YAN CHEN

Phone #: 1 (845) 303-9339 • email: yanchen0219@gmail.com

EDUCATION

The University of Chicago, Chicago, IL **March 2022**
MS, Analytics
CGPA: 3.98/4

Bard College, Annadale-on-Hudson, NY **May 2020**
BA, Mathematics
Thesis: A Multi-Centerpoint Theorem via Fourier Analysis on the Torus
Advisor: Dr. Steven Simon, Assistant Professor of Mathematics
CGPA: 3.6/4
Honors List, 2017

RESEARCH INTERESTS

Generative AI & Human Creativity, Interpretable Machine Learning for Business, Casual Inference, Deep Learning, Time Series Analysis, Reinforcement Learning

RESEARCH EXPERIENCE & PROJECTS

Jude Bellingham AI Pilot : LLM Model Integration for EA Sports **The Mill, Aug 2024 - Sep 2024**

- Prepared and optimized large conversational datasets for LLM fine-tuning, ensuring accuracy and alignment with Jude Bellingham's real-life attributes.
- Implemented and fine-tuned the Llama 3 model with customized settings to achieve a high-quality, personalized model for Jude Bellingham's AI representation.
- Developed API-based connections between the LLM and text-to-speech systems, delivering real-time voice interaction for the AI persona.
- Established seamless LLM integration with Unreal Engine, deploying the Jude figure via a custom server setup to enhance player immersion.

OpenAI Equity Research Summarization Pilot **Macquarie Group, Mar 2023 - Dec 2023**

- Conceptualized and built a Retrieval Augmented Generation system interfacing with the OpenAI APIs using Python.
- Designed and implemented algorithms for question matching, text embedding, and chunking.
- Evaluated Large Language Models (GPT, Hugging Face, BERT) to compare performance in text embedding and tokenization efficiency.
- Researched and explored prospective OpenAI APIs application at Macquarie, including interpreting financial tables from various file formats, facilitating language translations, etc.

Foreign Bank Account Reporting Process Reimagined **Macquarie Group, Mar 2023 - Dec 2023**

- Redesigned the Financial Management Group's FBAR process, enhanced data source integration and quality checks,
- Designed and implemented a data pipeline using Python to perform comprehensive data cleaning and processing; Utilized Hue and AWS S3 to manage database.

Mask Detection **The University of Chicago, Mar 2021 - Jun 2021**

- Developed CNN, MobileNet, and VGG models using image datasets to identify individuals not wearing masks.
- Performed Fine-Tuning techniques on VGG models to achieve 99.1% test accuracy.

Netflix Recommendation System **The University of Chicago, Jan 2021 - Mar 2021**

- Developed a recommendation system to predict how each user will rate all the movies using Ensembled SVD based on over 100M+ customer ratings from the Netflix dataset.
- Established user clusters using k-means and generated recommendation based on movies by using cosine similarity.

A Multi-Centerpoint Theorem via Fourier Analysis on the Torus **Bard College, 2020**

Advised by: Steven Simon

- Developed a multi-centerpoint theorem, aiming to find points in a plane such that any hyperplane passing through them closely partitions a given shape.
- Used Fourier Series and Parseval's Identity to calculate Fourier coefficients and solve optimization problems related to the theorem.

PROFESSIONAL EXPERIENCE

Data Analyst | YZone Creative Studio, New York **Jan 2024 - Present**

- Developed a web scraping algorithm to collect varied product and marketing data from online platforms.
- Conducted thorough NLP analysis to extract actionable insights for leadership team. Designed and executed visualization dashboard for comprehensive representation of sales data.

Data Scientist | Macquarie Group, New York **Nov 2022 - Dec 2023**

- Engaged in a multitude of comprehensive data science and advanced analytics initiatives as a member of Macquarie's Digital Transformation team, contributing significantly to the organization's digitalization endeavors.
- Implemented data-driven solutions to drive process improvement, leading automation efforts in data ingestion and preprocessing, resulting in accelerated delivery timelines and boosted operational efficiency across business groups.
- Bridged the gap between technical and business stakeholders by defining and structuring data-driven solutions.

COMMUNITY INVOLVEMENT

Student Librarian | Bard College, Stevenson Library **Jan 2017 - May 2020**

- Actively engaged with students and faculty members to understand their research needs and objectives, identifying and recommending relevant journals, articles, and books.
- Conducted one-on-one and group sessions for visitors, including students and researchers, on how to navigate and utilize academic databases.

SKILLS

- **Coding Languages:** Python, Java, R, SQL, Spark, Hadoop, MATLAB, LaTeX, HTML
- **Machine Learning Tools:** Numpy, Pandas, Keras, scikit-learn, PyTorch, NLTK, Tensorflow
- **Tools:** Microsoft, Tableau, Power BI, MongoDB, Neo4j, Jupyter Notebook, Dataiku, Google Colab, AWS, Confluence, Google Cloud, Azure, Slack, Jira, Git, Stash

REFERENCES

Steven Simon
Associate Professor of Mathematics, Bard College
ssimon@bard.edu

Greg Green
Director of the MS in Applied Data Science Program, The University of Chicago
greggreen@uchicago.edu

James Williams
Senior Manager, Macquarie Group
James.williams@macquarie.com