Shen-En Chen (Andrew Chen)

Atlanta, GA | +1 (404) 376 0019 | achen353@gatech.edu
www.linkedin.com/in/achen353 | achen353.github.io

EDUCATION

Georgia Institute of Technology — Atlanta, GA

- M.S. in Computer Science (Specialization: Machine Learning), GPA 4.0/4.0 — Aug. 2021 – Dec. 2022
- B.S. in Computer Science (Threads: Intelligence & Info Internetworks), GPA 3.96/4.0 — Aug. 2018 – May 2021

SKILLS

- Programming/Markup Languages: Python, Java, SQL, C, HTML, CSS, JavaScript, LaTeX
- Libraries/Frameworks/Platforms: Numpy, Scipy, Pandas, scikit-learn, PyTorch, Tensorflow/Keras, OpenCV, NLTK, Spacy, gensim, gRPC, Django, FastAPI, Kubernetes, Docker, Google Cloud Platform (GCP), Azure Cognitive Services
- Relevant Coursework: Natural Language Processing, Machine Learning, Computer Vision, Deep Learning in Text

EXPERIENCE

GliaCloud Co., Ltd. — Taipei, Taiwan / Remote

AI Intern — Apr. 2021 – Present

- Fine-tuned a custom multilingual **BERT-CRF keyword model** to improve performances on French and Indonesian by 10% and 13.6%, respectively, while retaining 97% of the performances on already supported English, Mandarin Chinese, Japanese, Korean, and Vietnamese.
- Advised and leveraged **knowledge distillation** and **input embedding reduction** to perform **model compression** on the **transformer model**, reducing the model size by 48% with 97% of performances preserved.
- Proposed the application of **uniform length batching** and **shortest-pack-first histogram-packing (SPFHP)** algorithm and the **refactorization** of tokenization and batching processes, achieving a 20% and 15% reduction in the training and inference time, respectively.
- Built a **GPT-3**-powered copywriting assistant and created **REST APIs** in **Django** for integration.
- Deployed a containerized GAN-based solution to **Google Kubernetes Engine (GKE)** as a microservice with **FastAPI**.

Medical Informatic Research and Genetic Elucidation Lab, National Taiwan University — Taipei, Taiwan

Summer Research Intern — May 2019 – Aug. 2019

- Designed a **machine learning** classification model for 5 common lung tumor types using an ensemble one-vs-one **support vector machine (SVM)** classifier.
- Applied **3D** residual **convolutional neural networks (CNNs)**, using **Keras** and **scikit-learn**, on augmented Lung Image Database Consortium image collection (LIDC-IDRI) to classify benign and malignant lung tumors and achieved an accuracy, sensitivity, and specificity of **97.23%**, **95.54%**, and **98.12%**, respectively.

PROJECTS

Data Augmentation for Entity Matching using Consistency Learning — Atlanta, GA

Team Lead — Sep. 2021 – Present

- Devise a **data augmentation** framework for transformer-based **entity matching** solutions using **consistency learning**.
- Incorporate weighted **Jenson-Shannon divergence** and **semi-supervised learning** to further extend the framework.

Divide-and-Conquer BERT for Legal Document Summarization — Atlanta, GA

Team Lead — Sep. 2021 – December 2021

- Created **extractive summarizers** for US Congressional and California state bills using BERT language models.
- Adopted a **divide-and-conquer** approach that outperforms the benchmark extractive legal document summarization models by 12.41%, 10.95%, and 17.04% F1 score on ROUGE-1, ROUGE-2, and ROUGE-L metrics.

Sentiment Analysis and Topic Modeling on COVID-19 Vaccine Tweets — Atlanta, GA

Team Lead — Jan. 2021 – May 2021

- Performed **exploratory data analysis** and set up a **data preprocessing** script with **NLTK** and **spacy** libraries.
- Utilized **Latent Dirichlet Allocation (LDA)** and **Hierarchical Dirichlet Process (HDP)** to uncover inherent tweet topics.
- **Guided** a team of 5 to carry out **sentiment analysis** with sentiment dictionaries and **bi-directional LSTMs**.

ITS-Chatbot — Atlanta, GA

Team Lead & Member — Jan. 2020 – May 2021

- Established a **retrieval-based chatbot** model using **cosine similarity** on **fastText** word embeddings with **TF-IDF**.
- Incorporated and tested a **BERT** and an **ELECTRA question-answering** model using the **Hugging Face** library.

- **Led** a team of 4 to train and compare **genism LDA**, **Mallet LDA**, **tomotopy LDA**, and **CorEx** topic modeling algorithms on DSP First textbook data and improve them with key term mappings.