Integrating Cyber-Physical Passports and Generative AI for Explainable RUL Estimation in Battery Remanufacturing

Yanchao Tan
Purdue University
West Lafayette, USA

Xingyu Li
Purdue University
West Lafayette, USA

Ragu Athinarayanan*
Purdue University
West Lafayette, USA

Abstract

Remanufacturing is a critical pathway for advancing circular economy (CE) goals by extending the product life of high-value components and reducing material waste. A prominent example of CE adoption is the rapid growth of electric vehicles (EVs), which has simultaneously increased the demand for remanufacturing end-of-life lithium-ion batteries. Accurate prediction of the remaining useful life (RUL) of returned batteries is essential at the inspection stage to guide reuse, replacement, or disposal decisions. However, considering that batteries may experience multiple use cycles, obtaining and interpreting the required multicycle data, and helping human operators to gain the actionable remanufacturing insights, remains a challenge.

Cyber-Physical Passports (CPPs) provide a way to retain product lifecycle information such as physical attributes, manufacturing data, and components details, which are important for product RUL estimation. However, the heterogeneous and often unstructured nature of this data limits its direct applicability. Additionally, conventional AI models for RUL prediction typically yield single-point estimates (e.g., cycles to failure), which capture degradation but often fail to represent uncertainty or provide actionable context for remanufacturing decisions. Generative AI (GenAI), particularly large language models (LLMs), offers a complementary capability by analyzing diverse data formats, interpreting results into natural language, and improving human–data interaction during the decision process.

This study focuses on investigating how Generative AI (GenAI) could expand the capabilities of traditional RUL estimation and remanufacturing decision-making. We will integrate GenAI with two Artificial Intelligence (AI) predictive models: a decision tree model, which produces interpretable rule-based outputs, and a Long Short-Term Memory (LSTM) model, which captures temporal degradation patterns and outputs time-to-event distributions. Using battery datasets containing features such as cycle count, discharge capacity, current, and voltage, we assess model performance with root mean square error (RMSE) and Mean Absolute Error (MAE). Explainability is addressed through SHapley Additive exPlanations (SHAP) values to identify the most influential factors driving predictions.

Building on these predictive models, we will further demonstrate how fine-tuned LLMs interpret the outputs into domain-specific natural language explanations, such as clarifying why certain features contribute to observed degradation trends. We will also conduct a case study to illustrate how these explanations support decision-making in representative degradation scenarios and strengthen human-data interaction. The findings will provide a comparative assessment of interpretable and survival-based models, while demonstrating how LLMs can extend model outputs to enhance transparency, explainability, and human-data interaction in RUL estimation for battery remanufacturing.

Keywords: Circular Economy; Remanufacturing; Cyber Physical Passports; Large Language Models; Decision Tree Model; LSTM

1. Introduction and Motivation

The rapid increase in EV adoption is a representative example using CE, which will be over 145 million on the road by 2030. But it also brings the challenges that an estimated 12 million lithium-ion batteries are expected to reach the end of their first life over the next decade [1, 2]. Remanufacturing has emerged as a potentially impactful CE strategy for high-value components by enabling environmental and economic benefits through material recovery and reuse, through second-life energy storage applications [3].

One of the key steps in battery remanufacturing is the inspection and evaluation of returned units. The accurate estimation of the RUL plays an important role in determining whether a battery should be reused, balanced, downgraded, or disassembled [4]. However, batteries may undergo heterogeneous usage conditions, which often result in incomplete, inconsistent, or missing multicycle data [5]. Additionally, interpreting voltage-current behaviors and

degradation signatures requires specialized electrochemical knowledge, while the industry to facing a widening skills gap, as experienced battery remanufacturing technicians remain in demand [6].

CPPs, which can provide a digital "birth certificate" and life record for products, enabling secure identification, tracking, and verification across the lifecycle. It can offer trusted essential metadata such as manufacturer information, product specifications, chemistry, historical diagnostic tests, and safety records, which are particularly valuable during RUL estimation and condition assessment [7]. However, the data might be heterogeneous and unstructured, making them difficult to interpret by both conventional AI and human operators [8]. Moreover, conventional RUL prediction models typically generate a single-point estimate of remaining cycles, often in a black-box manner, without explaining which degradation indicators drive the prediction or how the result should be interpreted in a remanufacturing context [9]. This lack of interpretability limits trust, reduces decision transparency, and constrains the ability of human operators to translate model output into concrete remanufacturing actions.

