REINFORCEMENT LEARNING FOR SAFE & EFFICIENT FREIGHT TRAIN OPERATIONS

Statement of Work

Authors:

J. Brooks, J. Wakeman, M. Pietrzykowski, and M. Karunaratne

Digital Advanced Technologies

2025-10-23 *– Issue 1*

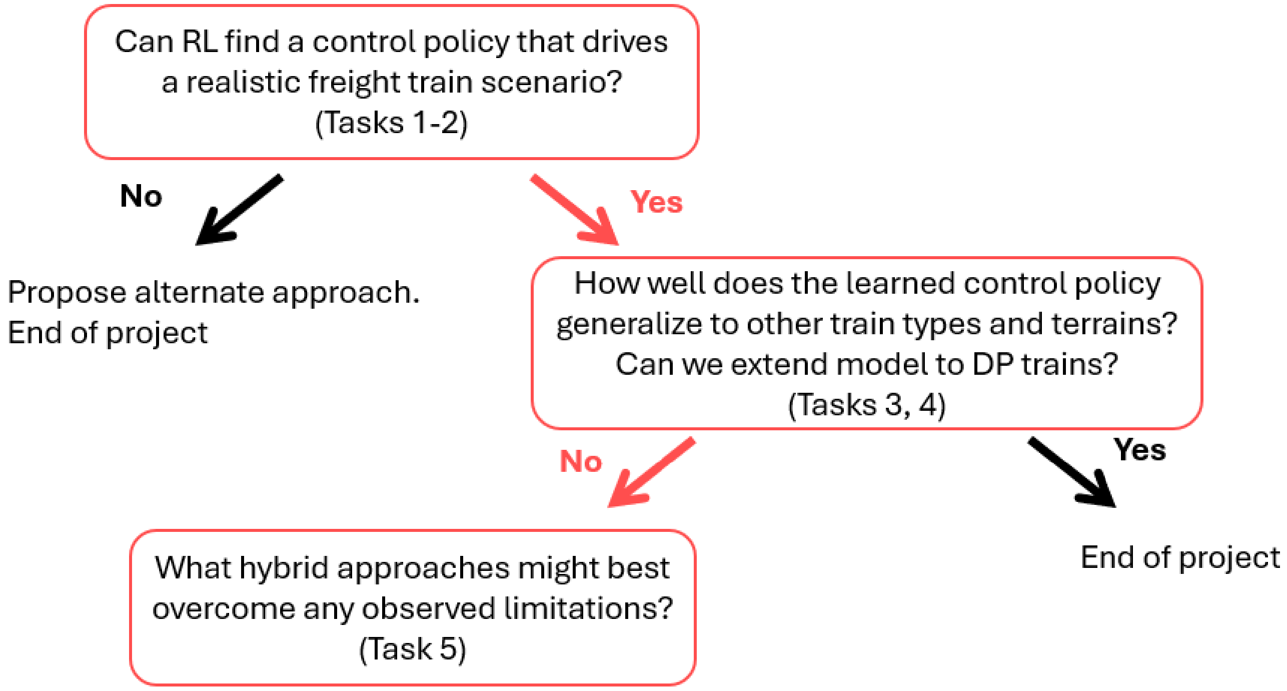
/// SOW SUMMARY

Wabtec has long been the industry leader in energy management systems with their Trip Optimizer™ product. With recent advances in machine learning and other hybrid approaches, Wabtec would like to explore the opportunities and limits of these methods in a scoped research project with FEV. This project seeks to answer a few key research questions which form the structure for the work:

1) Can state of the art reinforcement learning (RL) techniques find a control policy that reasonably drives a realistic conventional freight train scenario without airbrake?
2) How well does the learned control policy generalize to other train types and terrains?
3) What extensions are needed to handle distributed power trains?
4) What hybrid approaches might best overcome any observed limitations while providing safe and explainable system behavior?

The Wabtec team has done some initial exploration in developing a simulation environment and training a basic RL algorithm. Candidate train/track combinations for the various testing stages have also been selected. This initial research and development work would extend our understanding of what performance is possible with machine learning techniques, understand limitations and any other risks, and propose hybrid approaches that may be explored in a follow up project. The decision tree below outlines the flow of the work, key go/no-go decisions, and hypothesized answers to the various research questions.

