Max Highsmith

636-346-9886 | maxrhighsmith@gmail.com | LinkedIn | GitHub | max-highsmith.com

SUMMARY

Machine learning scientist and engineer with a Ph.D. in ML, 10+ years of software experience, and deep expertise spanning genomics, generative models, operations research, and LLM evaluation. Published researcher with first-author papers in deep learning for computational biology and independent work on AI safety and LLM generalization.

PROFESSIONAL EXPERIENCE

Synthesize Bio — AI Engineer **2025 – Present**

- Develop novel machine learning architectures for biological sequence generation, contributing across the full research lifecycle from data preparation through model training, evaluation, and reporting.
- Design benchmarks and analysis pipelines that drive model selection and iteration decisions for state-of-the-art deep-learning models in biological research.

Benson Hill — Genomic Data Scientist **2021 – 2025**

Ag-biotech company applying machine learning to crop genomics, breeding optimization, and food-trait prediction.

Target Discovery & MLOps

- Productionized neural networks within an MLOps framework on AWS (SageMaker, Step Functions, Lambda, Glue, Redshift) at scale.
- Built bioinformatics pipelines using boto3 and awswrangler to populate the data lake with Hi-C, structural variant, orthology, GWAS, TWAS, and gene expression datasets, formatted for queryable Athena tables.
- Trained weak-labeling trait discovery models that incorporated diverse genomic features to identify novel genes with predicted impact on key phenotypes (protein, oil, yield) and food-processing traits.
- Analyzed orthology datasets to extract synteny maps, leveraging genomic information across crop species for cross-species trait discovery.
- Supported a proprietary literature-mining database tracking co-occurrence of ontology terms and prospective gene targets across scientific articles, used as a feature in trait discovery models.
- Trained NIR spectroscopy models predicting wet-lab food traits from near-infrared spectral features of soybeans; high-performing NIR predictions were then used as ground-truth labels in downstream EBV genomic models, creating a two-stage prediction pipeline from spectra to genotype-based trait forecasting.
- Contributed to multiple production, cloud-based ML pipelines driving high-impact business decisions across gene-target discovery, NIR phenotype prediction, genomic selection, and breeding advancement; built SageMaker inference endpoints.

Analytics & Data Engineering

- Performed analytics on in-house wetlab food data to support product development decisions.
- Conducted haplotype analysis on crop genomic data to inform breeding strategies.
- Hydrated the data lake with curated datasets to support the Target Discovery team's machine learning pipelines.
- Built OrthoHunter, a data visualization application used by the Biological Science team to inform expert selections for gene target discovery.

Operations & Optimization

- Built and productionized estimated breeder value (EBV) models in SageMaker — genotype-to-phenotype ML models predicting fat, protein, and yield profiles from genetic data; automated pipeline predicted on all new genetic material, writing results to S3/Athena for downstream querying by breeding teams.
- Developed a multi-stage Monte Carlo genetic simulation modeling soybean breeding pipelines — simulating selfing and crossing stages from starter genotypes with known trait markers to predict final nursery genotype distributions, integrated with EBV models to forecast nursery phenotype specs before planting.
- Designed a mixed-integer linear programming model (PuLP) to schedule and spatially place all breeding nurseries across a 47,000 sq-ft Crop Accelerator greenhouse, incorporating constraints for chamber capacity, pipeline stage sizes, specialized labor availability for crossing operations, and work-schedule restrictions.

Conduce — Machine Learning Engineer **2019**

- Developed warehouse-optimization software at a 9-person startup, tracking unit movements over time intervals and reconstructing visual analytics of movement patterns.

- Built full-stack features end-to-end — Go backends and JavaScript UI — delivering performant analytics calculations including hotspot identification of warehouse movement.
- Worked with Cassandra for high-throughput storage and retrieval of time-series movement data.

MidwayUSA — Software Engineer **2018**

- Built ML-based credit-card fraud-detection improvements; estimated $250K/year savings to the company.
- Developed web applications in C# to support call center operations.

FlightSafety International — Software Engineer **2015 – 2016**

- Developed CUDA graphical shaders, internal developer tools, and simulation scripts for flight-training systems — direct GPU programming exposure at the shader level.

PROJECTS

PANOPTICON **2025 – 2026**

Creator / Sole Developer | panopticon.network | GitHub

- Built end-to-end (front-end, back-end, evaluation harness, multi-provider model integration) an open-source benchmarking framework for frontier LLMs across Anthropic, OpenAI, Google, and xAI.
- Supports queryable data layers across text, video, audio, and geospatial modalities with turn-based, real-time, and agentic tool-use execution modes.
- Renders scenario simulations on a real-time 3D globe, exposing model behavior across modalities.

GAMBIT **2026**

Sole Author — Independent Research | Paper | GitHub

- Designed multi-trial experimental protocols (N=50/condition) to study frontier-LLM behavior across 10 models and 7 providers in chess variant contexts.

