

# Thomas Shen

San Jose, CA, 95133 • yutang.shen.mle@gmail.com • (916) 937-4875 • <https://yutan9.github.io/>

## SUMMARY

Data engineer with proven success in delivering data-driven solutions. Proficient in Python, Java, SQL, machine learning, and data structures. Passionate about latest ML technologies.

## EDUCATION

<b>Master's, Computer Science</b>	<i>San José State University</i>	CA, USA	May 2023
<b>Bachelor's, Computer Science</b>	<i>The Chinese University of Hong Kong</i>	Hong Kong	May 2019

## PROFESSIONAL EXPERIENCE

<b>Machine Learning Intern</b> <i>Pure Storage</i>	June 2022–August 2022 Mountain View, CA
---	--

- Estimated products' lifetime with a 3-month error margin with a customized NLP model

<b>Machine Learning Engineer</b> <i>Winbond</i>	January 2021–August 2021 Taiwan
--	------------------------------------

- Enhance marketing department's efficiency by 200% with visualized trend extracted by a NER model
- Led discussions with patent experts, parsed patents with regular expressions, and visualized results
- Built a website with JavaScript to visualize patent relationships in Neo4j database

<b>Software Engineer</b> <i>Cellmax Life</i>	July 2019–November 2019 Taiwan
---	-----------------------------------

- Improved worker efficiency by 96% and increased accuracy with the report generator project
- Boosted efficiency on data querying by 98% with the MySQL database

## PROJECTS

<b><u>Field Lifetime Estimator (FiLteR)</u></b>	<i>Pure Storage</i>
---	---------------------

- Reduced input complexity with Drain3 and TF-IDF algorithms
- Developed customized attention layer class, aligning with the Big Bird attention design
- Delivered a prototype estimator with mean absolute error of 12 weeks

<b><u>Named Entity Recognition</u></b>	<i>Winbond</i>
--	----------------

- Sped up data labeling by 90% by adopting word embedding to tag synonyms
- Fine-tuned ALBERT NER model with customized labels to detect trending technologies in news
- Visualized trend with Power BI word cloud enhancing marketing team's efficiency by 200%

<b><u>MySQL Database</u></b>	<i>Cellmax Life</i>
------------------------------	---------------------

- Designed a MySQL database satisfying medical data storage regulations and BCNF
- Implemented the database and imported existing data in Google Sheet
- Delivered a responsive front-end interface with Vue.js for users to access the database

## SKILLS

<b>Languages</b>	Python, Java, C, SQL, HTML, JavaScript, React, Shell script, Android Java
<b>Technology and tools</b>	Git, Agile Methodology, Docker, AWS, GCP, Linux, FastAPI
<b>Machine Learning</b>	TensorFlow, PyTorch, Transformer, NLTK, Seaborn, PyPlot, OpenCV