# James (Xiangyu) Dong

## EDUCATION BACKGROUND

**University of Notre Dame, IN** 09/2016 – 05/2020
B.S. in Computer Science | Overall GPA: 3.84/4.0 (Cum Laude) | Notre Dame's Greater China Scholars | Dean's List
**Stanford University, CA** 09/2020 – Present
SCPD NDO program, AI track | Current GPA: 3.94/4.0

**Relevant Coursework:** Multivariable Calculus | Linear Algebra and Differential Equations | Discrete Mathematics | Probability | Data Structures | Design & Analysis of Algorithms | Theory of Computing | Operating System | Quantum Computing | Database Concepts | Computer Architecture | Advanced Machine Learning | Computer Vision | Network Science | Programming Paradigms | Artificial Intelligence | NLP with Deep Learning | Accounting for Managers and Entrepreneurs

## STANDARDIZED TESTS

**GRE General:** V: 162 / Q: 170 / AW: 4.5 / Total: 332

## SKILLS

**Programming & Software:** C/C++ | C# | MATLAB | JavaScript | Python | HTML | CSS | SQL | MongoDB
**Libraries:** Sci. Computing (*NumPy, SciPy*) | Data Manipulation (*Pandas*) | Machine Learning (*Scikit-Learn, PyTorch, Keras, TensorFlow*) | NLP (*NLTK, spaCy, CoreNLP, Texar-PyTorch, Hugging Face*) | Visualization (*Matplotlib, Seaborn*)
**Hobbies:** Tennis | Soccer | Reading | Photography

## PUBLICATIONS

- **Dong, X. Y.**, Yu, W. H., Zhu, C. G., Jiang, M., *Injecting Entity Types into Entity-Guided News Generation*, Accepted to the 2021 Empirical Methods in Natural Language Processing (EMNLP 2021, Oral Presentation)
- Zhang, Y., **Dong, X. Y.**, Rashid, M. T., Shang, L. Y., Han, J., Zhang, D., Wang, D., *PQACNN: Towards Perceptual Quality Assured Single-Image Super-Resolution in Remote Sensing*, 2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS)
- Zhang, Y., **Dong, X. Y.**, Shang, L. Y., Zhang, D., Wang, D. *A Multi-modal Graph Neural Network Approach to Traffic Risk Forecasting in Smart Urban Sensing*, 2020 IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON)
- Zhang, Y., **Dong, X. Y**., Zhang, D., Wang, D., *A Syntax-based Learning Approach to Geo-locating Abnormal Traffic Events using Social Sensing*, 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Vancouver, Canada

## RESEARCH EXPERIENCES

**Data Mining towards Decision Making Laboratory, Notre Dame, IN** 09/2019 – 05/2020
***Research Assistant | Advisor: Prof. Meng Jiang, Dr. Chenguang Zhu (Microsoft Research)***

- Systematically designed, prototyped and characterized an array of natural-language generation (NLG) models for producing a target sequence (i.e., a news story) based on a given list of entities:
  - Studied the impact of using different levels of input entity representations (name, mention, type), models, and mechanisms on text generation quality.
  - Developed a type generation model within sequence-to-sequence (Seq2Seq) learning framework to (1) improve the accuracy of contextual word generation, and (2) maintain entity order in the generated news story, followed by benchmarking the new model against multiple baseline methods on two public datasets, obtaining favorable results; **a first-author research paper accepted to EMNLP 2021, selected Oral Presentation, top 3%** of all submissions

**Notre Dame Social Sensing Laboratory, Notre Dame, IN** 08/2018 – 10/2019
***Research Assistant | Advisor: Prof. Dong Wang***

