Jingyu (Jack) Zhang

Department of Computer Science · Johns Hopkins University · Baltimore, MD

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

EDUCATION

Johns Hopkins University **2023 - Present**
Ph.D. in Computer Science
Advisors: Daniel Khashabi and Benjamin Van Durme

Johns Hopkins University **2019 - 2023**
B.S. in Computer Science *GPA: 3.97/4.00*
Additional Majors: Mathematics; Applied Mathematics & Statistics. Minor: Economics

PUBLICATIONS & PREPRINTS

1. ***k*-SemStamp: A Clustering-Based Semantic Watermark for Detection of Machine-Generated Text**
   Abe Bohan Hou, ***Jingyu Zhang****, Yichen Wang, Daniel Khashabi, Tianxing He.*
   In Submission.
2. **The *Language Barrier*: Dissecting Safety Challenges of LLMs in Multilingual Contexts**
   Lingfeng Shen, Weiting Tan, Sihao Chen, Yunmo Chen, ***Jingyu Zhang****, Haoran Xu, Boyuan Zheng, Philipp Koehn, Daniel Khashabi.*
   In Submission. *arXiv preprint*: `https://arxiv.org/abs/2401.13136`.
3. **SemStamp: A Semantic Watermark with Paraphrastic Robustness for Text Generation**
   Abe Bohan Hou,* ***Jingyu Zhang***** Tianxing He*, Yichen Wang, Yung-Sung Chuang, Hongwei Wang, Lingfeng Shen, Benjamin Van Durme, Daniel Khashabi, Yulia Tsvetkov.*
   In Submission. *arXiv preprint*: `https://arxiv.org/abs/2310.03991`. [*Equal Contribution]
4. **On the Zero-Shot Generalization of Machine-Generated Text Detectors**
   Xiao Pu, ***Jingyu Zhang****, Xiaochuang Han, Yulia Tsvetkov, Tianxing He.*
   In *Proc. of Findings of EMNLP 2023.*
5. **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.*
   In *Proc. of ACL 2023.* ***Oral Presentation****.* [*Equal Contribution]
6. **Geo-Seq2seq: Twitter User Geolocation on Noisy Data through Sequence to Sequence Learning**
   Jingyu Zhang*, Alexandra DeLucia, Chenyu Zhang, Mark Dredze.*
   In *Proc. of Findings of ACL 2023.*
7. **PCFG-based Natural Language Interface Improves Generalization for Controlled Text Generation**
   Jingyu Zhang*, James Glass, Tianxing He.*
   In *Proc. of *SEM 2023.* Preliminary version accepted at *2nd Workshop on Efficient Natural Language and Speech Processing (ENLSP), NeurIPS 2022.* ***Best Paper Award****.*
8. **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.*
9. **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]

RESEARCH EXPERIENCE

Center for Language and Speech Processing at Johns Hopkins University *Fall 2023 - Present*
PhD Researcher
Advisors: Daniel Khashabi and Benjamin Van Durme

University of Washington *Summer 2022 - Spring 2023*
Undergraduate Research Intern
Advisors: Yulia Tsvetkov and Tianxing He

MIT Computer Science and Artificial Intelligence Laboratory *Spring 2022 - Summer 2022*
Undergraduate Research Intern
Advisors: James Glass and Tianxing He

Center for Language and Speech Processing at Johns Hopkins University *Spring 2021 - May 2023*
Undergraduate Researcher
Advisors: Mark Dredze and Benjamin Van Durme

INDUSTRY EXPERIENCE

ByteDance Ltd. **C++ Development Intern**
Lark Explorer Department *May 2020 - May 2021 (P/T after Summer 2020)*
- 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

AWARDS & HONORS

- **Best Paper Award** - ENLSP Workshop at NeurIPS 2022
- **Michael J. Muuss Research Award** - 1 out of 744 JHU CS undergraduates (2023). Press coverage:
- **CRA Outstanding Undergraduate Researcher Award Nominee** - 4 out of 744 JHU CS undergraduates (2022)
- **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.
- **National Olympiad in Informatics in Provinces (NOIP)** - National 1st Prize Certification (2018)

PRESENTATIONS

SemStamp: A Semantic Watermark with Paraphrastic Robustness for Text Generation
Lightning Talk, JHU CLSP Seminar, Baltimore, Maryland Dec 2023

On the Blind Spots of Model-Based Evaluation Metrics for Text Generation
Oral Presentation, ACL 2023, Toronto, Canada July 2023

PCFG-based Natural Language Interface Improves Generalization for Controlled Text Generation
Spotlight Presentation, ENLSP Workshop, NeurIPS 2022, New Orleans, USA Dec 2022

ACADEMIC SERVICE

Reviewing
- Reviewer, ACL 2023 (Generation track)
- Reviewer, Workshop on Instruction Tuning and Instruction Following, NeurIPS 2023

Outreach
- Application Mentor, JHU CLSP pre-application support program (2023)
- Curriculum Committee, Department of Computer Science, Johns Hopkins University (2023)
- Recruitment Committee, Center for Language and Speech Processing, Johns Hopkins University (2023)

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 and 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

SKILLS & COURSEWORK

Technical Skills
- Programming Languages: Python, C/C++, Java, OCaml, 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

Related Coursework

- NLP/ML: Natural Language Processing, Machine Translation, Multilingual NLP, Artificial Agents, Machine Learning, Deep Learning, Human-Computer Interaction, Probabilistic Models of the Visual Cortex
- Math/Stats: Real Analysis, Abstract Algebra, Topology, Differential Equations, Probability, Statistics, Optimization, Time Series Analysis, Game Theory, Mathematical & Computational Foundations of Data Science