Lextra Wallet

Creator | GitHub

- Developed an MCP-based wallet integrating Cardano functionality into AI desktop clients (ChatGPT, Claude), supporting account creation, transactions, and staking through a Python/PyCardano and Blockfrost API stack.
- Demonstrated seamless AI-to-Cardano transaction flow, enhancing onboarding for AI-first users.

Adaquarium

Creator | adaquarium.org | GitHub

- Implemented a 3D aquatic visualization system mapping Cardano blockchain network activity in near-real time, providing both educational and developer-oriented perspectives on transaction flow.

EDUCATION

Ph.D., Machine Learning — University of Missouri **2018 – 2021**

Dissertation: *Structural modeling of the 3D genome using machine learning* | Advisor: Jianlin Cheng
NIH T32 Fellowship (5T32GM008396-29)

Teaching: Pre-Algebra, Algebra, Finite Math, Databases, Intro to Computer Science

M.S., Applied Mathematics — University of Missouri **2016 – 2018**

Thesis: *Modeling sales pipelines for large-scale business transactions* | Advisor: Sam Walsh

- Modeled sales pipelines for large-scale business transactions using computational methods for differential equations and probability theory.
- Worked in the Distributed and Intelligent Computing Lab on aerial image recognition using deep learning for bird counting in aerial imagery (Advisor: Yi Shang).

B.S., Mathematics — Truman State University **2011 – 2015**

SKILLS

Languages:	Python (expert), Go, JavaScript, C#, CUDA
Frameworks:	PyTorch (expert), PyTorch Lightning, Weights & Biases
Cloud / Infra:	AWS (SageMaker, Step Functions, Lambda, Glue, Redshift, S3, Athena), GCP
Tools:	Docker, Git, SQL, boto3, awswrangler, PuLP
Databases:	Cassandra, Redshift, Athena
Domains:	Machine Learning, Deep Learning, Genomics / Bioinformatics, Computer Vision, Generative Models, Operations Research / Optimization, MLOps

CERTIFICATIONS

- AWS Certified Solutions Architect Associate (SAA-C03) 2024 – 2025
- AWS Certified Data Engineer Associate (DEA-C01) 2024 – 2025

PUBLICATIONS

1. Highsmith, M. (2026). “Responsibility Laundering: Agents Circumvent Safety Rails by Delegating Violations to Other Agents.” *Working paper.*

2. Highsmith, M. (2026). “GAMBIT: Generalization Assessment via Modified Boards and Irregular Topologies for LLMs in Chess.” *Working paper.*

3. Koytiger, G., Walsh, A. M., Marar, V., Johnson, K. A., Highsmith, M., et al. (2025). “Generative genomics accurately predicts future experimental results.” *bioRxiv.* doi: 10.1101/2025.09.08.674753

4. Li, Y., Boadu, F., Highsmith, M. R., Hagen, D. E., Cheng, J., & Rivera, R. M. (2022). “Allele-specific aberration of imprinted domain chromosome architecture associates with large offspring syndrome.” *iScience*, 25(5), 104269.

5. Highsmith, M., & Cheng, J. (2021). “VEHiCLE: A Variationally Encoded Hi-C Loss Enhancement Algorithm for Improving and Generating Hi-C Data.” *Scientific Reports*, 11(1), 8880.

6. Highsmith, M., & Cheng, J. (2021). “Four-Dimensional Chromosome Structure Prediction.” *International Journal of Molecular Sciences*, 22(18), 9785.

7. Highsmith, M., & Cheng, J. (2021). “An Introduction to Computational Approaches for 3D Genomic Modeling.” *ACM-BCB.*

8. Highsmith, M. (2020). “Iterative Approximation of Markov Process Parameters in a Model of Large Scale Business Purchases.” *University of Missouri* (M.S. Thesis).

9. Oluwadare, O., Highsmith, M., Turner, D., Lieberman Aiden, E., & Cheng, J. (2020). “GSDB: A Database of 3D Chromosome and Genome Structures Reconstructed from Hi-C Data.” *BMC Molecular and Cell Biology*, 21(1), 62.

10. Al-Azzawi, A., Ouadou, A., Highsmith, M., Duan, Y., Tanner, J. J., & Cheng, J. (2020). “DeepCryoPicker: Fully Automated Deep Neural Network for Single Protein Particle Picking in Cryo-EM.” *BMC Bioinformatics*, 21(1), 509.

11. Oluwadare, O., Highsmith, M., & Cheng, J. (2019). “An Overview of Methods for Reconstructing 3D Chromosome and Genome Structures from Hi-C Data.” *Biological Procedures Online*, 21, 7.

12. Lieu, Y., Sun, P., Highsmith, M., et al. (2018). “Performance Comparison of Deep Learning Techniques for Recognizing Birds in Aerial Images.” *IEEE DSC.*