Recent developments in GenAI, particularly LLMs, introduce new opportunities to address these challenges. LLMs have demonstrated strong capabilities in interpreting multimodal data, maintaining contextual awareness, and generating task-specific recommendations [10]. Their use of natural language also enables more intuitive information exchange, thereby supporting human operators in making informed decisions [11]. However, most LLMs are trained primarily on web-scale text corpora and therefore lack the inherent ability to perform precise, domain-specific analysis or extract high-fidelity insights directly from CPP data. How to effectively integrate the interpretability and reasoning strengths of GenAI with the analytical precision of traditional AI models remains an underexplored research direction in battery remanufacturing.

Motivated by these challenges and opportunities, this study investigates how CPP-integrated GenAI can enhance the transparency and decision relevance of RUL estimation in battery remanufacturing. We develop an integrated framework that couples explainable machine-learning models with LLM-based explanation modules to produce structured, mechanism-aware narratives grounded in electrochemical principles. Using a public battery dataset, we compare two predictive models, an interpretable tree-based model and an LSTM model, evaluate degradation signatures using feature-attribution methods and assess the capacity of LLMs to convert attribution patterns into domain-consistent explanations. An LLM reviewer is further incorporated to assess correctness, clarity, degradation-mechanism alignment, and remanufacturing usefulness, thereby quantifying explanation quality and identifying areas for improvement.

The main contributions of this work are twofold. First, we developed a CPP-aware RUL estimation framework that integrates traditional RUL prediction model and LLM-based explanation, targeting the practical needs of battery remanufacturing. Second, we demonstrate how explainable AI methods, such as SHAP values, can be combined with structured LLM prompts to produce coherent, physically grounded degradation analyses. The paper is structured as follows. Section 2 provides a review of related work. Section 3 presents the technological framework, detailing the RUL predictive model architecture and the framework that integrates GenAI and traditional AI components. **Section 4** introduces the case study based on a public battery dataset. Section 5 offers a discussion of the key findings, highlighting their implications for battery remanufacturing. Finally, Section 6 concludes the paper and provides recommendations for future research.

2. Related Work

Recent research on battery prognostics has examined a wide range of models for estimating RUL. Feature-based machine learning approaches, such as gradient-boosted decision tree (XGBoost), have been widely used due to their ability to capture nonlinear degradation patterns from engineered voltage–current descriptors [12]. Deep learning architectures, including LSTM networks, extend this capability by modeling temporal dependencies in multicycle aging behavior [13]. However, practical deployment remains challenging, as returned batteries often lack complete cycling histories, exhibit domain shifts between training and real-world operating conditions, and contain noisy or inconsistent measurements [14]. Moreover, most RUL models report only a single predicted value, providing little insight into the degradation mechanisms responsible for the estimate, which limits operational decision-making in remanufacturing facilities.

To address the interpretability limitations of estimating RUL models, explainable XAI techniques have been increasingly integrated into battery health assessment. SHAP has become a commonly used tool because it provides consistent and model-agnostic feature attributions [15]. Studies applying SHAP to batteries have identified degradation signatures linked to internal resistance growth, loss of lithium inventory, and changes in voltage plateau durations, offering deeper insight into the factors driving predicted RUL outcomes [16-18].

CPPs have been proposed as a potential to enhance traceability, data integrity, and lifecycle visibility across manufacturing systems. CPPs can combine multiple interconnected passports, such as product-item passports, machine passports, software passports, human-agent passports, and record passports, to provide a multi-perspective view of the physical and digital states associated with manufacturing processes [7,19]. Applications have been applied in areas such as additive manufacturing, CNC machining, and metal casting, where CPPs have supported part verification, cybersecurity, and process validation [20-22].

LLMs have shown their potential to support decision-making in industrial environments by synthesizing heterogeneous data, interpreting contextual information, and producing task-specific explanations. For example, Men et al. [23] showed that LLMs can interpret sensor anomalies and operational logs to support diagnostic reasoning in engineering systems. Chen et al. [24] integrated LLMs into predictive maintenance workflows, demonstrating that LLMs can summarize machine conditions and assist in scheduling maintenance actions based on historical and real-time data. However, limited research has been conducted on combining CPP, model-based RUL prediction, explainable AI, and LLM-generated degradation narratives within a unified framework to support human operator decision-making in battery remanufacturing.

