Jingyu (Jack) Zhang

9 E 33rd St, Ste 616B, Baltimore, MD 21218, USA

jackz.io jzhan237@jhu.edu 443-453-8579 jackjyzhang

EDUCATION

Johns Hopkins University **GPA: 3.99/4.00**

B.S. in Computer Science

Additional Majors: Mathematics; Applied Mathematics & Statistics. Minor: Economics

Expected Graduation: May 2023

- **Relevant Coursework**: Natural Language Processing, Machine Translation, Deep Learning, Machine Learning, Introduction to Algorithms, Honors Real Analysis, Honors Abstract Algebra, Introduction to Optimization
- **GRE**: 331 (V162, Q169, W4.0), **TOEFL**: 115 (R30, L29, S26, W30)

PUBLICATIONS and PREPRINTS Please see personal website for an up-to-date list:

1. **On the Blind Spots of Model-based Evaluation Metrics for Text Generation**
   Tianxing He, **Jingyu Zhang***, Tianle Wang, Sachin Kumar, Kyunghyun Cho, James Glass, Yulia Tsvetkov.*
   Manuscript in Preparation. To appear on arXiv around Dec 21st, 2022. [*Equal Contribution]
2. **PCFG-based Natural Language Interface Improves Generalization for Controlled Text Generation**
   ***Jingyu Zhang**, James Glass, Tianxing He.*
   arXiv preprint. Preliminary version accepted at *2nd Workshop on Efficient Natural Language and Speech Processing (ENLSP), NeurIPS 2022.* ***Best Paper Award**.*
3. **Changes in Tweet Geolocation over Time: A Study with Carmen 2.0**
   ***Jingyu Zhang**, Alexandra DeLucia, Mark Dredze.*
   In *Proc. of the 8th Workshop on Noisy User-generated Text (W-NUT), COLING 2022.*
4. **Study of Manifestation of Civil Unrest on Twitter**
   Abhinav Chinta, **Jingyu Zhang***, Alexandra DeLucia, Anna L. Buzcak, Mark Dredze.*
   In *Proc. of the 7th Workshop on Noisy User-generated Text (W-NUT), EMNLP 2021.* [*Equal Contribution]

AWARDS & HONORS

- **Best Paper Award** - ENLSP Workshop at NeurIPS 2022
- **CRA Outstanding Undergraduate Researcher Award Nominee** - Top 4 out of 744 JHU CS undergraduates
- **Pistritto Research Fellowship** - $4000 grant, Fall 2022. Press coverage:
- **Bloomberg Distinguished Professor (BDP) Summer Program Recipient** - $6000 grant, Summer 2021
- **Upsilon Pi Epsilon** - International Honor Society for the Computing and Information Disciplines
- **Dean's List** - All available semesters
- **National Olympiad in Informatics in Provinces (NOIP)** - National 1st Prize Certification (2018)

RESEARCH EXPERIENCE

Meta-evaluation of Automatic Text Generation Evaluation Metrics **Research Intern | UW/MIT**

Advisors: Yulia Tsvetkov, Jim Glass, and Tianxing He *July 2022 - Present*

- Evaluating the robustness of common automatic text generation metrics based on pretrained LMs
- Discovered "blind spots" in metrics using simple checks based on synthetic errors in consistency and fluency
- Identified inductive bias in pre-trained languages models that caused problems in automatic metrics

Improving Commonsense Reasoning by Stratified Training **Research Intern | JHU**

Advisor: Benjamin Van Durme *May 2022 - Present*

- Constructed a cleaner version of the ATOMIC knowledge graph through few-shot denoising
- Designed a stratified training loss to improve the coherence of sequence-to-sequence models
- Applying stratified training to commonsense models such as COMET for improved commonsense reasoning

Natural Language Interface for Controlled Text Generation **Research Intern | MIT**

