

SHIKHA BORDIA

bordiasikhha06@gmail.com • 1-609-907-3642 • [Google Scholar](#) • [Webpage](#) • [LinkedIn](#) • New York

PROFESSIONAL SUMMARY

- 4+ years of research experience with world class research groups; worked in the areas of bias/fairness, reasoning, natural language understanding, natural language generation, reasoning and information extraction.
- Current research focus towards Machine Learning applied to NLP, factuality and evaluation metrics in NLG
- Lead several collaborations with industrial clients or business units and academic partnerships
- Filed 3 US Patents(in process), 7 Publications in top Conferences and workshops (NAACL, EMNLP, SDP)

CURRENT WORK

Verisk Analytics (NLP Research Engineer, R&D Lab, Aug 2019-present, Jersey City)

Research and Development in Machine Reading, Natural Language Generation, Information Extraction.

Collaborated with universities (UNC, IITB, IIITH) and published research papers

- **Information Extraction** Implemented term-tagging module that uses contrastive learning to extract medical specializations with model accuracy of 96% on synthetic data and high recall on domain data (Build Demo)
Deployed NER model of 87% model accuracy to production (Patent to be filed)
- **Concept Search** Implemented concept search pipeline on legal data using fastText representations, attention based sampling, text classification and active learning that incorporates user feedback (Patent under review)
- **Text classification** Deployed xgboost model that uses fastText representations with F1-score 83%
- **Many-hop reasoning** Introduced a challenging dataset and the accompanying evaluation task to encourage research in many-hop fact retrieval and information verification¹
- **Summarization** Augment sentence level SOTA model by varying its submodules
Investigated the influence of intermediate pretraining step on extractive summarization in different domains²
Reproduce and augment transformer based fact aware summarization model (current ongoing project)*

EDUCATION

Courant Institute of Mathematical Sciences, NYU, New York, U.S.A.

MS in Computer Science • Spring 2017 - Spring 2019 • Cumulative GPA: **3.7/4.0**

Machine Learning For Language ([ML²](#), [CILVR Lab](#))^{3,4,5,6,7} Advisor: Samuel Bowman, Lab Head: Yann Lecun

Selected Coursework: Algorithms · Statistical NLP · Programming Languages · Cloud Computing and Machine Learning · NLU And Computational Semantics · Social Networks · Seminar in Semantics

Indian Institute of Technology (IIT), Kharagpur, India

Bachelor in Technology, Fall 2007 - Spring 2011

RESEARCH PROJECTS/PUBLICATIONS

1. HoVer: A dataset for many-hop fact extraction and claim verification, (*Findings at EMNLP, 2020*) [[Paper](#)]
2. The Effect of Pretraining on Extractive Summarization for Scientific Documents (*Proceedings of the Second Workshop on Scholarly Document Processing, 2021*) [[Paper](#)]
3. Identifying And Reducing Gender Bias In Word Level Language Models (*NAACL-SRW 2019; WiML, NeurIPS 2018*) [[Slides](#)][[Paper](#)][[Poster](#)]
4. Measuring Social Biases in Sentence Encoders (*NAACL-HLT 2019*) [[Paper](#)]
5. Do Attention Heads in BERT Track Syntactic Dependencies? (*Natural Language, Dialog and Speech (NDS) Symposium, The New York Academy of Sciences. 2019*) [[Poster](#)][[Paper](#)]
6. Investigating BERT's Knowledge of Language: Five Analysis Methods with NPIs (*EMNLP 2019*) [[Paper](#)]
7. Contributed to [jiant](#) codebase ([ML²](#), [CILVR Lab](#))

PAST WORK EXPERIENCE

Deutsche Bank (Financial Analyst, EMEA Markets, June 2011-Jan 2015, Mumbai/Bangalore)

Received fast-track promotion and firm-wide recognition for excellent client services

- Modeled Structured Finance products (CBOs, CLOs, Synthetics) leading to \$2.5bn in transactions
- Built automated prototypes to perform end-to-end calculations for testing structured products and reporting
- Collaborated with technology teams for Development, UAT and Production of the product

IndustryBuying (Business Intelligence Lead, Mid-level Management, Feb 2015 -Sept 2015, Delhi)

Instrumental in startup's growth from 30 to 300+ employees and from series A to series B funding in a short span; Conceptualized the design and implementation of two crucial e-commerce infrastructure:

- **Order Management System:** Single-handedly led its execution as in a cross-functional role as product owner
Improved order pick up rate and order cycle time by technology integration increasing order fulfillment by 60%
- **Business Intelligence Team Lead:** Built a team of five for BI and actionable analytics
Initiated competitor's pricing wars using web crawling and product categorization

SKILLS:

Python (scikit-learn, pytorch, huggingface, pytorch-lightning, spaCy), SQL, AWS, Git