3. Technology Approach

This study proposed a multi-stage analytical framework (shown in Figure 1) for interpreting the degradation state of lithium-ion batteries and supporting human operator decisions in battery remanufacturing. It comprises four components: (i) information provisioning through digital product records (i.e., CPP, on-site testing) and data processing, (ii) data-driven RUL prediction, (iii) feature-level explainability using interpretable AI models, and (iv) natural-language interpretation using LLMs for the human operator. The overall objective is to integrate data from sources such as CPP, on-site testing, then convert numerical model outputs into physically grounded, technician-interpretable insights that support decision-making during battery remanufacturing.

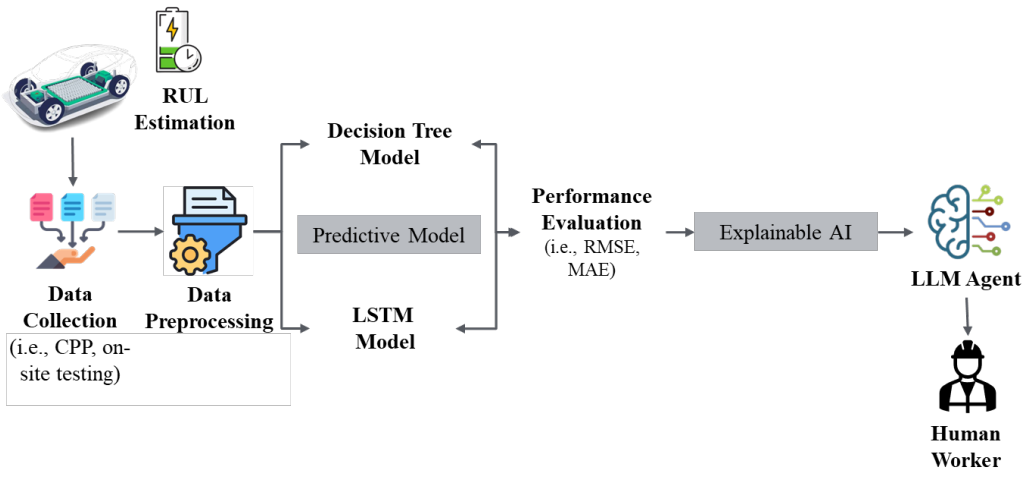

Figure 1. CPP-aware battery RUL estimation framework

3.1 Data Collection

The CPPs and on-site testing could provide such information for the following end-of-life battery data analysis. To prepare the dataset, let $I = \{\theta_{id}, \theta_{mfg}, \theta_{spec}, \theta_{type}, \theta_{diag} \ldots\}$, where denote the set of battery identification θ_{id},

manufacturing information θ_{mfg}, battery specifications θ_{spec}, type θ_{type}. Cycle-level diagnostic information is denoted by θ_{diag}, which could be obtained such as on-site testing. From θ_{diag}, we can extract a set of engineered cycle-level descriptors that characterize degradation-related electrochemical behavior, which we denote as feature vectors $\mathbf{x}_i \in \mathbb{R}^d, i = 1, \ldots, N$. For each cycle i, where d corresponds to engineered descriptors (e.g., discharge duration, voltage plateau duration) and $N$is the total number of observed cycles across the dataset. The corresponding ground-truth RUL is denoted by $y_i \in \mathbb{R}_+$.

3.2 RUL Prediction

We consider a predictive model $\hat{f}: \mathbb{R}^d \to \mathbb{R}_+$, is trained to predict RUL based on engineered features: $\hat{y}_i = \hat{f}(\mathrm{x}_i)$. Model parameters are optimized by minimizing a squared-error loss:

$$L(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{f}(\mathrm{x}_i))^2 .$$

Multiple candidate models (e.g., decision tree, LSTM models) can be evaluated. The model with the lowest held out error (e.g., RMSE) is subsequently selected for downstream explainability.

3.3 Explainable Feature Attribution

For the selected predictive model, we apply SHAP to quantify feature-level contributions. For each sample i, the SHAP decomposition is:

$$\hat{y}_i = \phi_0 + \sum_{j=1}^{d}\phi_{i,j},$$

where ϕ_0 is the model baseline prediction and $\phi_{i,j}$ represents the marginal contribution of feature j for sample i.

To identify the dominant degradation indicators, we extract the indices of the top-k contributors:

$$F_i^{\text{top}} = \arg\max^{(k)}_{j\in\{1,\ldots,d\}} |\phi_{i,j}|.$$

The resulting set F_i^{top}, along with the associated feature values and SHAP scores, provides a compact and interpretable signature of cycle-level degradation behavior and serves as structured input to the LLM.

