

EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) FÜR REINFORCEMENT LEARNING (RL) BASIERTE BETRIEBSSTRATEGIEOPTIMIERUNG

Aufgabenstellung

Für die **Optimierung der Betriebsstrategie von industriellen Versorgungssystemen** wird in der **ETA-Fabrik** das **Reinforcement Learning (RL)** eingesetzt. Forschungen haben bereits gezeigt, dass dadurch der **Energieverbrauch** und die **CO₂-Emissionen** der Anlagen deutlich gesenkt werden können. Die fehlende Interpretierbarkeit der optimierten Strategien solcher Black-Box KI-Modelle führt in der Praxis allerdings zu fehlender Akzeptanz und stellt damit eine Hürde für die Anwendung am realen System dar. Mit dieser Problematik beschäftigt sich das Forschungsfeld **Explainable Artificial Intelligence (XAI)**. Dabei sollen Entscheidungen von Black-Box KI-Modellen durch den Einsatz von **XAI-Methoden** erklärbar gemacht werden.

Das Ziel dieser Arbeit besteht darin, **geeignete XAI-Methoden zu identifizieren** und in ein bereits bestehendes **Python-Framework zur Betriebsoptimierung** mittels **Reinforcement Learning** zu **implementieren**. Anschließend soll in verschiedenen Versuchen der **Ansatz und dessen Interpretierbarkeit** evaluiert werden. Die Aufgabenstellung lässt sich folgendermaßen untergliedern:

- Literaturrecherche zu Interpretable Machine Learning, Explainable Artificial Intelligence (XAI), Extended Explainable Artificial Intelligence (XXAI), Explainable Reinforcement Learning (XRL)
- Einarbeitung in das vorhandene Python-Framework
- Auswahl und Implementierung geeigneter Methoden / Ansätze
- Anwendung der implementierten Methoden zur Interpretierbarkeit der optimierten Betriebsstrategien
- Evaluierung und Dokumentation der Ergebnisse

Voraussetzungen

- Programmierkenntnisse in Python

Beginn

Ab **sofort**

Kontakt

Arthur Stobert
a.stobert@ptw.tu-darmstadt.de