/// RL for Safe & Efficient Train Operations Project Decision Tree



Can RL find a control policy that drives a realistic freight train scenario? (Tasks 1-2)

No

Propose alternate approach. End of project

Yes

How well does the learned control policy generalize to other train types and terrains? Can we extend model to DP trains? (Tasks 3, 4)

No

What hybrid approaches might best overcome any observed limitations? (Task 5)

Yes

End of project

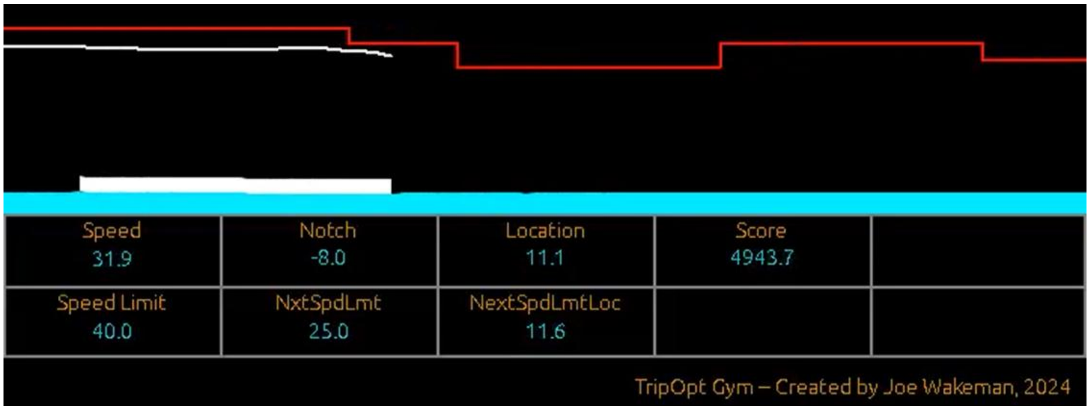
The scope of work is as follows:

Contents

Section 1: Simulation Framework

RL techniques require a robust simulation environment wherein algorithms can learn and explore possible control actions to find acceptable control policies. Wabtec has created an initial test system, called TO Gym, that trains a deep Q-network agent to control locomotive throttle and dynamic braking to minimize trip time while respecting speed limits and managing energy consumption. A sample output trace is shown in Figure 1 below. It currently has the following features:

- Realistic lumped-mass train physics simulation with trapezoidal integration
- Multi-locomotive consist model (3x ES44AC locomotives)
- Route-based scenarios with elevation profiles and speed limits
- Visual rendering with rolling map display and data grid
- DQN agent with experience replay and target network
- Configurable hyperparameters via YAML files
- Centralized device management (CUDA/CPU auto-detection)



Speed 31.9	Notch -8.0	Location 11.1	Score 4943.7	
Speed Limit 40.0	NxtSpdLmt 25.0	NextSpdLmtLoc 11.6		

TripOpt Gym – Created by Joe Wakeman, 2024

Figure 1. Sample output of current RL exploration system developed by Wabtec.

For this section of the work:

- Wabtec shall provide FEV:
  - Existing models (gitlab.corp.wabtec.com/Joseph.Wakeman/tripoptgym) – note these are distance-based.
  - Access to detailed physics models that generate time-based in-train coupler force estimates. (https://gitlab.corp.wabtec.com/Joseph.Wakeman/pysim)
  - Detailed track database (Fresno subdivision): includes grades, curves, speed limits for 2 representative regions
  - Detailed train makeup information: locomotive performance characteristics, car by car weights, lengths, coupler type
  - SMEs for weekly technical meetings and to support initial understanding of prototype system.
- FEV shall:
  - Extend the simulation environment to include train forces and other operational variations (grade errors, high wind)
  - Extend simulation to take detailed loco characteristics and train makeup (car by car) inputs.
  - Extend the simulation environment to include any other features needed for the remaining work.

Section 2 : Reinforcement Learning Algorithm Development – POC v1.0

This portion of the work would work to develop an RL-based autonomous driving engine that leverages deep reinforcement learning to dynamically navigate and optimize freight train operations. Unlike traditional rule-based and physics-modeling approaches, this new architecture either independent of or combined with existing physics based systems will adapt and learn from real-world operational data to unlock a scalable autonomous platform that can work across the rail network.

This approach would leverage a continuous training pipeline, utilizing field data to progressively enhance the AI's operational capabilities. A Q-Network is a neural network that learns to predict the expected future reward (Q-value) for each action in a given state, enabling reinforcement learning agents to make optimal decisions in complex environment. Various training techniques would be explored including experience replay and epsilon-greedy exploration with a decay schedule.