3.4 LLM-Based Interpreter

Let the structured explanation prompt for sample i be defined as

$$P_i = (y_i,\ \hat{y}_i,\ F_i^{\text{top}},\ \mathbf{x}_i, I^*),$$

which encodes (i) the true and predicted RUL y_i, $\hat{y}_i$, (ii) top k contributors set F_i^{top}, (iii) associated feature values $\mathbf{x}_i$, and (iv) other data from the CPP or on-site testing, denoted by $I^* \in I$. An LLM-based interpreter $E: P_i \to T_i$, generates a technical explanation T_i consisting of four-part sections: (i) RUL Interpretation: relationship between y_i and $\hat{y}_i$, (ii) Feature-Level Insights: explanation of SHAP-derived contributors, (iii) Mechanistic Reasoning: mapping features to known degradation modes (e.g., SEI growth, cathode fatigue), (iv) Remanufacturing Suggestion: recommended decisions (e.g., balancing, diagnostic testing).

3.5 LLM-Based Reviewer

To assess explanation quality, an independent LLM reviewer R, which evaluates T_i on four dimensions:

$$S_i = (s_i^{corr}, s_i^{clar}, s_i^{mech}, s_i^{dec}),$$

corresponding to (i) technical correctness, (ii) clarity, (iii) mechanistic completeness, and (iv) decision usefulness, each on a predefined scale. This produces an explainability-quality dataset to quantify performance of T_i.

```
prompt = f"""

You are an expert in lithium-ion battery degradation and remanufacturing. The battery sample
under analysis is an NMC–LCO 18650 cell with a nominal capacity of 2.8 Ah, cycled over 1000
times at 25 °C under CC–CV charging (0.5C) and 1.5C discharging.


### Battery Condition

- True RUL: {true_rul} cycles

- Predicted RUL: {pred_rul} cycles

1. {top1} (value = {val1:.3f}, SHAP = {shap1:+.3f})

2. {top2} (value = {val2:.3f}, SHAP = {shap2:+.3f})

3. {top3} (value = {val3:.3f}, SHAP = {shap3:+.3f})


Provide a concise technical explanation with the following sections:

#### 1. RUL Interpretation

Interpret the predicted RUL relative to the true RUL and describe the battery's current aging
stage.

#### 2. SHAP Insights

Explain what each top indicator suggests about the degradation process.

#### 3. Degradation Mechanism

Relate the SHAP patterns to known mechanisms, such as SEI growth, cathode fatigue, or
polarization.

#### 4. Remanufacturing Action

Provide a practical recommendation for action: reuse, balancing, downgrade, diagnostic, or
disassembly.

"""
```

Figure 3. LLM-based interpreter prompt example

```
evaluation_prompt =

You are acting as an expert battery reviewer evaluating the correctness of an RUL explanation
for a lithium-ion battery degradation scenario. Use domain knowledge to rigorously score the
explanation. The Battery sample is from an NMC–LCO 18650 with a nominal capacity of 2.8 Ah,
which are cycled over 1000 times at 25 °C under CC–CV charge (0.5C) and 1.5C discharge.

Evaluate the explanation on the following four criteria:

1) Technical correctness (0–5):

   - Is the explanation physically plausible for NMC-LCO lithium-ion cells?

   - Does it reflect the SHAP direction (positive vs negative impact) correctly?

2) Clarity and structure (0–5):

   - Is the explanation organized, readable, and coherent?

   - Does it correctly follow the four required sections?

3) Mechanistic completeness (0–5):

   - Does it correctly map indicators to mechanisms such as SEI growth,

     active material loss, cathode fatigue, or polarization?

4) Decision usefulness (0–5):

   - Does it provide a plausible remanufacturing action (reuse, balancing,

     downgrade, diagnostic testing, or disassembly)?


Provide your answer in the following format:

TECHNICAL_CORRECTNESS: X.X

CLARITY: X.X

MECHANISM: X.
```

Figure 4. LLM-based reviewer prompt example

4. Case Study

In this study, a public battery dataset is used to demonstrate how LLMs can support human-centered decision-making in battery remanufacturing. The dataset consists of 14 lithium-ion batteries from the Hawaii Natural Energy Institute (HNEI) and is used as a representative example of on-site diagnostic testing data available during end-of-life inspection. For each cycle, voltage-time and current-time profiles were summarized into seven engineered descriptors that capture key electrochemical behaviors: discharge time, 4.15V plateau duration, constant-current charging duration, mid-voltage decrement (3.6-3.4V), maximum discharge voltage, minimum charge voltage, and charging time. For each cycle i, these features form a vector $\mathbf{x}_i \in \mathbb{R}^7$, while the corresponding ground-truth RUL y_i denotes the number of cycles remaining until the battery reaches its end-of-life criterion. These $(\mathbf{x}_i, y_i)$ pairs form the supervised learning samples used for model training and evaluation. In addition to diagnostic features, contextual information such as $\theta_{mfg}, \theta_{spec}, \theta_{type}$ can be obtained from the CPPs. This CPP-provided information can be used as trusted contextual input for the LLM agent to support interpretation and remanufacturing recommendations.

