

# Integrating AI for Improved Correction of E-commerce Product Data

Naveen Koka

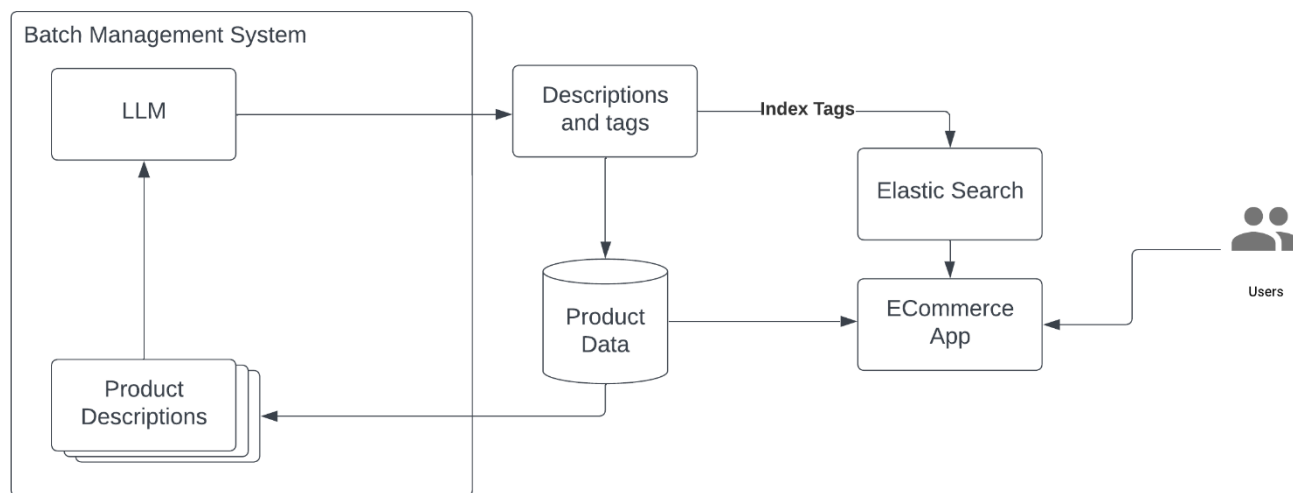


Fig 1: Description Optimizations

## ABSTRACT

In today's competitive e-commerce landscape, the accuracy and appeal of product descriptions are critical factors influencing customer purchasing decisions and platform performance. This paper explores innovative approaches to enhance e-commerce product data through the integration of advanced technologies such as Large Language Models (LLMs) and efficient batch processing methods. By categorizing products effectively and extracting pertinent details from product information and customer feedback, businesses can streamline the process of generating compelling and SEO-friendly product descriptions.

The implementation strategy involves leveraging AWS Batch for scalable batch processing and OpenAI's LLMs for natural language understanding. These technologies enable automated analysis and optimization of product data, ensuring consistency and relevance across vast product catalogs. Moreover, integrating libraries like LangChain enhances the system's elasticity and adaptability, facilitating seamless data management and correction without disrupting operational workflows.

Case studies in automotive and clothing e-commerce sectors illustrate the practical application of these methodologies. By harnessing AI to refine product descriptions and extract meaningful search tags, businesses can enhance product discoverability and customer engagement. This holistic approach not only improves the efficiency of data processing but also positions companies to meet the dynamic demands of modern consumers, ultimately driving sales and fostering brand loyalty in the competitive digital marketplace.

## Keywords

Ecommerce, Auto, Clothing, AI, Search, LLM, Improve descriptions

## 1. INTRODUCTION

In the dynamic world of e-commerce, the quality of product data plays a pivotal role in driving sales and enhancing customer experience. Accurate and engaging product descriptions are essential for capturing customer interest, ensuring searchability, and catering to specific buyer needs. This paper explores the integration of AI, particularly Large Language Models (LLMs), and batch processing techniques to improve the correction and enhancement of e-commerce product data. By efficiently categorizing products, extracting relevant details from product information and reviews, and leveraging advanced technologies for data processing and description enhancement, businesses can significantly optimize their product listings. This leads to improved visibility, higher engagement, and increased sales across various e-commerce sectors, including automotive and clothing.

