Derek Deming

derekdeming17@gmail.com
760-525-0871
• https://github.com/derekdeming
• https://www.linkedin.com/in/derek-deming/

EDUCATION

Washington State University
Masters in Chemistry with emphasis in Computer Science

- PNNL-WSU Distinguished Graduate Research Program Fellow
- Research Assistant (RA): DOE Fellow on the Data Science and Engineering Traineeship
- Bioinformatics Research Fellow at the Orthopedic Surgery Specialty Clinic

Concordia University Irvine
Graduated with **B.A. Biology and Chemistry** – *Emphasis in Chemical Biology Research*

TECHNICAL SKILLS AND PERSONAL PROJECTS

- **Programming Languages & Frameworks:** Proficient in Python, PySpark, SQL, and Spark SQL, with working knowledge of React.js, Next.js, Django, Node.js, DrizzleORM,, Express, Git, Bash, HTML, Tailwind, CSS, JavaScript, and C++. Experience with RESTful API development, microservices architecture, and server-side rendering

- **Machine Learning Libraries & Techniques:** Expertise in using Pandas, Numpy, Scikit-learn, PyTorch, TensorFlow, Keras, XGBoost, Statsmodels, and Prophet for developing ML algorithms. Advanced understanding of deep learning architectures and algorithms, including CNNs, RNNs, transformers (such as GPT-3 and GPT-4), and transfer learning techniques. Proficient in using large language models for various applications

- **Machine Learning Techniques & Algorithms:** Proficient in a range of machine learning techniques including:
  - **Supervised and Unsupervised Learning:** Experienced in implementing transfer learning, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and clustering algorithms (K Means, Hierarchical Clustering, etc.).
  - **Deep Learning & Transformers:** Deep understanding of state-of-the-art architectures like Transformers and application in text embeddings, generative NLP, and large language models like GPT.
  - **Statistical Analysis & Forecasting:** Experienced in forecasting methodologies utilizing Prophet and XGBoost, and statistical techniques including principal component analysis, logistical regression, and Monte Carlo methods.
  - **Computer Vision & OCR:** Practical experience with Optical Character Recognition and image processing techniques with CNNs.

- **Cloud & Distributed Systems:** Extensive experience with cloud computing platforms (AWS, Azure), High Performance Computing, Docker, and Databricks. Certified as an AWS Databricks Platform Architect and in Databricks Lakehouse Platform Fundamentals.

- **Personal Projects:**

  - **Machine Learning:** Developed an in-depth understanding of GPT / Transformer Architectures, leading to innovative projects involving retrieval augmented generation and deep learning. Explored Large Language Model Deployment and leveraged these models for various Natural Language Processing tasks such as text classification, sentiment analysis, and summarization. Explored cutting-edge techniques like self-supervised learning and adversarial training.

  - **Software Engineering (Full Stack):** Developed complete React.js frontend application with low latency communication between Node.js / express backend. Developed RESTful APIs and designed efficient data storage solutions with SQL and NoSQL databases. Worked on microservices architecture to build scalable and resilient systems. Experience in implementing containerization and orchestration techniques using Docker and Kubernetes

---

EXPERIENCE / RESEARCH

Open-Source Generative AI Research Software Engineer December 2022 to *Present*
Independent Contributor

- Contributed to the development and refinement of open-source projects in the LLM and generative AI space, similar to LangChain, focusing on the practical application and composability of LLMs to build transformative applications in the space of biology and high-throughput experimentation.
- Built out end to end retrieval augmented generation (RAG) pipelines utilizing data framework Llama Index to utilize personalized data with LLMs.
- Designed and implemented machine learning pipelines to support the embedding of biological data for LLMs to interact with. Utilized LLMs to generate novel biological sequences and predict potential functional implications, providing a powerful tool for experimental design and hypothesis testing.
- Implemented advanced features such as callback handlers for the collection of nested run objects, enabling more sophisticated evaluation and testing procedures for LLM-based applications.
- Built websites around implementing LLM and Gen AI technology into the frontend and backend. The tech stack used for this ranges from: React.js, Next.js 13, MongoDB, PostgreSQL, Tailwind, Prisma, OpenAI APIs, Langchain APIs, Llama Index APIs and much more.
- Collaborated in the development of AI-powered chatbots for biological research, providing natural language processing capabilities that can answer complex queries, summarize research findings, and facilitate knowledge sharing within the scientific community.

Data Platform Software Engineer April 2023 to *Present*
Securian Financial

- Infrastructure Management & ML Implementation: Design, build, and manage robust and scalable AI-focused infrastructure, and leverage MLOps principles to implement machine learning solutions into production systems with a focus on continuous integration, testing, and deployment
- Data & ML Architecture: Develop and maintain secure, efficient architectures including databases and large-scale processing systems. Collaborate with data scientists to design and implement robust, scalable machine learning models
- Cloud & Version Control: Use AWS services and GitHub for deploying scalable solutions, version control, code reviews, and collaborative development. Expertise in automating data and ML pipelines, with experience in tools like AWS Glue, Apache Kafka, and Apache Airflow
- Containerization & Microservices: Manage containerized applications and microservices using Docker and Kubernetes, ensuring efficient resource utilization and system performance.