To identify the model most suitable for downstream explainability, two predictive models were selected to estimate the battery RUL: a XGBoost and a model based on bidirectional long–short-term memory (BiLSTM) networks. They capture two complementary perspectives on degradation behavior. The XGBoost model operates on individual cycle-level feature vectors and can provide a structured, interpretable mapping from features to predicted RUL. In contrast, the BiLSTM model captures temporal dependencies by learning from sequences of feature vectors over fixed-length sliding windows, enabling the representation of degradation trends across multiple cycles. SHAP was selected to analyze feature contributions for the battery RUL value, which decomposes each prediction into additive contributions from individual features, enabling consistent comparisons across samples. GPT 4.1 is selected for the battery RUL results interpreter and GPT 5 mini as a rigorous reviewer, which is used to provide reasoning and assess the technical quality.

5. Results and Discussion

5.1 Predictive Model Evaluation

The XGBoost and BiLSTM models were trained on the HNEI battery dataset to estimate RUL. Their prediction errors on the held-out test set are summarized in Table 1. XGBoost achieved substantially lower estimation error with an RMSE of 29.73 cycles given around 1000 cycles and an R^2 of 0.991, indicating a very good alignment with empirical degradation patterns in the test data. The BiLSTM model, while still achieving a strong R^2 of 0.9452, showed substantially higher error magnitudes (RMSE = 76.51 cycles) given the HNEI dataset. Therefore, XGBoost is selected as the reference model for downstream interpretability and LLM-based reasoning, given this dataset.

Table 1. RUL prediction model performance.

Model	**RMSE (cycles)**	**MAE (cycles)**	$\boldsymbol{R^2}$
XGBoost	29.73	17.89	0.991
BiLSTM	76.51	60.27	0.947

5.2 Feature Attribution

Feature-attribution analysis was conducted using SHAP to identify the dominant predictors driving the XGBoost model's RUL estimates. Figure 2 illustrates the SHAP summary distribution across all test samples, where each dot represents the marginal contribution of a feature to the predicted RUL and the color scale indicates the corresponding feature value. For example, Discharge time emerges as the most influential variable, with a wide spread of both positive and negative SHAP values. High discharge-time values (shown in red) will increase predicted RUL, whereas reduced discharge duration (blue) strongly decreases the predicted RUL. This suggests that discharge duration can serve as a proxy for underlying electrochemical health and inform early screening decisions for battery reuse or further diagnostics.

The voltage-decrement feature (3.6-3.4 V) shows the opposite pattern: larger decrements (blue) correspond to

negative SHAP contributions. This behavior is consistent with increased internal resistance or degradation effects and provides interpretable signals for identifying batteries with limited remanufacturing potential. Other voltage- and time-based features exhibit comparatively smaller but consistent contributions, indicating secondary influence on RUL estimates given this dataset. Overall, the SHAP analysis not only enhances model interpretability but also supports actionable remanufacturing decisions by linking measurable electrical features to predicted degradation states. These feature-level attributions form the structured explanatory inputs used by the LLM interpreter to recommend justifications for reuse, refurbishment, or recycling pathways for human operators.

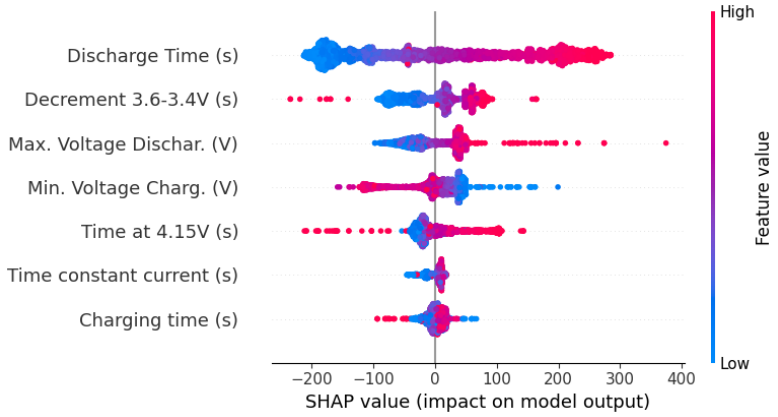

Figure 2. SHAP summary for XGBoost-based RUL prediction.

