

# Zhuoya SHI, PhD

[Google Scholar](#) | [LinkedIn](#) | [✉zhuoyashi@gmail.com](mailto:zhuoyashi@gmail.com) | ☎(213)706-8729 | 📍Jersey City, NJ

## SUMMARY

---

Highly motivated and detail-oriented researcher with expertise in data analysis, machine learning algorithms, NLP, image & video processing, and data visualization. Proven track record of optimizing processes, achieving 93% accuracy in façade inspection reports analysis and 11% reduction in reporting time. Strong collaboration and leadership abilities. Seeking a data scientist role to apply analytical prowess and domain knowledge to real-world challenges.

## EXPERIENCE

---

### Graduate Student Researcher, NYU, New York, NY

Sep. 2017—May.2024

- Identification of façade inspection taxonomies and defect-component mapping using Natural Language Processing.
  - Developed a generic taxonomy for façade defect by analyzing regulations and expert interviews
  - Established a natural language dataset of 3,000 façade inspection reports using Selenium.
  - Leveraged a machine learning-based approach (bi-LSTM) and achieved 93% accuracy in key information identification.
- Geo-spatial analysis of defect patterns in NYC neighborhoods using spatial clustering approach based on report analysis
  - Extracted façade condition data from historical inspection reports via an NLP analysis.
  - Prepared building general information from open datasets.
  - Conducted a two-step spatial clustering to identify building groups and the existing defect patterns within each group.
- Generation of component-based checklist for inspection guidance with developed algorithm
  - Leveraged UML diagram and capture a domain ontology with major entities, their attributes, relationships, and behaviors .
  - Designed a checklist generation algorithm and implemented in Python, achieving 82% precision and 100% recall.
- Identification of effective integrated visualization techniques in improving inspectors' understanding of façade conditions
  - Defined information requirement and developed low-fidelity prototypes to capture user preference.
  - Implemented high-fidelity prototypes, improving user assessment accuracy by up to 42%.
- Quantification of architecture impact on humans in the built environment
  - Conducted hypothesis testing using R for 356 responses and identified 9 large impact architecture design features.
  - Collected 40 participants' body area sensor data with virtual reality setting to quantify human experiences.

### Graduate Student Mentor, NYU, New York, NY

May 2018—Aug. 2020

#### K-12 STEM Applied Research Innovations in Science and Engineering (ARISE) Programs

- Led Python & machine learning bootcamp for 10+ high school and graduate students in 3 summer sessions.
- Organized brainstorm meetings to troubleshoot research bottlenecks and track research progresses.

### Graduate Research Assistant, USC, Los Angeles, CA

May 2016—May 2017

#### Proactive 2D model-based scan planning for existing buildings

- Developed and implemented the 2D-based building scan planning algorithm and achieved 98.5% scan completeness.
- Collaborated with the research group for project reports updates and journal publication.

## EDUCATION

---

### New York University

Sep. 2017—May.2024

Doctor of Philosophy in [Civil Engineering](#)

New York, NY

Thesis: A BIM-based and data-driven approach for comprehensive façade inspection guidance in cities

### University of Southern California

Sep. 2015—May. 2017

Master of Science in [Construction Engineering](#)

Los Angeles, CA

## TECHNICAL SKILLS

---

**Programming language:** (Proficient)Python, MATLAB, SQL, R; (Familiar) Java, JavaScript, HTML, CSS

**Libraries & Tools:** Pandas, NumPy, Scikit-learn, NLTK, TensorFlow, Keras, OpenCV, Selenium, MongoDB, Git, Tableau

**Methods:** Predictive Modeling, Natural Language Processing, Hypothesis Testing, Statistical Modeling

**Certifications:** Google Data Analytics, DeepMind Deep Learning, DeepMind Natural Language Processing