Advisors: Jim Glass and Tianxing He *Jan 2022 - June 2022*

- Developed a natural language interface for controlled text generation models using probabilistic context-free grammar to embed desired attributes (sentiment, topic, etc.) into natural language commands
- Designed challenging experiments, and empirically showed the natural language interface enables controlled text generation models to generalize to attributes and attribute combinations unseen during training
- Work in submission, preprint available at `https://arxiv.org/abs/2210.07431`

Study and Modeling of Geolocation Behavior on Twitter **Research Intern | JHU**
Advisor: Mark Dredze *Sep 2021 - May 2022, Sep 2022 - Present*

- Enhanced the performance of Carmen, a Twitter geotagging tool, on non-US and non-English data
- Designed specialized metrics and performed extensive evaluation on the geotagging performance
- Studied how the distribution of geolocation changed over the past decade with the improved Carmen geotagger. Work published in 7th Workshop on Noisy User-generated Text (W-NUT) at COLING 2022
- Developing a location name transduction model that translates informal location strings into formal location entries via tree-based constrained decoding on sequence-to-sequence models trained with Twitter location data

Civil Unrest on Twitter **Research Intern | JHU**
Advisor: Mark Dredze *Dec 2020 - Jun 2021*

- Performed classifier-based predictions and data science analysis of civil unrest events using Twitter data
- Conducted experiments on model interpretability using algorithms such as LIME and SHAP
- Developed a novel feature extraction method based on attention mechanism using BERTweet, which produces an interpretable and stable feature weighting paradigm
- Work published in 7th Workshop on Noisy User-generated Text (W-NUT) at EMNLP 2021

INTERNSHIP EXPERIENCE

ByteDance Ltd. **C++ Development Intern**
Lark Explorer Department *May 2020 - May 2021*

- Carried out C++ cross-platform development interacting with MacOS kernel and Windows Win32 API
- Performed client-side development with Electron and Node.js related to performance optimization
- Conducted data science analytics on extensive user-generated data with Apache Hive and Python
- Part-time after Aug 2020

TEACHING EXPERIENCE

EN.601.465/665 Natural Language Processing **Course Assistant**
JHU Department of Computer Science *Fall 2021, Fall 2022*

- EN.601.465/665 is a mixed graduate / upper-level undergraduate course in NLP taught by Prof. Jason Eisner
- Conducted grading of homework and exam papers
- Held review sessions and office hours on a weekly basis
- Received an average score of 4.80/5.00 on student TA evaluation (100% “Good” or “Excellent” rating)

Code in Place 2021 **Section Leader (Volunteer)**
Stanford University Department of Computer Science *April 2021 - May 2021*

- Worked with a team of more than 50 teaching leads and 1000 section leaders to support 10,000+ students across the world as they navigate the first five weeks of CS 106A: Programming Methodology course
- Prepared materials and taught a Python programming section of 10 students on a weekly basis

OPEN SOURCE PROJECTS

Biomedical Image Captioning

- Implemented a multimodal VGG-LSTM based model to generate medical reports from X-ray images
- Enhanced model performance by utilizing attention and conducted experiments with vision transformer
- Achieved a BLEU score of 0.399 on the IU X-Ray dataset, outperforming baseline BLEU score of 0.152

Fashion Shopping Bot

- Implemented a web scraper with BeautifulSoup and scrapped clothing items from Urban Outfitters, H&M, Abercrombie & Fitch, etc., and constructed a database of fashion items with retrieved data
- Built a search engine for fashion items that supports both text and image queries using FAISS

Multilingual Speech-to-Speech Translation System

- Combined an automatic speech recognition model, a machine translation model, and a text-to-speech model
- Capable of translating speech to speech directly between a large number of languages

TECHNICAL SKILLS

- **Programming Languages**: Python, C/C++, Java, HTML, CSS, Javascript, MATLAB, SQL
- **Frameworks**: Huggingface, PyTorch, Sklearn, Pandas, Numpy, Electron, Windows/macOS native APIs
- **Development Workflow**: LaTeX, Bash, Emacs, Git, Makefile, GN build