## 2. PROBLEM STATEMENT

In the e-commerce industry, product data is a crucial driver of sales. Elements such as accurate images, detailed product descriptions, and proper categorization significantly influence customer purchasing decisions. For product manufacturers, having control over the correct placement and presentation of this information is essential to ensure that their products attract the right audience and achieve optimal sales performance.

However, e-commerce platforms often aggregate products from various manufacturers and distributors, resulting in a highly volatile environment. This volatility can lead to inconsistencies such as incorrect images or misleading descriptions, which negatively impact the platform's overall user experience. These inaccuracies can deter potential customers, leading to lost sales and diminished brand trust.

Consequently, e-commerce platforms are forced to allocate substantial time and resources to manually correct this data, involving considerable human intervention.

The challenge becomes even more daunting when dealing with millions of products, as seen in industries like clothing and automobile parts. The sheer volume of data makes it impractical to rely solely on manual correction processes. The need for efficient, scalable solutions to manage and rectify product data is evident, highlighting the importance of integrating advanced technologies to streamline these operations and maintain data integrity across vast product catalogs.

### 3. SOLUTION

Leveraging programming and AI, particularly Large Language Models (LLMs), offers a promising solution to correct product data based on names, descriptions, and reviews. By harnessing the power of LLMs, we can automate the process of analyzing and rectifying product information with greater accuracy and efficiency. These models can understand and interpret product data, identifying discrepancies and inconsistencies that might be missed by manual review processes.

LLMs can analyze product names and descriptions to ensure they are accurate and properly categorized. By comparing the provided information against a vast dataset of similar products, LLMs can suggest corrections and improvements, enhancing the overall quality and consistency of the product data. Additionally, by examining customer reviews, LLMs can identify common issues or complaints related to product descriptions or images, allowing for proactive adjustments to be made.

The integration of LLMs in e-commerce platforms not only reduces the reliance on manual data correction but also significantly enhances the speed and scalability of these operations. As a result, platforms can maintain high-quality product data across extensive catalogs, improving user experience and driving sales. The use of advanced AI technologies thus represents a transformative approach to managing and optimizing product data in the e-commerce industry.

#### 3.1 Batches

Dealing with millions of product data entries requires a robust and efficient approach to ensure accuracy and consistency. One effective method is to use the concept of batches, a tried-and-true technique that allows for large-scale data processing without disrupting production systems. By running the data correction processes in batches, we can handle vast amounts of information systematically and efficiently.

##### 3.1.1 Scheduled Batch Processing

To minimize the impact on production systems, batches are executed during non-business hours. This scheduling ensures that the heavy computational load associated with data correction does not interfere with the platform's regular operations, such as customer transactions and real-time data updates. By running these processes when user activity is low, we can maintain optimal performance and user experience during peak hours.

##### 3.1.2 Focused on New Product Data

The batch processing approach targets new product data, ensuring that recent additions to the catalog are promptly and accurately corrected. This method allows for continuous improvement and maintenance of data quality, keeping the product information up-to-date and reliable. By systematically addressing new data in batches, we can efficiently manage and rectify inaccuracies, ultimately enhancing the overall quality and integrity of the e-commerce platform's product listings.

#### 3.2 LLM

Selecting the appropriate Large Language Model (LLM) library is crucial for the success of our data correction efforts. The right library will provide the necessary tools and capabilities to accurately process and enhance product data, ensuring it meets the high standards required for effective e-commerce operations.

