

# Yanjie He

Website: <https://yanjiehe.github.io/>

LinkedIn: Yanjie He

Email : [heyanjie0@outlook.com](mailto:heyanjie0@outlook.com)

Github: YanjieHe

## INTRODUCTION

---

I am an engineer who is interested in Data Science and Machine Learning. Currently, I am a graduate student in the Data Analytics program in the George Washington University. My goal is to be a Data Scientist in the technology field.

## EDUCATION

---

- **The George Washington University** Washington D.C., USA  
*Master of Science in Data Analytics, School of Engineering and Applied Science* Aug. 2018 – Dec. 2020
- **Shanghai University of International Business and Economics** Shanghai, China  
*Bachelor of Arts in Economics; GPA: 3.71/4.00* Sept. 2013 – June. 2017

## WORK EXPERIENCE

---

- **Kantar Media CIC** Shanghai, China  
*Data Engineer Intern* July 2016 - Feb 2017, 8 months
  - **Data Collection:** Designed data collection and data cleaning solution for Chanel APAC project.
  - **Text Mining:** Developed text mining system with team members. The system were used in making data analysis solution for L'Oréal, Chanel, Volkswagen and Dell.
    - \* The system contains a rule parser and an evaluator, which are able to let the data analysts to define the patterns that they want to match.
    - \* The system can provide basic-level functions for sentiment analysis
  - **Data Visualization:** Developed Data Visualization Solution for GroupM television show.

## PROJECTS

---

- **Compiler and Virtual Machine:** The project contains a compiler for a statically typed language, a byte code disassembler and a virtual machine. Besides, I designed a byte code instruction list, which is similar to the JVM instructions. And also, my implementation of the virtual machine taking JVM as reference. The compiler and disassembler are written in C++ and the virtual machine is written in ANSI C.
- **Matrix Library:** The matrix library is implemented in C#, which can be used for matrix computation and solving linear equations. Since the numerical calculation packages on the .Net platform are very limited, some developers used this library as a lightweight choice.
- **Tic Tac Toe A.I.:** Developed a Tic Tac Toe game A.I. by using neural network and reinforcement learning.
- **Reversi A.I.:** Reversi is a board game. Due to the huge amount of possible states of the game, it is next to impossible for computer to enumerate every situation. Therefore, I implemented the minimax algorithm with alpha-beta pruning. The project is written in C++ and has a GUI which is build using Qt5 Framework.
- **Text Co-occurrence Network Analsis for The Hunger Games:** Completed a text analysis project where implemented text co-occurrence network to visualize the relationship between the main characters in the novel *The Hunger Games*. Written the program in Python and used packages including nltk, pandas, matplotlib and wordcloud. Plotted the network by using Gephi.

## PROGRAMMING SKILLS

---

- **Languages:** C, C++, C#, Java, Scala, Python, R, SQL, Scheme/Racket, Html/CSS
- **Technologies:** Data Analysis, Qt 5 Framework, Linux, MySQL, Sqlite3
- **Software:** Gephi, Minitab