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**: 116 (R30, L30, S26, W30)

PUBLICATIONS

PCFG-based Natural Language Interface Improves Generalization for Controlled Text Generation

Jingyu Zhang, James Glass, Tianxing He.

ArXiv preprint.

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.*

Study of Manifestation of Civil Unrest on Twitter

Abhinav Chinta, Jingyu Zhang*, Alexandra DeLucia, Anna L. Buzcak, Mark Dredze.*

*(Equal Contribution)

In *Proc. of the 7th Workshop on Noisy User-generated Text (W-NUT), EMNLP 2021.*

RESEARCH EXPERIENCE

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

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

- Evaluating the robustness of common automatic text generation metrics that are based on pre-trained language models by designing simple adversarial attacks on coherence, fluency, and factuality
- Discovered inductive bias in pre-trained languages models that cause problems in automatic metrics

Improving Coherence of Text Generation Models by Stratified training **Reseach Intern | JHU**

Advisor: Benjamin Van Durme *May 2022 - Present*

- Designing a stratified training loss to improve coherence of generated text for sequence-to-sequence models
- Developed an automatic coherence evaluation method by prompting autoregressive language models

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

Advisor: Jim Glass *Jan 2022 - June 2022*

- Developed a natural language interface for controlled text generation models by using a 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, with preprint available at `https://arxiv.org/abs/2210.07431`

Study and Modeling of Geolocation Behavior on Twitter **Reseach 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
- Studied how the distribution of geolocation has changed during the past decade with the improved Carmen geotagger. Work published in 7th Workshop on Noisy User-generated Text (W-NUT) at COLING 2022
- Designed specialized metrics and performed extensive evaluation on the geotagging performance
- Developing a location name transduction model that translate informal location strings into formal location entries via tree-based constrained decoding on sequence-to-sequence models trained with Twtter location data

Civil Unrest on Twitter **Reseach Intern | JHU**

Advisor: Mark Dredze *Dec 2020 - Jun 2021*

- Machine learning 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**
Department: Lark Explorer *May 2020 - May 2021*

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

TEACHING EXPERIENCE

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

- Instructor: Jason Eisner | Division: Upper Level | Programming Language: Python
- Conducted grading of homework and exam papers
- Held review sessions and office hours on a weekly basis
- Scored 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
- Prepared materials for 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 that 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 the baseline BLEU score of 0.152

Fashion Shopping Bot
- Implemented a web scraper with BeautifulSoup and scrapped clothes 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

AWARDS & HONORS

- **CRA Outstanding Undergraduate Researcher Award Nominee** - 2022
- **Pistritto Research Fellowship** - $4000, Fall 2022
- **Bloomberg Distinguished Professor Summer Program Recipient** - $6000, Summer 2021
- **Upsilon Pi Epsilon** - International Honor Society for the Computing and Information Disciplines
- **Dean's List** - All semesters
- **National Olympiad in Informatics in Provinces (NOIP)** - National 1st Prize Certification (2018)
- **American Mathematics Contest 12** - Global Top 5%
- **American Regions Mathematics League** - Individual Round Regional Top 10

TECHNICAL SKILLS

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