Feiyang Chen

raechen0717@gmail.com | (616) 227-6289 | Ann Arbor, MI | LinkedIn | Github

EDUCATION

University of Michigan, Ann Arbor Ann Arbor, MI

Bachelor of Science *in Data Science & Mathematics | University Honors 2023- 2025* Dec.2025

TECHNICAL SKILLS

Languages: *Python, SQL, R, C++* | Database Management: *MongoDB*

Data Manipulation & Analysis | Data Visualization: *Tableau, Looker, Matplotlib*

Machine Learning & AI | SDE: *Object-oriented programming (OOP), API development*

Statistical Analysis & Mathematics: *Hypothesis testing, A/B testing, Bayesian, probability theory*

WORK EXPERIENCE

Allada Lab | Michigan Neuroscience Institute Ann Arbor, MI

Data Scientist Jul. 2025 – Present

- ❖ Led the design of end-to-end RNA-seq analysis workflows in Python and R, processing 8 sleep studies (10K–20K features per sample) to support cross-study biological discovery
- ❖ Benchmarked 3 differential expression pipelines (Sleuth, edgeR, meta-analysis framework) by reproducing results and measuring cross-method agreement, revealing 18% variance in gene-level significance calls
- ❖ Standardized multi-study preprocessing and normalization procedures, enabling direct comparison across independent experiments and reducing downstream reprocessing time by ~30%
- ❖ Authored an evaluation report detailing model assumptions, discordant results, and validation metrics, now used by lab researchers to guide analysis strategy selection
- ❖ Drove a large-scale meta-analysis identifying 182 reproducible gene-level signals, increasing statistical power and improving robustness beyond any single-study result

University of Michigan College of Engineering Ann Arbor, MI

Data Engineering Intern Jun. 2024 – Aug. 2024

- ❖ Structured 1,200+ facility images in Amazon S3, reducing retrieval time by ~40%
- ❖ Automated CSV/JSON-to-image linking in Python, saving 10+ hours/week
- ❖ Established standardized naming conventions and validation scripts, lowering labeling inconsistencies by ~25%
- ❖ Performed large-scale image quality screening (resolution, completeness, duplicate detection), removing 15%+ unusable files and increasing dataset reliability
- ❖ Built visual comparison tools for facility change tracking, cutting review time by ~20%

TECHNICAL PROJECTS

Predictive Analytics for Diabetes Progression | R | github.com/RaeChen0717/Diabetes_Progression

- ❖ Modeled disease progression using 500+ patient records and 10+ clinical variables, comparing LASSO, Ridge, and stepwise regression approaches
- ❖ Boosted prediction accuracy by 17% via feature engineering and model tuning
- ❖ Identified significant predictors using ANOVA, improving clinical interpretability
- ❖ Verified generalizability using 10-fold cross-validation and out-of-sample error analysis
- ❖ Created visual summaries of risk factors (BMI, glucose, age), supporting clearer clinical interpretation

Natural Language Text Coherence Analysis | Python, NLP | github.com/RaeC0717/Python-NLP_Language_Analysis

- ❖ Processed 2,000+ machine-generated text sequences to identify semantic coherence and logical discontinuities
- ❖ Conducted qualitative analysis on inconsistent outputs to refine GPT-based text generation models
- ❖ Identified the longest coherent sequence per document (avg. 8 sentences) and addressed limitations in n-gram models
- ❖ Improved coherence scoring by 15% and reduced inconsistencies in long-form text generation by leveraging bigram, trigram, and transformer-based models

Database Query Processing with JDBC | Java & SQL | github.com/RaeC0717/JDBC-DatabaseQueryProcessing

- ❖ Implemented 10+ SQL queries in a Java-based application to efficiently retrieve and process relational data using JDBC
- ❖ Developed 5+ stored procedures to analyze user interactions, friendships, and event participation, optimized query performance using B+ tree indexing, multi-way joins, and sub-queries, achieving a 20% reduction in query execution time
- ❖ Enhanced data retrieval speeds by 2x through engineered advanced SQL solutions within a Java framework