)

13. **Generating similes ~~effortlessly~~ *like a Pro*: A Style Transfer Approach for Simile Generation**
Tuhin Chakrabarty, Smaranda Muresan and Nanyun Peng
Proceedings of The 2020 Conference on Empirical Methods in Natural Language Processing (Long)

14. **Content Planning for Neural Story Generation with Aristotelian Rescoring**
Seraphina Goldfarb-Tarrant,**Tuhin Chakrabarty**, Ralph Weischedel and Nanyun Peng
Proceedings of The 2020 Conference on Empirical Methods in Natural Language Processing (Long)

15. **R^3: Reverse, Retrieve, and Rank for Sarcasm Generation using CommonSense Knowledge**

Tuhin Chakrabarty, Debanjan Ghosh and Smaranda Muresan and Nanyun Peng
Proceedings of The 2020 Annual Conference of the Association for Computational Linguistics (Long)

16. **DeSePtion: Dual Sequence Prediction and Adversarial Examples for Improved Fact-Checking**
Christopher Hidey, **Tuhin Chakrabarty**, Tariq Alhindi, Siddharth Varia ,Kriste Krstovski, Mona Diab and Smaranda Muresan
Proceedings of The 2020 Annual Conference of the Association for Computational Linguistics (Long)

17. **AMPERSAND: Argument Mining for PERSuAsive oNline Discussions**
Tuhin Chakrabarty, Christopher Hidey, Smaranda Muresan, Kathy McKeown and Alyssa Hwang
Proceedings of The 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong
* Accepted as a poster at Natural Language, Dialog and Speech (NDS) Symposium , New York Academy Of Sciences (Long)

18. **IMHO Fine Tuning Improves Claim Detection**
Tuhin Chakrabarty, Christopher Hidey, Kathy McKeown
Proceedings of The 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics, (NAACL) (Short)

19. **Discourse Relation Prediction: Revisiting Word Pairs with Convolutional Networks**
Siddharth Varia, Christopher Hidey, **Tuhin Chakrabarty**
Proceedings of The 20th Annual meeting of Special Interest Group on Discourse and Dialogue (SIGDIAL), 2019 (Long)

WORKSHOP, DEMO

1. **Identifying Distributional Perspective Differences from Colingual Groups**
Yufei Tian, **Tuhin Chakrabarty**, Fred Morstatter, Nanyun Peng
Proceedings of The 9th International Workshop on Natural Language Processing for Social Media, NAACL 2021 (Long)

2. **DiSCoL: Toward Engaging Dialogue Systems through Conversational Line Guided Response Generation**
Sarik Ghazarian, Zixi Liu, **Tuhin Chakrabarty**, Xuezhe Ma, Aram Galstyan, Nanyun Peng
Proceedings of The 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demo)

3. **Pay "Attention" to your Context when Classifying Abusive Language**
Tuhin Chakrabarty, Kilol Gupta and Smaranda Muresan
Proceedings of The 3rd Abusive Language Workshop, ACL 2019 (Long)

4. **The Answer is Language Model Fine-tuning**
Tuhin Chakrabarty and Smaranda Muresan
Proceedings of The 13th International Workshop on Semantic Evaluation, NAACL-HLT, Minneapolis, NAACL 2019 (Short)

5. **Robust Document Retrieval and Individual Evidence Modeling for Fact Extraction and Verification**
Tuhin Chakrabarty, Tariq Alihindi and Smaranda Muresan
Proceedings of The 1st Workshop on Fact Extraction and Verification, EMNLP 2018 (Short)

PAPERS IN SUBMISSION

1. **NORMSAGE: Conversation Grounded Socio-Cultural Norm Discovery**
Yi Fung, **Tuhin Chakrabarty**, Han Guo , Smaranda Muresan and Heng Ji

2. **Predicting and Explaining Message-level Disclosures of Opiate Use Disorder**
Tuhin Chakrabarty, Chenghao Yang and Smaranda Muresan

RESEARCH EXPERIENCE

Research Assistant   *August 2020 - Present*
Advisor: Smaranda Muresan
Columbia University

- Commonsense Reasoning

- Creative Text Generation (Long and Short Form)

- NLP for Public Health

Research Assistant   *August 2019 - August 2020*
Advisor: Nanyun Peng and Smaranda Muresan
Information Sciences Institute, University Of Southern California & Columbia University