The Wabtec team has implemented a basic proof of concept which will be shared with the FEV team. The observation space is currently composed of a 6-dimensional continuous vector:

- Train velocity (mph)
- Train acceleration (mph/minute)
- Current position in route (miles)
- Current speed limit (mph)
- Next speed limit (mph)
- Next speed limit location (miles).

The action space is simply notch change behavior (note: no airbrake actuation is expected in these first POCs; only motoring and dynamic braking at locomotives):

- 0: Hold current notch (no change)
- 1: Notch up (increase throttle or decrease brake)
- 2: Notch down (decrease throttle or increase brake)

Notch ranges from -8 (maximum dynamic brake) to +8 (maximum throttle). Actions are rate-limited to one notch change per 3 seconds.

And finally, the reward function has several components that include:

1. **Progress reward**: +100 points per mile traveled
2. **Speed compliance**:
    - Quadratic penalty for exceeding speed limits (-5 × error^1.5 for minor violations, up to -50 per mph for major violations)
    - Bonus for operating near speed limit (85-100% of limit)
    - Modest penalty for very slow speeds (<40% of limit) without acceleration

3. **Anticipation bonus**: Up to +10 points for slowing down early before speed limit reductions (within 2 miles)
4. **Terminal rewards**:
   - +500 for successfully reaching destination
   - -300 for stalling (speed drops below 5 mph)

For this section of the work:

- Wabtec shall provide FEV:
  - Existing models (github)
  - Representative operational data as needed (e.g., operational trip data – speed, notch, etc. – for up to 10 representative runs at one second rate)
  - Detailed track database: includes grades, curves, speed limits for 2 representative regions (Fresno and Pine Bluff subdivisions)
    - These tracks are benign enough that train airbrakes are not needed to traverse the route
  - Detailed train makeup information (2-5 in total): locomotive performance characteristics, car by car weights, lengths, coupler type (some EOCC units)
    - Conventional locomotive configuration (all at head end, common notch/brake command)
  - Trip Optimizer (TO) trip details and performance metrics for these representative cases (to match above information)
  - SMEs for weekly technical update meetings and to support initial understanding of prototype system.
- FEV shall:
  - Refine the RL modeling approach (model structure, observation space, reward/training method, etc.) to improve performance for conventional trains.
  - Generate test results for representative tracks and trains and compare performance metrics to existing TO system
    - Metrics to include at least: average velocity, maximum speed limit violation, in-train forces (min/max buff/draft forces), and fuel consumption

Section 3: Generalizability Testing

This portion of the work aims to understand the extent to which the initial RL agent has learned fundamental and transferable control policies. Additional track and train configurations will be provided. For this section of the work:

- Wabtec shall provide FEV:
    - 15-20 additional detailed track databases: includes grades, curves, speed limits
        - No airbrake use is expected to be necessary on the routes provided
    - 15-20 additional detailed train makeup information: locomotive performance characteristics, car by car weights, lengths, coupler type (conventional only; including EOCC)
    - Trip Optimizer (TO) trip details and performance metrics for representative cases (to match above information)
    - SMEs for weekly technical update meetings
- FEV shall:
    - Generate test results for the expanded set of representative tracks and conventional trains and compare performance metrics to existing TO system
        - Analyze RL agent performance and summarize deficiencies and proposed modifications.
        - Exploration of model parameter structure that predict/explain generalization results

Section 4: Distributed Power Extension – POC v2.0

Distributed power (DP) trains come in a variety of configurations which are designed to make large trains more fuel efficient and better manage in-train coupler forces by placing some of the locomotives in the middle or rear of the train. Key operational decisions with DP trains include whether the notch commands should be the same (synchronous) or different (asynchronous). There are various operating rules of thumb that operators use that we hope a RL agent can learn empirically.

The action space would need to be extended to provide a mechanism for increasing/decreasing the notch command of the remote units. Additional inputs are also likely to inform the agent about the location and size of each locomotive consist and the grade beneath the train extent.