5.3 LLM-Generated Explanations and Reviewer Assessment

To evaluate the interpretive capabilities of the LLM module, three representative cycles were selected spanning mid- to late-life degradation patterns. The LLM interpreter was tuned by a predefined prompt using a template specifying the required structure of the explanation. The interpreter produced four-part explanations that synthesized the numerical RUL prediction, feature attributions, and domain-consistent degradation reasoning.

An independent LLM reviewer was also used to assess the quality of each explanation under a standardized rubric evaluating technical correctness, clarity, mechanistic completeness, and decision usefulness. The reviewer was guided by a predefined evaluation prompt. Because the raw LLM outputs are lengthy, each explanation is summarized below in a condensed format.

		CASE 1
Interpreter Summary	RUL Interpretation	The predicted RUL (1046.4 cycles) closely matches the true value (1043 cycles), indicating the cell is near the end of its first life but not yet at failure threshold.
	Feature Insights	Positive SHAP contributions from discharge time (+245), 4.15 V plateau duration (+95), and the 3.6-3.4 V decrement (+76) increase the predicted RUL relative to the model baseline. Although the corresponding feature values are slightly below nominal, the model interprets them as indicators of comparatively stable mid-voltage behavior and moderate aging.
	Mechanistic Reasoning	The explanation links the attribution pattern to controlled SEI growth, limited impedance rise, and gradual cathode wear, though it does not fully reconcile why reduced discharge-time values yield positive SHAP contributions.
	Remanufacturing Action	Recommended reuse in low-demand applications following diagnostic confirmation and module-level balancing; disassembly is advised if subsequent tests reveal variance or safety concerns.
Reviewer Evaluation	Scores	Correctness 2.0, Clarity 4.5, Mechanism 3.0, Usefulness 4.0
	Justification	The explanation is clearly structured and broadly consistent with battery degradation behavior but misinterprets the SHAP direction for discharge time,

		reducing technical correctness. Mechanistic reasoning is moderate, while the remanufacturing guidance is appropriate and practical.

CASE 2		
Interpreter Summary	RUL Interpretation	Predicted RUL (471.6 cycles) closely matches the true RUL (469 cycles), indicating a late-life cell exhibiting advanced capacity and power fade.
	Feature Insights	Negative SHAP contributions from discharge time (-43.3), time at 4.15 V (-23.7), and maximum discharge voltage (-12.5) reflect reduced capacity, diminished high-voltage stability, and increased polarization
	Mechanistic Reasoning	Patterns are consistent with SEI thickening, lithium-inventory loss, and cathode structural degradation, all characteristic of battery cells nearing end-of-life.
	Remanufacturing Action	Recommended disassembly or materials recovery; temporary downgrade only if diagnostics confirm acceptable safety margins.
Reviewer Evaluation	Scores	Correctness 3.5, Clarity 5.0, Mechanism 4.0, Usefulness 4.0.
	Justification	Strong organization and plausible mechanistic mapping but identified minor misinterpretation of SHAP sign semantics and recommended inclusion of confirmatory diagnostics.

CASE 3		
Interpreter Summary	RUL Interpretation	Predicted RUL (186.7 cycles) slightly overestimates the true RUL (176 cycles), consistent with late-stage degradation where capacity fade accelerates.
	Feature Insights	Negative SHAP contributions from discharge time (-182.9), mid-voltage decrement (-58.2), and time at 4.15 V (-35.8) indicate declining capacity, rising impedance, and loss of high-voltage charge acceptance.
	Mechanistic Reasoning	The explanation attributes these patterns to cumulative lithium-inventory loss, SEI thickening, polarization growth, and cathode-material fatigue.
	Remanufacturing Action	Suggested low-power repurposing or disassembly following further diagnostics.
Reviewer Evaluation	Scores	Correctness 4.0, Clarity 5.0, Mechanism 4.0, Usefulness 4.0
	Justification	The reviewer affirmed the plausibility and structure of the explanation and recommended more explicit linkage between individual features and mechanisms.

