



AI Engineer, Biological Foundation Models

Location: San Carlos, CA

Company: Heritable Agriculture

U.S. work eligibility: Required

Heritable Agriculture is seeking a trailblazing AI Engineer to push the boundaries of regulatory genomics with foundation models. We're looking for a creative builder who can develop sequence-based models that learn the cis-regulatory code and propose precise DNA edits to increase or decrease expression of a target gene. You'll work with a growing portfolio of plant multi-omic datasets to train, interpret, and deploy models that turn sequences into actionable, testable hypotheses for crop improvement.

What You'll Do

You will be at the forefront of our efforts to apply state-of-the-art AI to solve fundamental challenges in agriculture and biology.

- **Build Foundation Models:** Design, pretrain, and fine-tune long-context DNA sequence models (conv-attention/transformer/Hyena-style) that predict gene expression/CRE activity and variant effects from promoters, enhancers, UTRs, and splice regions
- **Propose Edits, Not Just Scores:** Develop edit-design methods (in-silico mutagenesis, gradient-guided minimal edits, generative proposals) that up-/down-regulate specified genes under defined tissues/conditions, with built-in motif/grammar, conservation, and off-target guardrails.
- **CRISPR Edit Outcome Modeling:** Build guide design and probabilistic edit outcome models (e.g., for Cas9/Cpf1; support single and dual-guide strategies), plus edit efficiency prediction using NGS-derived evidence.
- **Interpretability to Insight:** Implement IG/DeepLIFT/SHAP, motif ablations, and base-level attribution maps; generate experiment-ready rationales and ranked edit shortlists.
- **Close the Loop with Wet Lab:** Partner with biologists to plan MPRA/CRISPRi/a/prime-editing assays; run active-learning/Bayesian loops that select the next edits to test and retrain models on assay results.
- **Validate & Benchmark:** Establish a model leaderboard across genes/tissues; benchmark against natural variation (eQTL) and promoter-bashing datasets; report Δ -expression correlation, direction accuracy, and hit-rate.
- **Build & Ship Software:** Deliver robust, documented Python/PyTorch/JAX libraries, containers, and reproducible training/inference pipelines (Docker, Nextflow/Airflow; GCP/AWS).



-
- **Scale Across Platforms:** Architect efficient MLOps for large-scale pretraining and evaluation; manage datasets spanning ATAC/MOA-seq, ChIP, RNA-seq (bulk/single-cell), eQTL/TWAS, and Ribo-seq.
 - **Collaborate & Discover:** Work closely with software engineers, breeders, and biotechnologists to frame biological questions as solvable modeling problems and drive discovery.
-

Who You Are

You are a builder, an innovator, and a scientist at heart. You don't just use tools; you create them.

Required Qualifications:

- M.S. or Ph.D. (or equivalent experience) in Computer Science, Machine Learning, Computational Biology, or related quantitative field.
- Demonstrated expertise training sequence models for genomics (promoter/enhancer/UTR/splice) and variant-effect prediction; comfort with long genomic contexts (≥ 64 kb).
- Proven experience turning model scores into edit proposals and evaluating impact with attribution/ablation or in-silico mutagenesis.
- Proficiency with PyTorch or JAX, and modern software engineering (version control, testing, documentation, CI, containers).
- Experience building scalable pipelines on GCP/AWS/Azure.
- Team-oriented, mission-driven, and impact-focused.
- Ability to relocate to the San Francisco Bay Area and work in the office at least 4 days per week.

Preferred Skills:

- Hands-on work with genomic foundation models (e.g., Evo 2-class DNA LMs) and/or regulatory predictors such as Enformer/Basenji/BPNet/DeepSTARR; familiarity with long-context backbones (e.g., Hyena-like architectures) and JAX/TPU or PyTorch/GPUs at scale.
- Experience integrating motif/grammar constraints, conservation, and off-target risk into edit design; familiarity with guide selection and edit outcome modeling.
- Experience with MPRA/CRISPRi/a/prime-editing assay design/analysis and active-learning (e.g., Bayesian optimization) for experiment selection.
- Background with plant multi-omics and molecular biology genomics a plus.
- Experience leveraging TWAS/eQTL/GWAS to inform locus priors and evaluation cohorts.
- A public portfolio (e.g., GitHub) showing released libraries/packages or significant contributions.



-
- Excellent communication skills for cross-disciplinary teams and external audiences.

To Apply:

- Send an email with your application materials to ai-engineer@heritable.ag
- Review of applications will begin on October 31, 2025