K CHALKLEY

Applied Scientist, Machine Learning Engineer,
Natural Language Processing Specialist

k@kchalk.com
585-209-3390

kchalk @ Github and LinkedIn

FAMILIAR WITH…

- Python, writing Python via Claude
- Generative Language Models, GPT 4+, LLMs, Classification, Entity Recognition, IR, tf-idf, SVM, Deep Neural Networks, linear regression…
- Spacy, Pytorch, HuggingFace, Transformers, nltk, OpenAI API
- Elasticsearch, Lucene, vector search
- Pyspark, Dask, Kubernetes, Microsoft and Amazon cloud platforms
- Science, research, theoretical linguistics

SEEKING…

- Opportunities to relocate or work while traveling outside the US
- Senior individual contributor or tech lead role, applying machine learning techniques to language
- Collaborative, open, and engaged team culture with a balanced approach to speed and quality.

LEAD MACHINE LEARNING ENGINEER (MLE)

Oct 2024 - Current

Clara Analytics

claraanalytics.com

Most used technologies: Python, OpenAI (GPT 4o-mini); AWS cloud services such as S3, Athena, Sagemaker

Technical Lead

- Collaboration with senior leadership, across teams and across projects. Project planning, estimation, resource allocation. Enabling a team to be effective while keeping leadership updated.
- Able to cover for any member of my team across products and tech stacks within 3 months of start. Picked up the work of other MLEs and data scientists and ML Ops code review after workforce reduction.
- Continuous monitoring of multiple products for failures and bug investigations. My team has by far the fastest incident response rate in the company.
- Mine is the final technical evaluation required of any code committed by data scientists and I own the process that prevents model quality regressions.

Claims Doc Intel Pro

- Reduced job failure rate in production environments from 2% to less than 0.1%
- Drastically reduced run times for large documents (e.g. 8 hours to 3 hours) and technical debt

Triage

- Engaged with board members and in-house SMEs to design modeling approaches in support of new features.
- Maintain a collection of custom trained models and artifacts in MLFlow, as well as an anonymized feature store.
- Redesigned anonymization pipelines to be more efficient, more effective, *and* more affordable, including creation of auditable documentation as well as the final implementation.

SENIOR APPLIED SCIENTIST, LEAD MLE — Jan 2023 – Oct 2024

Relativity ODC — relativity.com

Most used technologies: Python, GPT 4(+), Spacy, Scikit-learn; Pyspark, K8s, Azure Cloud, Open telemetry, New Relic, SVM, Classification, Ranking

Review Center – Monitoring, Tech Debt, Scaling

- Designed, implemented, & deployed the company's first model quality monitoring for our active-learning-based classifier. Collaborated legal and data governance committees to navigate requirements concerning client data.
- Resolution of tech debt resulted in reduction from multiple incidents per week to 99% stability, while also providing 10x scalability

Pilot Experimentation – Client Engagement, Objective Evaluation

- Engaged clients to evaluate and tune and advanced access stage modeling project, involving generative models, natural language prompting, and difficult to optimize ranking targets.
- Revised evaluation protocols to provide more comprehensive and generalizable metrics.

Responsible AI Inventory – Risk, Documentation

- Determined model documentation standards in accordance with NIST AI Modeling Risk Management Framework

SENIOR MACHINE LEARNING ENGINEER — Sept 2021 – Dec 2022

SparkCognition Government Systems — sparkgov.ai

Most used technologies: Python, Pandas, Spacy, Scikit-learn, Elastic Search, Azure DevOps, Dask, Document Similarity, Classification

Form Embedding

- Historical records containing heterogeneous data (long & short text, categorical fields, document codes, entities, etc.) embedded for 'similar document' ranking.
- Assist engineering in deploying vectorization apps and nearest neighbor search in databases

Data Interpolation

- Predict missing categorical data from text fields for use in downstream efforts
- Design modeling solutions for small (20k) training set

ENGINEER II, MACHINE LEARNING — Mar 2020 – Aug 2021

Comcast NBC Universal — comcast.com

Most used technologies: Python, PyTorch, AWS EC2 and S3, Gensim, Jenkins, Github, LSTM

Xfinity Assistant Ontology Development

- Redesigned ontology of intent classifications and extracted entity types to rebalance training data and improve scalability.
- Completion of each project milestone resulted in immediate and sustained improvement on key metrics (containment rate, classification accuracy).

...

Logical Form Prediction

- Upgraded model to support variable length utterances
- Improved prediction of misspelled queries from 23% to 42%, while maintaining 80% accuracy on the rest of the domain

Publication: Voice QAC @ EMNLP

Voice Query Auto Completion. Tang et al., EMNLP 2021. https://aclanthology.org/2021.emnlp-main.68

MS COMPUTER SCIENCE 2017-2019

Oregon Health and Science University

Most used technologies: Python, R, PySpark, GGplot, Seaborn, Bokeh

Project: Absolutist Language Use Across Subreddits

- Text representation by dictionary frequency (also LDA)
- T-SNE dimensionality reduction
- Bokeh for data visualization

Project: Aphasia Classification

- Bi-directional LSTM representing actual response and target response
- ARPAbet and CMU Dict phonemic transcription
- 6 class categorizer of error types (phonetic, semantic, mixed, etc.)

BA LINGUISTICS 2010-2015

Reed College

Most relevant skills: Pattern analysis, theory crafting, research papers, data analytics, wrangling people and ideas in meetings, LaTeX

- Thesis -- Applied Asymmetries: Syntax of applicative constructions in Tukang Besi
- Linguistic focus in Syntax, Morphology, Morphosyntactic typology
- Really loved being a nerd at nerd school

VALUES

- Honesty and transparency
- Seeking input from diverse perspectives
- Thinking! Theory, research, creative problem solving, etc.
- Being my whole, best self and doing my best work

K CHALKLEY

Applied Scientist, Machine Learning Engineer,
Natural Language Processing Specialist

k@kchalk.com
585-209-3390

kchalk @ Github and LinkedIn