RESOURCE MANAGEMENT

USER REQUIREMENT SPECIFICATION

TECHNICAL REQUIREMENTS FOR ARTIFICAL INTELLIGENCE MODELS FOR DATA ANALYTICS COLLABORATION

Author:	Engineer Train Design Technologies	Thenjiwe Mtsheku	03/02/2026
Approved:	Principal Engineer Train Design Technologies	Wynand Botha	
Authorised:	Executive Manager Technology Management	Justice Ngwenyama	

February 2026

Circulation Restricted To: Transnet, Relevant third parties

REVISION HISTORY

Rev 00	February 2026	Thenjiwe Mtsheku

TABLE OF CONTENTS

REVISION HISTORY 2
TABLE OF CONTENTS 3
GLOSSARY 4
1 INTRODUCTION 5
2 FUNCTIONAL AND TECHNICAL REQUIREMENTS 7

GLOSSARY

Term	Definition
AI	Artificial Intelligence
ADD	Auto Drop Device
API	Application Programming Interface
BI	Business Intelligence
BPO	Business Process Optimisation
CSV	Comma-Separated Values
ECPB	Electronically Controlled Pneumatic Brakes
GPU	Graphics Processing Unit
HPC	High-Performance Computing
HVAC	Heating, Ventilation, and Air Conditioning
KPI	Key Performance Indicator
LDS	Locomotive Data System
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MTBF	Mean Time Between Failures
MTTF	Mean Time To Failure
MTTR	Mean Time To Repair
MQTT	Message Queuing Telemetry Transport
OEE	Overall Equipment Effectiveness
OPC UA	Open Platform Communications Unified Architecture
POPIA	Protection of Personal Information Act
REST	Representational State Transfer
RM	Resource Management
RMSE	Root Mean Square Deviation
RUL	Remaining Useful Life
TAS & GS	Theoretical Analysis & Governance and Safety
TE	Transnet Engineering
TFR	Transnet Freight Rail
THG	Train Handling Guide
TM	Technology Management
VCB	Vacuum Circuit Breaker
VIS	Vehicle Identification System

INTRODUCTION

Transnet Freight Rail (TFR), through the Technology Management function in the Resource Management Division, seeks to collaborate with a Gauteng-based Academic Institution or a research and technology organization with advanced Artificial Intelligence (AI) and Machine Learning (ML) expertise, to research, design, and prototype Artificial Intelligence (AI) and Machine Learning (ML) analytical models that enhance the value, utilisation, and interpretation of TFR's operational and engineering datasets that will be used to provide a single, reliable source of information for the organization.

TFR generates large and diverse volumes of data across its locomotive, rolling stock, and infrastructure value chains. These include:

- on-board operational and event-recording systems
- telemetry, GPS and location-based feeds
- asset identification and tracking data
- condition-monitoring and diagnostic outputs
- maintenance, inspection and operational activity records
- specialised measurement, test-run and instrumentation datasets
- incident, exception and event-based operational logs

Extracting actionable insights from these complex, high-volume datasets using conventional tools has become increasingly challenging. To strengthen data-driven reliability, maintenance optimisation, and operational decision-making, TFR intends to establish a research collaboration with universities that possess advanced AI/ML capabilities.

The collaboration will focus on applied research and the development of analytical models, targeting priority engineering and operational domains such as traction performance, braking systems, bogie and wheel behaviour, coupler dynamics, energy management and incident analytics. Activities will include data preparation, modelling, feature engineering, analysis, visualisation, and knowledge transfer.

This initiative aims to:

- enable predictive, prescriptive, and diagnostic analytics to improve asset reliability;
- provide data-driven insights that support maintenance planning, operational optimisation and risk mitigation;
- promote innovation, skills development and student involvement; and
- establish a sustainable industry–academia partnership that supports TFR's operational excellence and contributes to South Africa's digital research ecosystem.

Overall, the collaboration reflects TFR's commitment to leveraging modern AI methodologies to advance next-generation rail analytics and support long-term technology transformation.

1.1 Scope of Collaboration

This specification defines the scope of a collaborative research partnership between Transnet Freight Rail (TFR) and a South African university with recognised Artificial Intelligence (AI) and Machine Learning (ML) research capability. The collaboration focuses on research, development, and prototyping analytical models that support TFR's operational, engineering, and asset-management objectives.

The scope covers research activities only; production deployment, enterprise integration, and software implementation remain out of scope unless explicitly contracted by TFR.

1.1.1 AI/ML Research and Prototype Development

Design, experimentation, and prototyping of AI/ML analytical models using TFR's operational and engineering datasets. Model outputs may include predictive, prescriptive, diagnostic, optimisation, classification, or anomaly-detection algorithms aligned to TFR's operational context.

1.1.2 Data Preparation and Enrichment

Preparation of datasets required for modelling, including:

- data cleaning, synchronisation, and validation
- transformation, harmonisation, and feature engineering
- enrichment with contextual or environmental variables where appropriate and approved by TFR

Any field instrumentation, new sensors, or hardware trials must be pre-approved in writing by TFR and, if required, separately contracted.