- Developed SyntaxLoc, a syntax-based probabilistic learning framework using social sensing to extract the

location entities from social media context, focusing on addressing the challenges associated with the limited and unstructured content of social media posts:
  - Applied SyntaxLoc to geo-locate abnormal traffic events in NYC and Los Angeles; crawled traffic accident tweets, followed by extracting locations with n-gram model and Bayesian method based on text syntax content.
  - Benchmarked SyntaxLoc against various state-of-the-art baselines (Google, Stanford NER, spaCy), showing significant improvement in accuracy.
  - A second-author paper presented in the ASONAM'19.
- Developed PQA-CNN, a perceptual quality-assured convolutional neural network framework, to solve the single-image super-resolution (SISR) problem in remote sensing, i.e., reconstructing a high-quality satellite image from a low quality one:
  - Implemented a multi-Gaussian deep neural network to 1) enable SISR for meeting the human perceptual requirement, and 2) quantify the uncertainty level of the synthesized image with the absence of ground truth data.
  - Evaluated PQA-CNN's performance in land usage classification based on a real-world remote sensing application; benchmarked PQA-CNN against an array of state-of-the-art super-resolution models (e.g., SRGAN), showing 5%+ improvement in terms of MSE and MAE metrics.
  - A second-author research paper published at IWQoS'20, **best demo award**.
- Developed GraphCast, a multimodal graph neural network framework, to predict the urban traffic risks at a fine-grained spatial scale by jointly exploiting the data from social media sensing and remote sensing:
  - Developed strategies to 1) reconcile different modalities and characteristics of two data sources, 2) denoise social media data for capturing spatial-temporal dynamics of traffic accident occurrences, and 3) use complex visual features.
  - Experimentally characterized the new framework's performance based on real-world data from NYC, showing its advantage in addressing the limitations of facility availability, privacy, and legal constraints.
  - Published a second-author paper in the SECON 2020; gained familiarity with various tools associated with computer vision, NLP, and graph mining.

**Notre Dame iCeNSA Lab, Notre Dame, IN** 01/2019 – 05/2019
***Research Assistant | Advisor: Prof. Nitesh Chawla***
- Developed a graph-based workflow to mitigate the majority invasion problem associated with training a supervised classifier (e.g., SVM) against an imbalanced dataset:
  - Key steps include 1) building a graph from the imbalanced raw dataset, and 2) employing global neighborhood features of minority nodes for classification.
  - Compared the new method to an array of resampling methods, such as synthetic minority oversampling technique (SMOTE), showing improved AUROC metric for handling two imbalanced datasets.

**Personal Robot Group @ MIT Media Laboratory, MA** 06/2018 – 08/2018
***Undergraduate Research Assistant | Advisor: Prof. Cynthia Breazeal***
- Developed an integrated workflow to enable a social robot learning companion (used in education as a tutor or a peer learner) to perform active role adaptation based on users' feedback:
  - Established a machine learning pipeline to analyze feature importance, covering data cleaning, feature extraction, feature ranking/selection, and classification for 50 robot-human interaction features and facial expression features tied to children's final performance in an interactive game.
  - Provided quantitative guidance to promote a social robot's interactivity and user engagement.

***Deep Multi-Task and Meta Learning*** **Course Project, Stanford University, CA** 09/2020 – 11/2020
- Joined two Stanford graduate students to develop a new method for multi-task representation learning with adversarial regulation.
- Compared the newly developed method to two commonly used multitask representation learning algorithms based on soft-parameter and hard-parameter sharing, respectively.

## WORK EXPERIENCES

***Software Engineer*** **| Advertisement Campaign Platform, Google, Mountain View** 09/2022 – Present
- Onboarding

***Software Engineer*** **| Repo Ticketing and Lifecycle, Bloomberg L.P., NYC** 12/2020 – 09/2022
- Led the revamping of financial figures calculation of Repo ticket by leveraging an internal third-party API; hosted several design meetings within the team and documented project progress; engaged with API owner for question clarifications; prototyped a design to prove the feasibility.
- Acted as the engineering lead of a redesign of an overnight carry-over rate calculation API for Repo (repurchase agreement) ticket; engaged with clients to clarify the functionality of the API; dissected the project into

deliverable stories and prioritized based on Agile development method; followed through the lifecycle of the project development after launching to production.

