DNAnexus®

AI/ML Accelerator by DNAnexus: Streamline, Secure, and Scale Your AI and Machine Learning Workflows

Accelerating Insights from Complex, Multimodal Data

Life sciences and healthcare organizations face a unique set of challenges when incorporating artificial intelligence (AI) and machine learning (ML) into their research and clinical workflows. Complex, multimodal datasets—including omics, clinical, and real-world data—often lead to bottlenecks that slow down decision-making and innovation. Organizations struggle with lengthy data preparation times, complicated exploratory data analysis, and difficulties ensuring compliance and scalability when deploying AI/ML models.

DNAnexus AI/ML Accelerator is specifically designed to resolve these challenges, enabling faster and secure insights from multimodal biomedical data within an integrated platform.
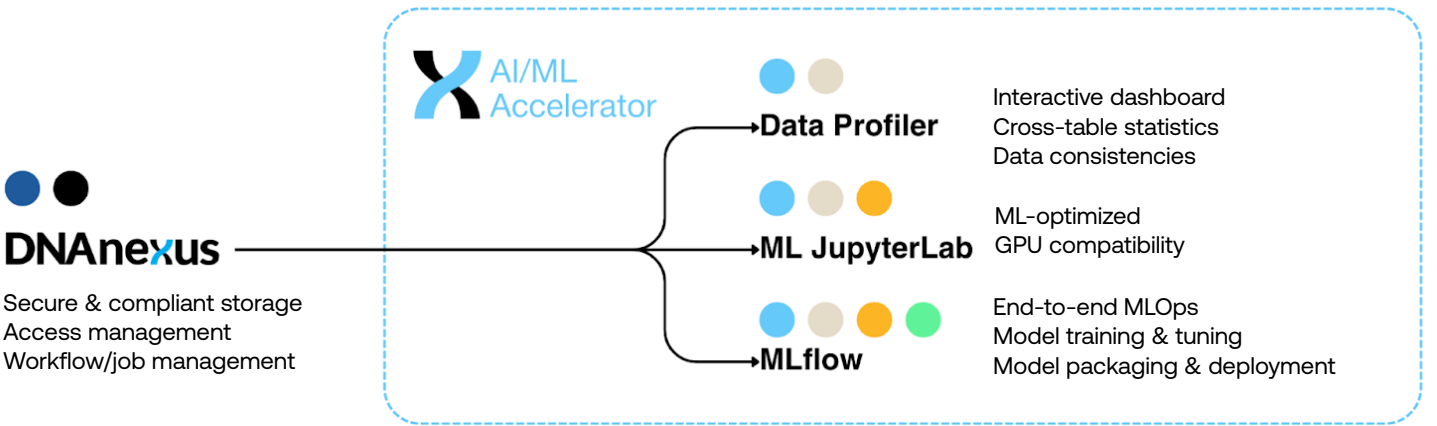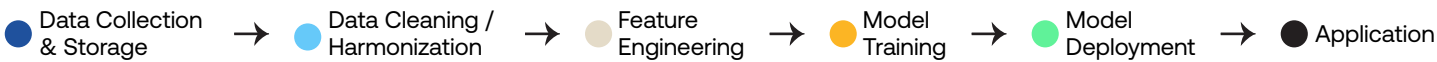
Powerful Tools for *Every Step* of Your Workflow

The AI/ML Accelerator is a robust suite of AI/ML tools tailored for healthcare and life sciences organizations. It supports data scientists, clinical informaticians, and researchers by significantly simplifying the development, deployment, and scaling of AI and machine learning workflows.

KEY CHALLENGES ADDRESSED

☑ **Data Complexity and Readiness:** Many organizations grapple with inconsistent, incomplete, or poorly structured data, which hampers effective analysis and model accuracy.

☑ **Cumbersome Exploratory Data Analysis (EDA):** Traditional EDA approaches require extensive coding expertise and substantial manual effort, slowing down crucial stages of data preparation and insight generation.

☑ **Security and Compliance Risks:** Life sciences companies face rigorous regulatory demands (e.g., HIPAA, GDPR), making it critical to securely manage sensitive data without excessive overhead.

☑ **Scaling AI/ML Projects:** Many organizations struggle to efficiently scale data processing and machine learning workloads, often facing bottlenecks due to resource constraints and infrastructure complexity.

Standard ML Development Workflow

Data Collection & Storage → Data Cleaning / Harmonization → Feature Engineering → Model Training → Model Deployment → Application

AI/ML Accelerator

Data Profiler
Interactive dashboard
Cross-table statistics
Data consistencies

ML JupyterLab
ML-optimized
GPU compatibility

MLflow
End-to-end MLOps
Model training & tuning
Model packaging & deployment

DNAnexus
Secure & compliant storage
Access management
Workflow/job management

Introducing DNAnexus AI/ML Accelerator

Data Profiler

User-friendly tool for comprehensive visualizations and insights

- Easily perform EDA for AI/ML with a "point and click" interface.
- Quickly explore data quality, data coverage, and descriptive statistics to identify potential data issues.
- Interactively examine data and create custom visualizations using an intuitive UI.
- Generate consistent Profile Reports from CSV, TSV, TXT, Parquet files, or a DNAnexus Apollo Dataset.

ML JupyterLab

Efficient ML, distributed and Ready—All Within JupyterLab

- Instantly access popular ML libraries and ready-to-use environments
- Effortlessly manage demanding tasks with scalable solutions powered by Ray, a distributed computing framework.
- Boost efficiency by processing large datasets in DNAnexus projects using parallelized operations on a Ray cluster.

MLflow

Instantly simplify, manage, and secure your ML workflow

- Oversee the entire ML lifecycle—from experimentation to deployment—within the DNAnexus platform.
- Securely collaborate on data and ML artifacts across projects, minimizing exposure risks.
- Eliminate tedious manual tasks with dataset, parameter, and metric logging automation for TensorFlow, PyTorch, and Scikit-learn.
- Manage model versions effortlessly with the model registry for saving, loading, and deploying models on DNAnexus.

Key Benefits for Your Organization:

Ease of Use and Accelerated Time to Insight:

- Intuitive, point-and-click interfaces like Data Profiler allow users to perform EDA effortlessly, rapidly assessing data quality, completeness, and structure through visualizations and customizable dashboards.
- ML JupyterLab provides ready-to-use, preconfigured machine learning environments with popular frameworks (TensorFlow, PyTorch, PyCaret), removing tedious setup steps and accelerating project initiation.

Dedicated Expert Support:

- Access DNAnexus Care and Professional Services tailored specifically for healthcare and life sciences use cases.

Comprehensive Data Profiling Capabilities:

- Quickly assess and visualize complex datasets, including clinical, omics, and real-world data.
- Interactive tools like PygWalker support customized, drag-and-drop data exploration, enabling non-technical users to easily generate insightful visualizations.

Seamless Scalability and Collaboration:

- Scale effortlessly using distributed computing via Ray Clusters, ensuring efficient processing of large datasets.
- Collaborative project spaces support secure, controlled sharing of data, models, and insights across diverse roles within your organization, from data engineers to clinical informaticians.

Robust Security and Built-in Compliance:

- Secure, compliant platform adhering to global regulatory standards such as HIPAA and GDPR.
- Centralized ML workflows on the DNAnexus platform eliminate data transfer risks, ensuring sensitive data remains within a secure, audited environment throughout the ML lifecycle.

Accelerate your AI and ML journey today—

confidently, compliantly, and collaboratively. Contact DNAnexus to learn more or schedule a demo.

DNAnexus®