- Creative text generation

Research Assistant   *August 2017 - August 2019*
Advisor: Kathy Mckeown and Smaranda Muresan
Columbia University

- Learning from less labeled data for improved Dialogue and Discourse Understanding

- Argumentation (Structures and Semantics)

- Adversarial Attacks in Veracity/ Fact checking models

WORK EXPERIENCE

Summer Intern PhD, Allen Institute of Artificial Intelligence   *May 2021 - August 2021*
Explanatory models for Narrative Understanding involving figurative language.
Mentors : **Vered Shwartz and Yejin Choi**

Summer Intern PhD, IBM Research   *April 2020 - July 2020, YorkTown Heights, NY*
Generative models for Question Answering.

Applied Scientist Intern, Amazon.com   *May 2018 - August 2018, Boston, MA*
Improving classification results of Alexa NLU utterances via diverse data generation using Weighted
Finite State Transducers.

Machine Learning Engineer, UBER   *Jan 2016 - August 2017, Hyderabad, India*
Improving transparency by preventing partner and credit card fraud in emerging markets using machine
learning techniques.

Software Engineer, Amazon.com   *July 2014 - Jan 2016, Hyderabad,India*
Backend Engineer in Customer Returns and Refunds team responsible for handling resolutions such as
Return, Exchange, Replacement.

MENTORING EXPERIENCE

Arkadiy Saakyan (https://asaakyan.github.io/)   *August 2020 - August 2021*
PhD student, Columbia University.
Topics : Counter-claim Generation for Fact-Checking and Neural Machine Translation
Publication Venue : **ACL 2021, EMNLP 2021**

Aadit Trivedi (https://www.linkedin.com/in/aadit-trivedi)   *January 2021 - August 2021*
Georgia Tech Freshman.
Topics : Implicit Premise Generation for Enthymeme Reconstruction
Publication Venue : **EMNLP 2021**

Julia Guo   *August 2021-Present*
Columbia College Senior.
Topics : Neural Machine Translation

INVITED TALKS

Controllability in Figurative Language, COMS 6998 Language Generation and Summarization, September 2022

Fine-tuned Language Models are Continual Learners, ETH Zurich NLP group, August 2022

CONSISTENT: Controllable Open Ended Question Generation from News Articles, New York Times, January 2022

The curious case of Figurative Language University of Utah Data Science Seminar, April 2022

Controlled Generation for Figurative Language and Argument Reframing, ETS Research, November 2021

RELEVANT COURSEWORK

Machine Learning, Natural Language Processing, Deep Learning, NLP in Context: Computational Models of Social Meaning, Analysis Of Algorithm, Fundamentals of Speech Recognition, Conversational AI, Multilingual Technologies and Language Diversity, Language and Mind

PROFESSIONAL ACTIVITIES

- Program Committee at ACL 2021, EMNLP 2021, AAAI 2022, ACL 2020, EMNLP 2020 ,EMNLP 2021, NAACL 2021, *SEM 2021 , AAAI 2020, *SEM 2020

- Program Committee at 1st,2nd and 3rd Workshop on Fact Extraction and Verification

- Program Committee at 4th Abusive Language Workshop.

- Program Committee at 6th Workshop on Noisy User-Generated Text (W-NUT)

REFERENCES

- **Smaranda Muresan**
  Research Scientist, Data Science Institute
  Adjunct Associate Professor
  Columbia University, New York, NY 10027
  Email: sm761@columbia.edu

- **Yejin Choi**
  Brett Helsel Professor
  Office: 578 Allen Center
  Paul G. Allen School of Computer Science & Engineering, University of Washington
  Fax: 206-685-2969
  email: yejin@cs.washington.edu

- **Nanyun Peng**
  Assistant Professor,Department of Computer Science
  University of California, Los Angeles
  Email: violetpeng@cs.ucla.edu

- **He He**
  Assistant Professor,Department of Computer Science
  New York University
  Email: hehe@cs.nyu.edu

- **Kathy McKeown**
  Henry and Gertrude Rothschild Professor of Computer Science
  Columbia University, New York, NY 10027
  Email: kathy.cs@columbia.edu

- **Vered Shwartz**
  Assistant Professor, Dept Of Computer Science
  University of British Columbia
  Email: vshwartz@cs.ubc.ca