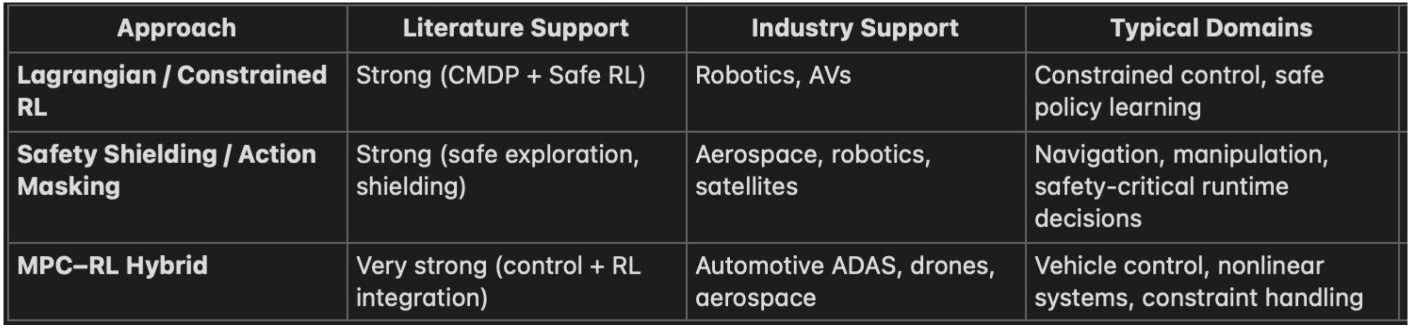
For this section of the work:

- Wabtec shall provide FEV:
  - 15-20 additional detailed train makeup information: locomotive performance characteristics, car by car weights, lengths, coupler type (including EOCC)
    - To include distributed power (DP) locomotive configuration (some locos in middle or end of train consist which can be commanded a power/brake level different from the locomotives in the front) – maximum of two locomotive consists
    - Again, no airbrake use is expected to be necessary for these trains on the routes provided
  - Trip Optimizer (TO) trip details and performance metrics for representative cases (to match above information)
  - SMEs for weekly technical update meetings
- FEV shall:
  - Extend design of RL agent to include DP train makeup (extend dimensionality of action space to include independent remote locomotive notch command).
    - Select subset of trains for training; test generalizability on remainder
    - Analyze RL agent performance and summarize deficiencies and proposed modifications.

Section 5: Hybrid Approach Exploration

We recognize that unconstrained machine learning approaches may not be ideal for critical operations like freight train operations. The Wabtec team has done some initial literature review of hybrid approaches shown in Table 1 below. The goal of this portion of the work is to make a recommendation to the Wabtec team about potential hybrid approaches that have promise in overcoming any limitations found in Sections 2-4.

Table 1. Representative hybrid approaches and their characteristics.



Approach	Literature Support	Industry Support	Typical Domains
Lagrangian / Constrained RL	Strong (CMDP + Safe RL)	Robotics, AVs	Constrained control, safe policy learning
Safety Shielding / Action Masking	Strong (safe exploration, shielding)	Aerospace, robotics, satellites	Navigation, manipulation, safety-critical runtime decisions
MPC–RL Hybrid	Very strong (control + RL integration)	Automotive ADAS, drones, aerospace	Vehicle control, nonlinear systems, constraint handling

- Wabtec shall provide FEV:
  - Known research literature
  - SMEs for discussion and debate
- FEV shall:
  - Provide a presentation summary of available hybrid methods and rationale for pros/cons relative to our problem context and the results of Sections 2-3 of the work.

Section 6: Deliverables and Timing

It is expected that the first two tasks be completed in parallel. Each Go/No-Go milestone review will be planned with review materials sent to the Wabtec technical team at least one week in advance. Total period of performance not to exceed 10 months. FEV shall deliver the following over the course of this work:

- **Machine Learning Simulation Framework (Section 1)**
  - Software implementing complete training environment.

- **POC v1.0 Details and Results (Section 2)**
  - Software implementing RL agent
  - Report describing design details and theoretical background of method as needed
  - *Go/No-Go Milestone:* Results of training and testing (presentation and raw data)

- **POC v1.0 Generalizability Results (Section 3)**
  - Software implementing extended test set for RL agent
  - *Go/No-Go Milestone:* Results of training and testing (presentation and raw data)

- **POC v2.0 Details and Results (Section 4)**
  - Software implementing extended RL agent for distributed power trains
  - Report describing design details and theoretical background of method(s) as needed
  - *Go/No-Go Milestone:* Results of training and testing (presentation and raw data)

- **Hybrid Approaches (Section 5)**
  - Report/presentation of state-of-the-art hybrid approaches and recommended approach for the freight rail case.