1.1.3 Analytics, KPIs, and Visual Insight Development

Creation of research-level analytics, including:

- exploratory data analyses
- model-based diagnostic indicators
- prototype KPIs and visualisations
- conceptual dashboards or insight views

These artefacts are intended to support engineering and operational decision-making.

1.1.4 Operational and Maintenance Workflow Concepts

Development of conceptual workflow models to illustrate the potential operational application of AI/ML outputs (e.g., incident triage, maintenance prioritisation, anomaly alerts, or reporting automation).

1.1.5 Functional Domains of Application

The collaboration may span multiple domains, including but not limited to:

- traction and braking systems
- bogies, running gear, and wheel–rail interaction
- couplers and drawgear
- energy and auxiliary systems
- condition monitoring and diagnostic subsystems
- operational performance and fleet reliability
- maintenance planning and asset-health modelling

Specific research topics will be agreed with TFR for each work package.

1.1.6 Integration and Architecture Research Concepts

Where relevant, the university may propose integration-ready research artefacts, such as:

- Conceptual API designs
- open-format data-exchange schema
- reference architectures or data pipelines suitable for future implementation

1.1.7 Governance, Security, and Model Lifecycle Practices

The university must apply research-level model-governance principles, including:

- experiment tracking and reproducibility
- versioning of models, datasets, and code
- documentation of data lineage
- explainability and transparency of model behaviour

- secure storage and handling of sensitive TFR datasets
- compliance with POPIA and TFR ICT security requirements

Where applicable, the university's practices should align with recognised standards such as ISO/IEC 27001 for information security.

1.1.8 Collaboration Deliverables and Responsibilities

The university must:

- produce prototype models and research artefacts as defined with TFR
- provide reproducible methodologies and clear documentation
- participate in technical review sessions with TFR
- support knowledge transfer, including postgraduate student involvement
- deliver all outputs in an open, transparent, and implementation-ready format

2 FUNCTIONAL AND TECHNICAL REQUIREMENTS

The selected university must demonstrate the capability to undertake research-grade AI/ML modelling, advanced data analysis, and prototype development using diverse operational and engineering datasets from Transnet Freight Rail (TFR). All work shall be performed in a secure, POPIA-compliant research environment and must contribute directly to improving TFR's operational insight, reliability, maintenance planning, and engineering decision-making.

2.1 Functional Requirements

The university shall deliver research projects that achieve the following outcomes:

2.1.1 Core Analytical Capabilities

- **Ingest, process, and analyse TFR datasets** originating from on-board recorders, asset-tracking systems, telemetry feeds, diagnostic platforms, maintenance systems, test-runs, and incident logs.
- **Develop predictive, diagnostic, prescriptive, optimisation, classification, or anomaly-detection models** relevant to TFR's assets and operating environment.
- **Implement and evaluate automated anomaly-detection and early-warning methods** for latent faults or emerging failure patterns.
- **Provide prototype KPI and analytical visualisations**, including reliability indices, utilisation measures, and energy-performance views.
- **Develop integration-ready research artefacts**, such as API specifications, data schemas, or open-format data-exchange models suitable for future incorporation into TFR enterprise systems.
- **Produce insight-generation outputs** such as comparative analyses, fault clusters, trend evaluations, and investigative summaries for engineering teams.
- **Support AI-assisted reporting and simulation concepts** for maintenance, energy optimisation, and operational decision support.
- **Use modular, extensible modelling approaches** that support future expansion, new datasets, and iterative improvements.
- **Ensure secure dataset handling, controlled access, and compliance with POPIA and TFR data-governance policies.**

2.2 Data and Analytical Capability Requirements

2.2.1 Data Sources and Acquisition

The bidders should be capable of processing data from:

- on-board operational data recording systems
- asset identification and tracking datasets

- telemetry, GPS, and location-based data streams
- condition-monitoring, diagnostic, and telemeter systems
- maintenance and inspection records (SAP/Oracle/DB2 extracts)
- specialised measurement systems (QuantumX, accelerometers, strain/vibration loggers, rope transducers)
- production-train datasets (air-brake, ECP, pneumatic logs)
- incident, exception, and event-based operational logs

The bidder shall perform cleaning, validation, synchronisation, harmonisation, and feature engineering necessary to prepare data for modelling.