Across the three samples, the LLM interpreter generated structured explanations that generally reflected the dominant SHAP attributions, presented them in a predefined four-part format that understanding of RUL predictions and underlying degradation patterns. Another LLM reviewer, who played as the "experienced technician" assessed the explanation from LLM interpreter. Explanations for Cases 2 and 3 demonstrated good alignment between attribution direction, feature values, and known late-life degradation mechanisms such as SEI growth, polarization, and cathode fatigue, resulting in higher correctness and mechanistic scores. In contrast, Case 1 showed unclear in interpreting attribution direction, particularly the positive SHAP contribution associated with reduced discharge time, leading to its substantially lower correctness rating.

6. Conclusion & Recommendations

This study presents an integrated CPP-aware, RUL prediction, and LLM-augmented framework to support explainable RUL estimation and decision-making in EV battery remanufacturing. Model-based RUL estimation adds value in scenarios where degradation arises from multiple interacting mechanisms and cannot be reliably inferred from remaining capacity alone. By leveraging battery information recorded in CPPs in combination with on-site diagnostic testing and integrating predictive models such as XGBoost and BiLSTM, the framework demonstrates the feasibility of estimating RUL and identifying key degradation indicators. In the presented case study using the HNEI public dataset, the XGBoost model achieved an RMSE of 29.73 cycles and an R^2 of 0.991, serving as a reference

model for downstream explainability analysis. Explainable AI analysis using SHAP identified dominant degradation-related features, which were further contextualized through LLM-generated explanations. These explanations linked numerical predictions to physically mechanistic understanding and actionable remanufacturing recommendations, such as reuse, downgrade, or disassembly. An LLM-based reviewer was further employed to assess technical correctness, clarity, mechanistic alignment, and decision usefulness, illustrating the potential of LLMs to bridge quantitative predictions and technician-interpretable reasoning.

The proposed framework is designed to support decision-making by allowing task-relevant engineered feature snapshots to be selected based on the human operator's objective, using relevant CPP data and on-site diagnostics. It represents a controlled data scenario intended to demonstrate framework feasibility. LLMs provide support for synthesizing heterogeneous data, contextualizing uncertainty, and translating numerical predictions into technician-interpretable remanufacturing recommendations.

Additionally, as product passport initiatives such as the CPP in the United States and the Digital Product Passport (DPP) in the European Union continue to mature and standardize lifecycle and operational data, the proposed framework provides a scalable foundation for leveraging richer information to support more robust, transparent, and trustworthy remanufacturing decisions. Future work will focus on incorporating human expert evaluation of LLM-generated explanations, extending the framework to multiple battery chemistries and failure modes, and validating robustness and usability through pilot studies in industrial remanufacturing settings.

References

1. A. A. Adejare, R. K. Tagayi, J. Kim, Enhancing power substation reliability with second-life battery energy storage systems for dynamic fault mitigation in grid-scale applications, Electric Power Systems Research 241 (2025) 111372.
2. Y. Tan, B. Wang, W. Yuan, Q. Zhou, Y. He, W. Ye, M. C. May, X. Li, S.-G. Kim, and R. Athinarayanan, *Generative manufacturing system for human-aligned and disruption-resilient production*, Proceedings of the ASME 2025 International Mechanical Engineering Congress & Exposition (IMECE2025), American Society of Mechanical Engineers (ASME), Memphis, TN, USA, Paper No. IMECE2025-167119, 2025.
3. S. Khan, A. Haleem, N. Fatma, Effective adoption of remanufacturing practices: a step towards circular economy, Journal of Remanufacturing 12 (2) (2022) 167–185.
4. Chirumalla, K., Kulkov, I., Vu, F., & Rahic, M. (2023). Second life use of Li-ion batteries in the heavy-duty vehicle industry: Feasibilities of remanufacturing, repurposing, and reusing approaches. *Sustainable Production and consumption*, *42*, 351-366.
5. Chen, Z., Liu, W., Zhou, D., Xia, T., & Pan, E. (2025). Inconsistency identification for Lithium-ion battery energy storage systems using deep embedded clustering. *Applied Energy*, *388*, 125677.
6. Chigbu, B. I., Nekhwevha, F. H., & Umejesi, I. (2024). Electric Vehicle Battery Remanufacturing: Circular Economy Leadership and Workforce Development. *World Electric Vehicle Journal*, *15*(10), 441.
7. U.S. Department of Energy, Sandia National Laboratories. *SBOM Use Cases and CyMANII Presentation*. Idaho National Laboratory, 2024.
8. J.-P. Kaiser, S. Lang, M. Wurster, G. Lanza, A concept for autonomous quality control for core inspection in remanufacturing, Procedia CIRP 105 (2022) 374–379.
9. Ramezani, S. B., Cummins, L., Killen, B., Carley, R., Amirlatifi, A., Rahimi, S., ... & Bian, L. (2023). Scalability, explainability and performance of data-driven algorithms in predicting the remaining useful life: A comprehensive review. *Ieee Access*, *11*, 41741-41769.
10. Hadi, M. U., Al-Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., ... & Shah12, M. (2024). LLMs: A Comprehensive Survey of Applications, Challenges, Datasets, Models, Limitations, and Future Prospects.
11. Favier, A., Verma, P., La, N., & Shah, J. A. (2025, May). Leveraging LLMs for Collaborative Human-AI Decision Making. In *Proceedings of the AAAI Symposium Series* (Vol. 5, No. 1, pp. 60-62).
12. Jafari, S., Shahbazi, Z., Byun, Y. C., & Lee, S. J. (2022). Lithium-ion battery estimation in online framework using extreme gradient boosting machine learning approach. *Mathematics*, *10*(6), 888.
13. Safitri, M., Adji, T. B., & Cahyadi, A. I. (2025). Enhanced early prediction of Li-ion battery degradation using multicycle features and an ensemble deep learning model. *Results in Engineering*, *25*, 104235.

