ELISA CHEN

+1-415-601-9082 | elisa96chen@gmail.com | linkedin.com/in/elisachenfin | https://github.com/celisa

EDUCATION

Duke University   *August 2022 - Present*
M.S in Data Science, GPA: 3.95   *Durham, NC*
- **Relevant Coursework:** Deep Learning, Data Engineering, Machine Learning, NLP, Time Series, Statistical Models, Reinforcement Learning

University of California, Berkeley   *May 2015 - May 2019*
B.A in Economics & Minor in Computer Science, GPA: 3.5   *Berkeley, CA*
- **Relevant Coursework:** Econometrics, Data Structures, Database Management, Linear Algebra, Discrete Mathematics, Game Theory

TECHNICAL SKILLS

Programming Skills: Python (Pandas, NumPy, PyTorch, HuggingFace, Sklearn, Dask), SQL, Apache Spark, R, Git, Bash

Software Tools: Databricks, AWS, Snowflake, Tableau, Looker (LookML), JIRA, Alteryx, Google Analytics, UI Path

Data Proficiencies: A/B Testing, Regression Analysis, Causal Inference, Deep Learning

Certifications: AWS Cloud Foundations Certificate, AWS Solutions Architect Certificate (In Progress)

Language Skills: English (Native), Finnish (Native), Chinese-Mandarin (Fluent), German (Beginner)

PROFESSIONAL EXPERIENCE

Data Engineer at Chan Zuckerberg Initiative (Investments Team)   *June 2023 - Present*
CZI is an organization focused on building a better future for everyone through technology and investments   *Redwood City, CA*
- Automating key daily operations leading to a time saving of ~10 hours / week for three stakeholders. This was achieved by developing and managing 5+ ETL and integration workflows referencing 5+ APIs using Python DAGs with Apache Airflow
- Improving investor productivity by building an internal LLM tool. Responsibilities include scoping the systems requirements, sourcing potential vendors (e.g fine-tuning platforms, out-of-the-box solutions), and using LangChain to build an internal platform
- Spearheading the build of our data warehouse solution in Databricks from 0 to 1 to support our data analytics and machine learning initiatives by integrating with 3 external data sources, optimizing cluster configurations, and enacting best security practices
- Implemented automated monitoring and error-handling mechanisms in Databricks resulting in 50%+ improvement in downtime

Data Scientist at GoPuff (Product Analytics Team)   *August 2021 - August 2022*
GoPuff is an American multinational consumer goods and food delivery company   *San Francisco, CA*
- Recommended a product strategy on free samples for the marketing team to increase new user retention by ~5% using A/B testing and NLP (text classification) to analyze user feedback on the free samples program
- Led the A/B testing initiative and authored the experimentation requirements document to optimize the homepage banner sequence, successfully boosting order volume by ~$7.5M in a cross-collaboration effort with engineering, design, and product management
- Conducted a causal inference study using a propensity score model in Python to analyze the impact of category promotions on purchasing decisions. This analysis helped shape the product placement strategy of future category promotions
- Developed a SQL script and built a Looker dashboard (utilizing LookML) for leadership to monitor user consent rates and compliance violations, playing a key role in upholding EU GDPR standards and averting potential fines of up to ~$20M

Data Analytics Associate at PwC (Digital Risk Solutions Team)   *August 2019 - August 2021*
PwC DRS team offers professional services with expertise in systems design, data analytics and automation   *San Francisco, CA*
- Expedited software testing time (e.g unit testing) by >50% using SQL queries, Python scripts, and Alteryx workflows to create automated pipelines for big data extraction, validation, and analysis processes for 3 Fortune 500 companies
- Built a transformer-based ML model using NLP techniques like named entity recognition, OCR, and text classification in Python to identify contract clauses and automate the data migration of 15k+ legal contract documents for a financial services client
- Communicated with 10+ stakeholders to develop technical requirements documents and ensure alignment on the project deliverables and goals. Presented data analysis findings in 5+ client deliverables to assist with clients' strategic decision-making

PROJECTS

Large Language Model (LLM) Network Acceleration And Compression Research   *August 2023 - Present*
Tools & Frameworks: NLP (PyTorch, Transformers), AWS Cloud, Hugging Face, Git
Conducting research and experiments on neural network model compression techniques such as quantization, distillation, and pruning on various LLM models with ~3B parameters to increase inference speed by ~30% while preventing an accuracy loss of <3%. The goal of the model is to generalize the detection of malware threats more effectively using fine-tuning and compression techniques.

Humanitarian Data Exchange Project For The United Nations (link to research paper)   *February 2019 - May 2019*
Tools & Frameworks: Python (Scikit-Learn, PyTorch), Pickle, Text Vectorization (GloVe), Text Classification, Git
Developed an NLP text classification algorithm using PyTorch to reduce the manual field tagging effort of end-users by >60%. The goal of the model is to operationalize ad-hoc views and streamline the workflow of tagging datasets. The project was completed together with Microsoft AI Research Team and presented at KDD 2019 (leading international data science conference)