# Chengwei Ye

(508) 373-3796 | linkedin.com/in/chengwei-ye/ | cye3@wpi.edu
Actively seeking Full-time or Intern opportunities in Data Science

## EDUCATION

**Worcester Polytechnic Institute (WPI)**, *Worcester, MA*
*Master of Science, Data Science, GPA: 4.00/4.00* *May 2020*

**Tongji University**, Shanghai, China
*Bachelor of Engineering in Automation* *July 2018*

## SKILLS

**Programming:** Python, SQL, R, JavaScript, Java, Hive, Spark, Scala, CSS
**Library:** Numpy, Pandas, Scikit-learn, TensorFlow, Matplotlib, Seaborn, D3.js
**Others:** Machine Learning Algorithms, Git, AWS, Salesforce, Tableau, PowerBI

## WORK EXPERIENCE

**Data Analyst Intern**
*Genscape, Boston, MA* *Feb. 2020 - Present*
- Analyzing disparate datasets to improve foundational metadata for Power software applications as well as improving tools to automate this process
- Cleaned data of substations across the United States

**Data Science Co-op** - Capstone Project, WPI
*Homesite Insurance, Boston, MA* *Jan. 2020 - Present*
- Used SQL and Python to explore and extract useful data from 38M data with 500+ features
- Detect anomaly with various supervised and unsupervised methods
- Identify properties where it's less confident to get an accurate replacement cost

**Data Analyst Intern**
*BraunWeiss, Wellesley, MA* *Sept. 2019 - Dec. 2019*
- Extract, clean and import large and complex datasets with Python
- Design ad-hoc data flow and integrate clients' applications with Salesforce
- Design and develop websites using HTML, CSS, JavaScript and Wordpress

**Data Science Co-op** - Capstone Project, WPI
*VMware, Boston, MA* *Aug. 2019 - Dec. 2019*
- Design and implementation of a protocol for privacy-preserving AI
- Built an installable Python package
- Enabled computation of values from multiple encrypted data sources without any party revealing data

## COURSE PROJECTS

**Web-based Data Visualization** *Aug. 2019 - Nov. 2019*
- Visualized weather and aviation data for Logan International Airport with d3.js
- Github: https://github.com/rohame/logan-weather-visualization

**Question-answering System** *Mar. 2019 - Apr. 2019*
- Built a neural network consisting of BiGRU and Attention layers with Tensorflow
- Achieved higher F1 and EM scores than the benchmark of Stanford NLP course

**Airbnb house price prediction**
*Kaggle Competition* *Oct. 2018 - Dec. 2018*
- Used renter information, property attributes and reviews to predict Airbnb rental price
- Used Python and R to explore data, fill null values, encode features and analyze text data
- Trained and tuned linear regression, random forest, gradient boosting and XGBoost models
- Outperformed 90% competitors on private leaderboard