14. Zhao, J., Feng, X., Tran, M. K., Fowler, M., Ouyang, M., & Burke, A. F. (2024). Battery safety: Fault diagnosis from laboratory to real world. *J. Power Sources*, *598*(234111), 10-1016.
15. Kumar Kamboj, R., Singh, M., Singh, A., & Bala, A. (2025). Explainable artificial intelligence driven estimation of remaining useful life for lithium-ion battery. *Ionics*, 1-18.
16. Lee, J., & Rew, J. Integrating SHAP with Vision-Language Models for Explainable Remaining Useful Life Prediction of Lithium-ion Batteries. *Available at SSRN 5537792*.
17. Chen, O., Reid, J., & Meier, A. (2025). Explainable AI for Battery Degradation Prediction in EVs: Toward Transparent Energy Forecasting. *Journal of Advances in Engineering and Technology*, *2*(3).
18. Wang, Y., Wang, J., Chen, X., Wang, Y., & Song, J. (2025). Feature Selection for Battery Lifetime Prediction using Explainable Machine Learning. *IEEE Transactions on Industry Applications*.
19. The Cybersecurity Manufacturing Innovation Institute (CyManII). *CyManII IFT Summary for RFP Updates to Peer Review*. CyManII, 2023.
20. Karlee Jenkins. *Increasing cybersecurity of metal casting*. The Ohio State University Center for Design & Manufacturing Excellence, 2024.
21. Grimes, H. D., Ciocarlie, G. F., Butler, R. J., & Austad, W. E. (2022). *Building Blocks for Secure and Prosperous Defense Critical Supply Chains: A Case Study from Microelectronics* (No. INL/MIS-22-70124-Rev000). Idaho National Laboratory (INL), Idaho Falls, ID (United States).
22. Kurfess, T., & Grimes, H. (2021). The role of the digital thread for security, resilience, and adaptability in manufacturing. *The Bridge (Washington, DC)*, *51*(1).
23. Men, C., Han, Y., Wang, P., Tao, J., & Huang, C. G. (2025). The interpretable reasoning and intelligent decision-making based on event knowledge graph with LLMs in fault diagnosis scenarios. *IEEE Transactions on Instrumentation and Measurement*.
24. Chen, X., Lei, Y., Li, Y., Parkinson, S., Li, X., Liu, J., ... & Zhao, Z. (2025). Large Models for Machine Monitoring and Fault Diagnostics: Opportunities, Challenges, and Future Direction. *Journal of Dynamics, Monitoring and Diagnostics*, *4*(2), 76-90.

About the Author(s)

Dr. Athinarayanan conducts interdisciplinary research in generative AI, digital twins, cyber-physical passports, and smart manufacturing systems. He directs Purdue's DOE/CESMII Smart Manufacturing Innovation Center and leads major DoD, NSF, DOE, and industry-supported projects in industrial sustainability, automation, and workforce innovation.

Dr. Li is currently an Assistant Professor in the School of Engineering Technology at Purdue University, West Lafayette. Before joining Purdue, he was an Adjunct Assistant Research Scientist at the Department of Mechanical Engineering at University of Michigan - Ann Arbor.

Yanchao Tan is a Ph.D. student in Engineering Technology with academic training in both Computer Science and Systems Engineering. Her current research focuses on integrating artificial intelligence techniques with industrial decision-support systems,