2.2.2 Analytical and Model Development Capability

Research outputs may include:

- supervised, unsupervised, and reinforcement learning models
- fault-pattern discovery and causal inference (where applicable)
- context-aware models (weather, gradient, load, traffic density)
- experiments using historical batch data or simulated real-time streams
- feedback-loop retraining and drift-assessment experiments

2.3 Functional Domains of Application

Models may be applied across multiple locomotive and rolling-stock domains, including:

- **Traction and Propulsion Systems**
  - analysis of traction-converter and inverter parameters
  - motor torque-temperature correlation, wheel-slip detection
  - energy-consumption modelling and regeneration efficiency
  - DC-link stability, surge-arrestor behaviour, cooling performance
- **Brakes Technology**
  - wear-rate estimation, brake-actuator diagnostics
  - pneumatic leak detection, pressure-decay deviations
  - ECPB log analysis, valve timing, communication errors
  - Remaining Useful Life (RUL) models
- **Bogies and Running Gear**
  - detection of bogie vibration, hunting, wheel polygonisation
  - fusion of accelerometer/axle-load/track-geometry data
  - reinforcement-learning-based overhaul optimisation
- **Couplers and Drawgear**
  - shock-logger force-signature analysis
  - lifecycle modelling and fatigue prediction
  - digital-twin concepts linked to SAP
- **Electrical, Power and Auxiliary Systems**
  - transformer, rectifier, inverter health monitoring
  - current/voltage imbalance, overheating, insulation tracking
  - compressor, motor, and auxiliary load monitoring
- **Pantograph, OHTE and ADD Events**
  - ADD/pantograph-strike detection and localisation

- line-voltage correlation, environmental context analytics
- Diesel Engine and Fuel Systems (where applicable)
- engine-health modelling, fuel-consumption optimisation
- degradation modelling for turbochargers, injectors, filters
- integration with SARS Diesel Rebate compliance datasets

- **HVAC, Cab and Crew Environment**
  - cab temperature/humidity performance
  - HVAC subsystem fault prediction
- **Energy Management and Regeneration**
  - regeneration efficiency analysis
  - braking-energy loss assessment
- **Protection, Safety and Control Systems**
  - VCB/circuit-breaker fault-log analytics
  - vigilance-overspeed-penalty event correlation
- **Simulation and In-Service Testing**
  - train-performance analysis and black-box event reconstruction
  - THG and driver-performance assessments
- **Maintenance Planning and Scheduling**
  - predictive maintenance KPIs (MTTF, MTBF, MTTR, OEE, availability)
  - spares forecasting
  - tool and calibration management concepts
- **Administrative and Process Optimisation**
  - AI-assisted routing, claims processing, documentation drafting
  - semantic search over historical records

2.4 Research Tooling, Infrastructure, and Integration Readiness

The bidder shall provide:

- GPU-enabled or HPC research computing environments
- secure, POPIA-aligned research data-storage environments
- model versioning, experiment tracking, reproducibility, and audit trails
- open-format integration artefacts (CSV, Parquet, JSON, REST, MQTT, OPC UA—where applicable)
- conceptual API designs or data schemas suitable for LDS/SAP/Telemetry/VIS
- automated data-quality scoring and inference-logging pipelines at research level
- low-/no-code prototypes where beneficial to engineering teams

Production system integration is not part of this scope unless expressly contracted.

2.5 Performance and Validation Requirements (Research-Level)

Each research project must define appropriate validation criteria, but the following baseline applies:

2.5.1 Data Quality & Processing Metrics

- ingestion and processing times
- completeness, error rates, and integrity checks
- multi-stream synchronisation accuracy and drift residuals
- scalability trials on HPC or equivalent environments

2.5.2 Model Accuracy & Reliability Metrics

- precision, recall, F1-score, AUROC (classification)
- MAE, RMSE, MAPE (regression)
- confusion-matrix analysis
- false-alarm analysis and drift monitoring
- benchmarking against baseline and peer models

2.5.3 Prototype Responsiveness & Visualisation

- insight refresh rates suitable for demonstration
- query turnaround time
- alert/event latency in research harnesses

2.5.4 Interoperability & Security

- conformance to data schemas and interface definitions
- test-integration success rates
- adherence to POPIA and TFR ICT security requirements

2.5.5 Operational Value Demonstration

- case studies showing potential improvements to reliability, availability, MTBF/MTTR, energy efficiency, or maintenance effectiveness
- engineering feedback from TFR review workshops

These are research metrics, not production SLAs.

2.6 Project Delivery, Governance, and Knowledge Transfer

The bidder shall demonstrate structured research-project governance, including:

- research plans, milestones, risk registers, and ethics/data-access approvals
- reproducible documentation (data schemas, feature lists, modelling assumptions)
- code notebooks, workflows, model cards, and reproducibility packs
- technical walkthroughs and knowledge-transfer sessions
- postgraduate student involvement where appropriate
- continuous improvement and research governance aligned with ISO/IEC 42001:2023
- periodic retraining or recalibration of models using updated datasets

2.7 Constraints and Compliance Requirements

The bidder shall operate within the following constraints:

- Data access: NDAs, POPIA compliance, and ethics clearance where necessary
- Security: all modelling and analysis must occur within secure, approved research environments
- Data availability: some datasets may be intermittent or time-bound; TFR will define access per project
- Publication: any academic output requires prior written TFR approval
- Intellectual Property: all foreground IP developed under this collaboration remains the property of TFR

This specification outlines the essential requirements for a research partnership that will enable Transnet Freight Rail to advance its AI and data-analytics capabilities. By providing a structured framework for secure, innovative,

and high-value model development, the collaboration aims to strengthen TFR's operational insight, support informed engineering decisions, and contribute to long-term digital transformation across the organisation