MateriAI

Transforming terabytes of unused corrosion data into actionable insights with AI

Alex Makogon
makogon422833@gmail.com

Paris
November 2024

Problem Statement:

The global cost of corrosion and material degradation is staggering, estimated at US$2.5 trillion annually, equivalent to 3.4% of the global GDP[2][5]. This enormous economic burden affects various industries, from infrastructure and transportation to energy and manufacturing. While traditional mitigation strategies have shown some success, there is a pressing need for more advanced, data-driven approaches to combat this pervasive issue.

Current Mitigation Strategies

Existing corrosion management practices include:

1. Risk assessment and corrosion mapping
2. Material selection and design considerations
3. Protective coatings and surface treatments
4. Cathodic protection systems
5. Regular monitoring and inspection programs
6. Specialized control measures for specific environments
7. Training and education programs[3]

These strategies have demonstrated potential savings of 15-35% of the total cost of corrosion, translating to US$375-875 billion annually on a global scale[2][5].

Transition to Data-Driven Approaches

Despite the effectiveness of current practices, the complexity of corrosion and material degradation processes necessitates more sophisticated, data-driven solutions. Advanced monitoring technologies, such as "smart pigs" used in pipeline inspections, can generate up to 1 terabyte of data per run[5]. This influx of big data, combined with information from scientific experiments, industrial monitoring, and computational simulations, creates a vast pool of potentially valuable insights[1].

Challenges in Data Analysis

However, the sheer volume of data generated poses significant challenges:

1. **Data Utilization**: Studies suggest that less than 50% of collected corrosion-related datasets are effectively used for actionable insights[1].
2. **Data Fragmentation**: Corrosion and degradation data are often siloed within industries or research groups, limiting accessibility and comprehensive analysis[1].
3. **Standardization Issues**: The lack of centralized global repositories and inconsistent data formats hinder large-scale analysis efforts[1][5].

4. **Limited Advanced Analytics**: Many sectors have not fully adopted machine learning or big data analytics tools, leaving valuable data unprocessed[1].
5. **Scope of Analysis**: Current approaches often focus on specific factors or environments, rather than comprehensive, multi-factor analyses[4].

The Need for Automated Analysis

The vast majority of corrosion and material degradation data remains underutilized due to these challenges. Advanced software solutions for automated analysis are crucial to:

1. Process and integrate diverse datasets from various sources
2. Apply machine learning and AI techniques to uncover patterns and improve predictive models
3. Standardize data formats and create accessible repositories
4. Enhance decision-making in corrosion management and material selection
5. Extend analysis beyond traditional corrosion to include degradation of advanced materials like battery components and photovoltaics

By addressing these issues, automated analysis software can unlock the full potential of the enormous amount of data generated in the field of material degradation, leading to significant cost savings, improved safety, and more sustainable material use across industries.

Solution

Our solution is an **agentic LLM framework** designed to revolutionize the analysis of corrosion and material degradation data across various industries, including energy, infrastructure, and manufacturing. This innovative system:

1. **Predefined Tools and Transformation Techniques**: Includes standard machine learning algorithms like dimensionality reduction, as well as specialized tools for processing and interpreting electrochemical impedance spectral data. For instance, we have tested the applicability of dimensionality reduction for processing large volumes of electrochemical impedance data in a recent study[8]. Other tools include:
    - **Corrosion Rate Prediction**: Using machine learning models like random forests for accurate corrosion rate prediction, as demonstrated in a study on low-alloy steels in marine environments[2].
    - **Material Property Analysis**: Quantifying material properties and their degradation rates through sophisticated algorithms.
    - **Environmental Impact Assessment**: Evaluating the influence of environmental factors on corrosion processes.
2. **Agentic Framework Decision-Making**: The framework autonomously determines the most appropriate tools for each dataset based on its characteristics and the analysis

goals. This includes:

- **Selection of Data Analysis Algorithms**: Choosing algorithms like principal component analysis, t-distributed stochastic neighbor embedding (t-SNE) or Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction[1].
- **Custom Tool Integration**: Incorporating specialized tools like electrochemical impedance spectroscopy (EIS) fitting and quantification techniques[4].

3. **Image Data Processing**: Equipped with an array of tools for handling image data:

    - **Optical Images**: Analyzing visual data from corrosion sites, including high-resolution images of corroded surfaces.
    - **SEM Images**: Processing scanning electron microscopy (SEM) images to identify microstructural changes or degradation patterns.

4. We leverage fine-tuned visual-language models to interpret these images in context, enhancing the understanding of corrosion mechanisms[5].

5. **Data Correlation and Transformation**: Once the tools are selected, the framework identifies the necessary transformations required to make the data compatible with these tools. This includes:

    - **Data Integration**: Combining datasets from different sources or formats, such as merging corrosion rate data with environmental exposure data[3].
    - **Feature Engineering**: Extracting relevant features from raw data to improve model performance[1].

6. **Correlative Analysis**: The framework allows for the correlative analysis of various data types, like:

    - **Images and Tabular Data**: Combining visual degradation patterns with material properties or environmental factors to uncover correlations previously hidden by data fragmentation[7].

7. **Visualization and Insights**: Finally, the agentic framework proposes visualizations to extract new insights from available corrosion data:

    - **Data-Driven Insights**: Visualizing corrosion trends, material degradation patterns, and predictive modeling results to facilitate decision-making[3].
    - **Interactive Dashboards**: Providing users with interactive tools to explore and understand their data in a more intuitive manner[5].

This agentic LLM framework represents a significant leap forward in the field of corrosion and material degradation analysis, enabling the efficient processing and integration of large volumes of unstructured or loosely structured data. By leveraging advanced machine learning techniques, it not only processes data more effectively but also draws actionable insights that can inform corrosion management strategies, material selection, and predictive maintenance, ultimately reducing operational costs and enhancing the longevity of critical assets.