

Data Versioning

Track and manage versions of datasets used for ML training and evaluation

Large Medical Images

CT/MRI scans are gigabytes per patient

HIPAA Compliance

Encrypted storage with audit logs

Reproducibility

Exact dataset for each model version

Tools & Platforms

DVC (Data Version Control)

Git-LFS

Pachyderm

Delta Lake