ML Data Scientist / Engineer II (MLOps Specialist) August 2022 to *June 2023*
Swire Coca-Cola, Salt Lake City, Utah

- Overview: Started and grew a team of data scientists and machine learning engineers at a startup initiative inside Coca Cola to handle supply chain issues.
- Machine Learning & Deep Learning: Built and implemented advanced machine learning models for diverse business applications, including sales demand forecasting via cutting-edge deep learning techniques. Designed and developed infrastructure for a trade promotion optimization application using genetic algorithms and Streamlit.
- MLOps & Data Engineering: Developed and streamlined CI/CD pipelines via Jenkins and Azure DevOps for seamless ML model deployment. Used Kubernetes for orchestrating containerized applications and Git for version control. Enhanced security and monitoring practices for robust ML application reliability.
- Real-world Business Solutions: Led projects addressing key business challenges such as **sales demand forecast** using time series analysis with deep learning models, **customer churn prediction** and intervention strategies using NLP analysis of feedback, **customer segmentation** through NLP-based clustering, and **supply chain optimization** using reinforcement learning.
- Technology Evaluation & Implementation: Identified and incorporated new software technologies to improve performance, maintainability, and reliability of ML systems. Tools included MLOps lifecycles, Streamlit, MLflow, Delta Lakes, Docker, Cloud development, Snowpark, Snowflake database, Databricks, and Azure.

Research Data Scientist II August 2020 to September 2022
University of California, Irvine

- Applied statistical and machine learning techniques, including deep learning, to create scalable

simulations for systems of interest. Employed deep learning models such as CNNs for the analysis of molecular structures, and unsupervised learning techniques for clustering and dimensionality reduction. Analyzed and understood large amounts of data for specific conditions and worked closely with collaborators to optimize the complexity of the simulations.

- Implemented atomic-scale molecular dynamics and multi-conformational Monte Carlo simulations, as well as machine learning techniques to simulate protein structures and optimized conformations of the protein structures. This required in-depth statistical analysis as well as dimensionality reduction analysis such as, KNN, regression, clustering, SVMs. The data analysis was performed in both Python and R scripting.
- Used a multiscale molecular simulation approach to gain atomic-level insight into the interprotein interactions that stabilize concentrated solutions of wild-type γ-crystallins and lead to the formation of aggregates in solutions of their cataract-related mutants. Utilized deep learning techniques for protein structure prediction and classification, enabling a better understanding of the underlying mechanisms.
- Translated statistical simulation analysis results to experimentalist collaborators to verify results and discuss further simulations and types of analyses that needed to be completed, including the potential integration of advanced machine learning methods for the interpretation of research findings and the identification of novel therapeutic targets.

Research Data Engineer / Scientist Fellow @ Department of Energy May 2018 to June 2020
Dr. Qiang Zhang, Washington State University, Pullman, WA

- Spearheaded multiple research projects in collaboration with Pacific Northwest National Laboratory to develop a data pipeline between WSU and National Laboratory for continuous research analytics in our computational models.
- Applied various statistical methods in computational design of Metal-Organic Frameworks (MOFs), including regression analysis, Principal Component Analysis (PCA), cluster analysis, machine learning algorithms, Bayesian optimization, Monte Carlo simulations, molecular dynamics simulations, genetic algorithms, and artificial neural networks (ANNs) for property prediction and optimization.
- Utilized these techniques to effectively explore the vast MOF design space, identify structure-property relationships, and guide experimental synthesis efforts towards optimal material designs.
- Deployed genetic algorithms as an optimization technique based on the principles of natural selection and genetics, used to search for optimal MOF structures by evolving a population of candidate materials through selection, crossover, and mutation operations.
- Exploited synthetic crystallographic techniques to understand MOFs as extrapolating agents in solid-solvent phase extractions to improve radioactive waste separations. Leveraged Monte Carlo simulations paired with experimental data to design the most 'optimized' material.
- Mentored undergraduate students pursuing research by teaching them laboratory techniques, conceptual understanding of research tactics, and leading them on projects they found interesting, including the application of advanced machine learning techniques in their research.

Undergraduate Research Assistant January 2016 – May 2018
Dr. John W. Kenney III, Concordia University, Irvine, CA

- Developed a machine learning program using python to determine the difficulty of a college course based on previous grades and the current grade distribution.
- Gained experience with experimental molecular and biomolecular techniques (i.e., DNA isolation, PCR, sub-cloning, microbial transformation, solution/media preparation, aseptic techniques) as well as computational protein modeling and statistical models (utilized python and R).
- Synthesized, purified, and spectroscopically characterized chromium transition metal complexes with acetylacetone, chloride and bromide ligands. Complexes were synthesized using controlled conditions.
- Determined degradation pathway of *Sphingomonas* bacterium of antibiotic resistant bacteria through Spectroscopic Techniques. Compared the localization of human and yeast copper-zinc superoxide dismutase (SOD1) in *Saccharomyces cerevisiae*.

RESEARCH PUBLICATIONS

- Li X., Ding G., Hao L., Deming, D. A,[b] and Qiang Zhang (2020). *ACS Appl. Mater. Interfaces* https://doi.org/10.1021/acsami.0c04961
- Hao, L., Ding, G., Deming, D. A. and Zhang, Q. (2019), *Eur. J. Org. Chem.* doi:10.1002/ejoc.201901303
- Derek Deming et. al (2019) A Facile Method to Introduce Iron Secondary Metal Centers into Metal–Organic Frameworks, *J. Organ. Chem*. doi.org/10.1016/j.jorganchem.2019.06.0

Trainings and Certifications

- Databricks Lakehouse Platform Fundamentals
- AWS Databricks Platform Architect
- AWS Databricks Networking and Security Fundamentals
- Technical Sales: Sales Engineer
- Large Language Models & Langchain & Vector Databases in Production