

AI Model Training Dataset Requirements & Governance Policy

Effective Date: January 15, 2024

Document Version: 2.0

Policy Owner: Summit Digital Solutions, Inc.

Last Reviewed: January 9, 2024

1. Purpose and Scope

1. This policy establishes mandatory requirements and governance procedures for the acquisition, validation, usage, and maintenance of datasets used in training artificial intelligence and machine learning models within Summit Digital Solutions, Inc.'s ("Company") Peak Performance Platform and related offerings.
2. This policy applies to all employees, contractors, and authorized third parties involved in AI/ML development, data science, and related activities utilizing training datasets.

2. Definitions

1. "Training Dataset" means any collection of data used to train, validate, or test AI/ML models.
2. "Sensitive Data" means any personal information, proprietary information, or regulated data as defined by applicable privacy laws and regulations.
3. "Model" means any artificial intelligence or machine learning algorithm, neural network, or computational system designed to learn from Training Datasets.
4. "Data Governance Committee" means the cross-functional oversight body responsible for enforcing this policy.

3. Dataset Requirements

1. Data Quality Standards

- All Training Datasets must meet minimum quality thresholds:
- Data completeness rate 95%
- Label accuracy rate 98%
- Documented data lineage

- Standardized format and structure
- Verified data integrity

2. Legal Compliance

- Training Datasets must:
- Be legally obtained and properly licensed
- Comply with all applicable data protection regulations
- Have documented rights for AI/ML training purposes
- Exclude any unauthorized third-party intellectual property
- Maintain audit trails of data source agreements

3. Bias Prevention

- Mandatory controls include:
- Demographic representation analysis
- Bias detection testing
- Regular fairness assessments
- Documented mitigation strategies
- Quarterly bias review reports

4. Governance Framework

1. Dataset Approval Process

Initial assessment by Data Science team

Legal review of data rights and compliance

Privacy impact analysis

Security risk evaluation

Final approval by Data Governance Committee

2. Documentation Requirements

- Each Training Dataset must maintain:
- Source documentation
- Usage rights verification
- Data dictionary

- Quality metrics
- Processing methodology
- Version control records

3. Access Controls

- Implementation of:
- Role-based access management
- Data encryption standards
- Access logging and monitoring
- Regular permission audits
- Secure transfer protocols

5. Operational Procedures

1. Dataset Validation

- Mandatory validation steps:

Statistical analysis

Outlier detection

Distribution verification

Cross-validation testing

Performance benchmarking

2. Maintenance Requirements

- Quarterly dataset reviews
- Annual comprehensive audits
- Regular updates to maintain relevance
- Version control management
- Retirement procedures for obsolete datasets

3. Emergency Procedures

- Immediate notification of data incidents
- Dataset quarantine protocols
- Investigation procedures

- Remediation requirements
- Stakeholder communication plan

6. Compliance and Monitoring

1. Audit Requirements

- Annual compliance audits
- Quarterly performance reviews
- Monthly monitoring reports
- Random spot checks
- External validation (as needed)

2. Reporting Requirements

- Regular compliance reports to:
- Data Governance Committee
- Executive Leadership
- Board of Directors (as requested)
- Regulatory bodies (as required)

7. Policy Enforcement

1. Non-compliance may result in:

- Immediate dataset suspension
- Corrective action requirements
- Project delays or cancellation
- Disciplinary action
- Legal remedies as appropriate

2. Exception Process

- Written request to Data Governance Committee
- Risk assessment requirement
- Time-limited approvals only
- Regular review of exceptions
- Documentation of justification

8. Policy Review and Updates

1. This policy shall be reviewed annually by the Data Governance Committee.

2. Updates require approval from:

- Chief Technology Officer
- Chief Innovation Officer
- Legal Department
- Data Governance Committee

9. Signature and Acknowledgment

By authority of:

Dr. Alexandra Reeves

Chief Executive Officer

Summit Digital Solutions, Inc.

Michael Chang

Chief Technology Officer

Summit Digital Solutions, Inc.

Date: _

Document Control Number: AI-POL-2024-001