- Trusted by manager to independently refactor the most heavily used ticketing screen shared across multiple teams. Updated the project from ES5 to TS; broke down the monolith repository (20k lines of code) consisting of several team's business logic into submodules. Reduced bug-related tickets by 30%, while getting a more extendable, testable codebase that significantly increased productivity.
- Introduced messaging middleware RabbitMQ to the ticketing process. Prevented ticketing timeout for online workflow caused by overloaded offline workflow with priority setting. Improved the stability of the ticketing process with the help of RabbitMQ's message acknowledging mechanism.
- Developed a new ticketing screen on a full-stack scale. Organized the code repository with a certain level of abstractness to balance maintainability and extendibility. Migrated the business logic on the front-end to a back-end service to keep the front-end light-weighted and easier for debugging.
- Mentored a new grad pretraining at the team for 3 months. Introduced him to the products, went through key repositories with him, gradually onboarded him with stories of increasing complexity while offering software design advice; the mentee has decided to join the team after training (first since 2019)

***Software Engineer* | Data Intelligence Group, Bloomberg L.P., NYC** 07/2020 – 12/2020

- Contributed toward building a new internal machine learning framework handling the entire development pipeline from data serialization to model deployment that allows more independent model development and more clear business responsibility dissection
- Independently worked on several projects of document classification and entity extraction from financial documents, leveraging tools including:
  - GiGwork and Snorkel for data collection and augmentation
  - PyTorch and sklearn for model development such as LSTM and Graph Convolutional Neural Network
  - Several internal tools for result visualization, and Kubernetes for automatic model deployment

***Software Engineer Intern* | Bloomberg L.P., NYC** 05/2019 – 08/2019

- Enabled memory-sharing among ticker message queues with atomic lock-free queue; allowed less active tickers to be grouped together and freed up memory resources to accommodate more clients, leading to 10X boost in resource utilization efficiency.

## EXTRACURRICULAR EXPERIENCES

***Teaching Assistant* | *Social Sensing / Cyber-Physical System* Course, Notre Dame** 01/2020 – 03/2020

- Held office hours, graded assignments, and managed various logistic items to improve the students' learning experience.
- Developed comprehensive teaching plan and curriculum syllabus with course instructors; practiced an array of course design techniques to enhance teaching effectiveness.

***Alto Saxophonist* | Notre Dame Marching Band & University Band, IN** Fall 2016 – Fall 2018

- Participated in the weekly rehearsal and training.
- Delivered performance on behalf of Notre Dame in an array of on-campus and off-campus events, e.g., the annual performance at DeBartolo Performing Arts Center.

***Team Leader* | Hesburgh Library Hackathon, Notre Dame, IN** 03/2017

- Led a team of three to design, implement and test a software application for generating and cancelling events on Google calendar based on the tone of text conversation with NLTK natural language processing libraries:
  - Employed unified modeling language to guide software design; followed an incremental and iterative workflow with agile retrospectives to drive development process.
  - Leveraged various software engineering concepts and design patterns to promote efficiency, modularity, extensibility, maintainability and robustness of the application.
  - Completed a systematic software testing to verify the functionality of individual components and overall performance.

***Team Leader* | Virtual Reality Internship, Notre Dame, IN** 06/2017 – 08/2017

- Coordinated a team of three to develop a virtual reality game in C# and Unity:
  - Allowed players to navigate through a maze using an HTC Vive helmet for immersive visualization.
  - Designed a robust and computationally efficient random maze creation algorithm using the L-System with local and global constraints as well as the priority queue algorithm, which are capable of generating intertwined road network and diverse buildings while realistically mimicking population distribution and geographic environment.
  - Implemented a pathfinding library using A* and Dijkstra's algorithms