JONATHAN FAN

17 Shagbark Ct, Iowa City, IA 52246 ● jonathanfan1256@gmail.com
● (540) 818-7855 ● https://www.linkedin.com/in/jonathanfaniowa/ ● https://www.jonathanfan.site/

EDUCATION

YALE UNIVERSITY — **New Haven, CT**
Department of Mathematics
Bachelor of Science, May 2027
- Emphasis in Economics, Mathematics, Statistics, and Data Science
- Coursework: Linear Algebra, Probability Theory
- Clubs: Yale Alternative Investments, Yale Student Quantitative Research Group, Yale Christian Union
- GPA: N/A

UNIVERSITY OF IOWA — **Iowa City, IA**
College of Liberal Arts and Sciences
Non-degree High School Student, May 2023
- Coursework: Calculus III, Discrete Structures, Introduction to Linear Algebra, Introduction to Numerical Methods: Analysis and Computation
- GPA: 4.00/4.00

PROJECT

AlgoFacto Hedge Fund — **New Haven, CT**
2023-Present **Project Creator**
- Created a python package that allows users to efficiently craft alpha factor gradient-boosted ML models, whether that means designing factors, predicting future returns, tuning hyperparameters, or optimizing portfolio weights
- Designed and Replicated 500+ factors (i.e., macro, value, momentum, etc.) via multi-factor dynamic regression models, PCA eigen-loadings, Cross-sectional K-means clustering, etc.
- Devised 6 profitable trading strategies (robust out-of-sample testing and stress tests), with a focus in ML-Based, Factor-Based, Smart-Beta, Statistical Arbitrage, and Trend-Following Strategies
- Coded fully-automated live-trading executional system using IBKR API and striving to run a strategic asset allocation portfolio of 10-15 profitable strategies live in the upcoming 6 months ($10K Capital)

EXPERIENCE

YALE SCHOOL OF MANAGEMENT — **New Haven, CT**
2023-Present **Research Assistant**
- Programming LLMs and RAG to process thousands of self-trained word embeddings (i.e., NYT, WSJ, etc.) to generate novel Uncertainty Indices correlated with current Uncertainty Measures (i.e., EPU). Researching under Professor Yinan Su's and Professor Leland Bybee's guidance.
- Published a paper with Professor Leland Bybee as a contributor. Created a repo to replicate Bubbles for Fama's Factor Characteristics and industry portfolios from Kenneth French's Data Library.

UNIVERSITY OF IOWA COLLEGE OF NURSING — **Iowa City, IA**
2020-2022 **Lead Algorithm Researcher**
- Published paper on using deep learning to effectively identify and extract symptom information from electronic health records (EHR) to allow physicians to automatically analyze any given EHR
- Executed 100+ computer simulations, self-trained and fine-tuned word embedding models (i.e., Word2Vec, BERT, etc.) on EHR specific datasets, and designed classification models utilizing LSTM-CNN.

YALE UNIVERSITY — **New Haven, CT**
2022-2022 **Research Assistant**
- Published paper under Professor Dragomir Radev's mentorship and his LILY Lab team to create FOLIO (First Order Logic) Dataset
- Assisted in data creation process by writing 20 First-Order-Logic & English stories for dataset, analyzed 180 natural language data structures, and devised conclusions from 75 premises in human performance task

UNIVERSITY OF IOWA COLLEGE OF ENGINEERING — **Iowa City, IA**
2021-2022 **Secondary Student Training Program**
- Crawled twitter data relating to recent tweets regarding the Supreme Court's Decision on overturning Roe vs. Wade using a Search API to analyze human reaction to the decision
- Developed a deep learning model to predict geo locations of where certain tweets came from and ran a sentiment analysis on the texts using the RoBERTa-model

NATIONAL ADVANCED DRIVING SIMULATION — **Iowa City, IA**
2020-2020 **Data Analyst Intern**
- Coded a Batch-Processing method using Python to preprocess and clean unstructured eye-tracking pixel data to determine areas drivers frequently viewed (e.g., phone, road, surrounding areas, etc.)
- Visualized data through graphs that contained quadrants to separate x-y pixel points and processed survey data and used SmartReader (AI Survey Machine) to analyze responses

SKILLS

- Fluent in Python, Ray, Pandas, Numpy, Tensorflow, Keras, PyTorch, Optuna, Scikit-Learn, HTML, CSS, LightGBM, XGBoost, CatBoost, SQL, Asyncio
- Adequate knowledge in machine learning, data visualization, and statistical analysis