Kathryn (Katy) Monopoli, PhD

+1 508 367 9077 | krmonopoli@gmail.com | linkedin.com/in/kathrynmonopoli | kmonopoli.github.io

PROFESSIONAL SUMMARY

Machine learning scientist with 10+ years of programming experience and expertise in developing predictive models for drug discovery and therapeutic development. Proven track record building production ready ML pipelines that directly impact pharmaceutical programs, with 2,000+ therapeutic designs and 34 molecules advancing to drug discovery. Strong combination of technical depth, biological domain knowledge, and cross functional collaboration skills. Experienced at bridging computational and experimental teams to deliver actionable insights from complex datasets.

CORE COMPETENCIES

Machine Learning and AI: Python (10+ yrs), TensorFlow, PyTorch, Scikit learn, XGBoost, deep learning, feature engineering, predictive modeling, model optimization, ensemble methods, cross validation, hyperparameter tuning

Data Science: Statistical modeling, high dimensional data analysis, data pipelines, ETL, data cleaning and preprocessing, quality control, experimental design, A/B testing, model validation

Drug Discovery: Therapeutic design, lead optimization, structure activity relationships, high throughput screening analysis, target selection, dose response modeling, translational research

Technical Infrastructure: AWS, Docker, Git, SQL, MySQL, Bash, HPC, cloud computing, CI/CD, database design

Domain Expertise: RNA therapeutics, genomics, transcriptomics, NGS data analysis, protein biophysics, antibody discovery

Soft Skills: Project management, cross functional collaboration, technical communication, team leadership, mentoring

PROFESSIONAL EXPERIENCE

Graduate Researcher, Computational Biology and ML 2019–2025
UMass Chan Medical School, Worcester MA

Machine Learning and Drug Discovery

- Developed and deployed AI and ML models enabling design of 2,000+ therapeutic molecules with 34 advancing to active drug discovery programs; models used across multiple therapeutic areas and disease targets.
- Pioneered a feature engineering approach using deep learning embeddings to capture complex sequence relationships, achieving a 5 fold improvement in predictive accuracy over baseline models.
- Built production ready ML pipelines for high throughput experimental data (NGS and screening assays) with automated QC, feature extraction, model training, and validation to ensure reproducibility and scalability.
- Applied diverse ML techniques including gradient boosting (XGBoost, LightGBM), ensemble methods, regularized linear models, SVMs, and neural networks with rigorous cross validation strategies.
- Optimized models for limited data scenarios by developing data partitioning and training strategies that enable predictive modeling on small biological datasets typical of early stage drug discovery.

Project Leadership and Collaboration

- Led software development as technical project manager, guiding an engineering team to integrate ML models into a production web application and coordinate timelines with experimental and computational stakeholders.
- Collaborated with medicinal chemists, biologists, and clinicians to define requirements, validate models, and translate predictions into experimental designs and go or no go decisions.
- Contributed to 4 patent applications and multiple high impact publications, including a manuscript in preparation for *Nature Biotechnology*.

Data Science and Analysis

- Designed analysis frameworks for statistical modeling, data cleaning, normalization, batch correction, and parameter optimization across diverse experimental platforms.
- Created publication quality visualizations that communicate complex high dimensional relationships to both technical and non technical audiences.
- Developed database schemas and APIs for data management, retrieval, and integration across research programs.

Associate Scientist, Bioinformatics 2016–2018
Advirna (Biotech startup), Cambridge MA

- Improved predictive algorithms for therapeutic design, increasing design throughput 5 fold while improving accuracy.
- Trained interpretable ML models optimized for small datasets using regularized approaches that balance performance and

explainability for stakeholder trust.

- Developed and deployed a cloud based MySQL database platform to streamline data organization, classification, and access across R and D teams while minimizing operational costs.

Researcher, Computational Biophysics 2016
Institute for Protein Design (Baker Lab), Seattle WA

- Implemented software components for the Rosetta protein folding suite in C++.
- Executed large scale molecular modeling campaigns using high throughput cloud computing infrastructure for protein design optimization.

Intern, Antibody Discovery 2013
Biogen, Cambridge MA

- Characterized therapeutic monoclonal antibodies using biophysical assays (BLI, ELISA) for binding, stability, and functionality.
- Developed Python tools for antibody sequence analysis that accelerated epitope mapping and CDR analysis workflows.

EDUCATION

PhD, Computational Biosciences and Bioengineering Nov 2025
UMass Chan Medical School and Worcester Polytechnic Institute
Advisors: Anastasia Khvorova (RNA therapeutics), Dmitry Korkin (Bioinformatics and CS)

MS, Molecular and Cellular Biology 2015
University of Massachusetts Amherst, Phi Beta Kappa, GPA 4.0

BS, Biochemistry and Molecular Biology 2014
University of Massachusetts Amherst, Honors College, *summa cum laude*, GPA 4.0

TECHNICAL SKILLS

Languages: Python, R, Java, C++, SQL, Bash
ML and AI Frameworks: TensorFlow, PyTorch, Scikit learn, XGBoost, LightGBM, Keras
Data Science: Pandas, NumPy, SciPy, Statsmodels, Jupyter
Visualization: Matplotlib, Seaborn, Plotly, ggplot2, Altair
Bioinformatics: Biopython, pysam, HTSeq, DESeq2, Scanpy
Infrastructure: Linux and Unix, Docker, Git and GitHub, AWS (EC2 and S3), Slurm and HPC
Databases: MySQL, PostgreSQL, SQLite
Other: Experimental design, statistical analysis, API development, technical writing

PUBLICATIONS AND PATENTS

Selected Publications

- Monopoli KR, Sostek B, Gross K, Alterman J, Korkin D, Khvorova A. Applying advanced AI methods to define features of functional therapeutic siRNAs. In preparation for *Nature Biotechnology*.
- Davis SM, Hildebrand S, Monopoli KR, et al. Systematic analysis of siRNA and mRNA features impacting fully chemically modified siRNA efficacy. *Nucleic Acids Research*, 2025.
- Monopoli KR, Korkin D, Khvorova A. Asymmetric trichotomous data partitioning enables predictive ML models using limited siRNA efficacy datasets. *Molecular Therapy: Nucleic Acids*, 2023.

Full publication list: Google Scholar profile

Patents

- Oligonucleotides for MAPT modulation (US 17/204,480)
- Oligonucleotides for SNCA modulation (US 17/204,483)
- Oligonucleotides for MSH3 modulation (US 63/012,603)
- Oligonucleotides for SARS-CoV-2 modulation (US 17/333,839)

LEADERSHIP AND PROFESSIONAL DEVELOPMENT

Software Development Project Manager 2023–2025
UMass Chan Medical School

- Managed the full product development lifecycle for an ML powered therapeutic design platform.
- Led technical recruitment, hiring, and onboarding; coordinated budgets and resources across teams.

- Mentored junior engineers and facilitated cross functional communication to ensure successful delivery.

Conference Presentations
Invited speaker at industry conferences (ASGCT, OTS) and academic meetings (ISMB, Cold Spring Harbor), presenting ML methods and drug discovery applications to diverse audiences.