When choosing an LLM library, performance and scalability are key considerations. The library must handle large volumes of data efficiently, processing millions of product entries without compromising speed or accuracy. It should also offer scalability to accommodate future growth, ensuring that as the product catalog expands, the library can continue to deliver reliable results.

Another critical factor is the ease of integration and compatibility with existing systems. The LLM library should seamlessly integrate with our current tech stack, including data storage, processing frameworks, and other AI tools. Compatibility ensures that we can implement the library without extensive modifications to our infrastructure, reducing deployment time and costs.

The chosen LLM library should offer advanced capabilities in natural language understanding and generation. It should be proficient in interpreting product names, descriptions, and customer reviews, identifying and correcting inaccuracies effectively. Additionally, the library should support custom training to fine-tune the model based on specific product categories and industry nuances, enhancing its precision and relevance.

### 4. IMPLEMENTATION

To kickstart our project, we need to set up the batch management tool and the LLM that will be the backbone of our application. Given the numerous tools available, we can leverage existing solutions to avoid reinventing the wheel and ensure robust performance.

For example:

AWS Batch allows us to efficiently run hundreds of thousands of batch computing jobs by dynamically provisioning the optimal quantity and type of compute resources based on the volume and specific requirements of the batch jobs. It simplifies the process of managing and scaling batch workloads, ensuring that our data correction tasks are executed smoothly and efficiently.

OpenAI offers powerful models that can be integrated into our system to handle natural language processing tasks with high accuracy. OpenAI's models are well-suited for interpreting and correcting product data based on names, descriptions, and reviews.

To ensure scalability and elasticity in our application, we can use libraries like LangChain. LangChain provides a flexible framework for building applications with language models, enabling us to handle complex workflows and large-scale data processing tasks effectively.

#### 4.1 Product Descriptions

A compelling product description can grab a customer's attention and create a positive first impression. By using engaging language, clear formatting, and highlighting key features, descriptions can make products more appealing. This increases the likelihood that customers will explore further and consider making a purchase. Emotional appeal, storytelling, and relatable scenarios can also play a significant role in capturing customer interest.

##### 4.1.1 Categorizing Products

To optimize product descriptions and enhance overall e-commerce performance, the first step is to categorize products efficiently. This involves identifying the product category, understanding the target audience, and extracting relevant details from the product information and customer reviews. This data can then be passed to a Large Language Model (LLM) to retrieve accurate categories and target demographics, followed by refining the product descriptions.

- Extract product information such as name, specifications, features, and usage.
- Analyze customer reviews and feedback to identify common themes and attributes.
- Use this data to determine the product category (e.g., electronics, clothing, automotive parts).

##### 4.1.2 Leveraging LLM for Categorization

- Provide the LLM with the extracted product information and customer reviews.
- Query the LLM to classify the product into appropriate categories.
- Retrieve insights on the target audience, including demographic details and preferences.

An example of llm prompt as below

Given the following product information and customer reviews, classify the product into a category and identify the target audience:

Product Information:

- Name: Roof Rack
- Features: Rack that fits any car. Extensibility based on the size of the car.
- Usage: Can carry luggage, Includes lock to secure contents in the box

Customer Reviews:

- we've traveled with it approx 2 6-8 hrs trips and it's held up well. The only reason we didn't give it 5 stars is that it said it was compatible with
- easily held all of our luggage for a weeks stay plus more and allowed us to have lots of room in the cabin of our car

##### 4.1.3 Enhancing Product descriptions

- Combine the product description with the categorized data and target audience insights.

Use the LLM to generate improved, engaging, and SEO-friendly product descriptions.

Improve the following product description using the product category and target audience details:

Product Category: Roof rack fits for any car

Target Audience: Supports car, suvs, etc.,

Original Description: The car roof bag adopts a smooth curved slope design, which is designed to reduce the air resistance of the vehicle while driving, thereby improving fuel efficiency and increasing stability. We also use orange and silver reflective strips to serve as warnings in dark environments to ensure your driving safety.

Description: The car roof bag adopts a smooth curved slope design, which is designed to reduce the air resistance of the vehicle while driving, thereby improving fuel efficiency and increasing stability. We also use orange and silver reflective strips to serve as warnings in dark environments to ensure your driving safety. Provide me enhanced description

Enhanced Description: Elevate your vehicle's functionality with our versatile roof rack, designed to fit any car, SUV, and more. Engineered with a sleek, smooth curved slope, this car roof bag minimizes air resistance, enhancing your vehicle's fuel efficiency and stability on the road. For added safety, we've incorporated high-visibility orange and silver reflective strips, providing clear warnings in low-light conditions to ensure safe driving. Whether you're embarking on a long road trip or just need extra storage for daily commutes, this roof rack offers the perfect blend of efficiency, safety, and style.

##### 4.1.4 Summary

Develop scripts to extract relevant product information and customer reviews.

Organize the data into structured formats for easy processing.

Use APIs to interact with the LLM for categorization and description enhancement.

Send product information and reviews to the LLM to retrieve categories and target audience details.

Pass the original descriptions along with categorized data to the LLM.

Generate and refine enhanced descriptions to ensure they are engaging and SEO-friendly.

Implement batch processing using tools like AWS Batch to handle large volumes of data.

Schedule processing during non-business hours to avoid impacting production systems.

#### 4.2 Search Tags

Effective search functionality is crucial for quickly presenting relevant results to consumers, capturing their attention, and driving sales. To optimize search, we need to extract relevant tags from product details and descriptions, enabling better indexing by tools like Elasticsearch. By enhancing our prompt to include these tags, we can improve search performance and user experience.

We'll adjust our approach to include both the optimized description and the extracted search tags in a JSON format.

Enhance the prompt like below

improve the following product description and extract relevant search tags:

Provide the output in JSON format with 'optimized\_description' and 'tags'.

providing clear warnings in low-light conditions to ensure safe driving. Whether you're embarking on a long road trip or just need extra storage for daily commutes, this roof rack offers the perfect blend of efficiency, safety, and style.", "tags": [ "roof rack", "car roof bag", "vehicle accessory", "SUV roof rack", "fuel efficiency", "aerodynamic design", "stability", "reflective safety strips", "road trip storage", "extra vehicle storage"]}

```
{
  "optimized_description": "Elevate your vehicle's functionality with our versatile roof rack, designed to fit any car, SUV, and more. Engineered with a sleek, smooth curved slope, this car roof bag minimizes air resistance, enhancing your vehicle's fuel efficiency and stability on the road. For added safety, we've incorporated high-visibility orange and silver reflective strips,
```

## 5. USES

### 5.1 Ecommerce for Auto

In the automotive e-commerce industry, precise and detailed product descriptions are essential for helping customers find the exact parts or accessories they need. This sector often deals with complex products that require accurate specifications and compatibility information. Enhancing product descriptions with LLMs can significantly improve the searchability and accuracy of listings. For instance, products like roof racks, tires, and engine components can have detailed descriptions that highlight features, compatibility, and benefits. Additionally, extracting and optimizing search tags ensures that customers can easily find the right products using search tools like Elasticsearch.

### 5.2 Ecommerce for Clothing

In the clothing e-commerce industry, compelling and informative product descriptions are crucial for capturing customer interest and driving sales. Descriptions need to convey not only the style and design but also the material, fit, and care instructions. By using LLMs, clothing descriptions can be enhanced to appeal to target demographics, highlight unique features, and include relevant keywords for better searchability. Extracting tags from descriptions helps in categorizing products effectively, making it easier for customers to find exactly what they are looking for.

## 6. CONCLUSION

In the highly competitive e-commerce industry, effective product descriptions are crucial for capturing customer interest, ensuring searchability, and meeting specific buyer needs. Leveraging advanced technologies such as Large Language Models (LLMs) and batch processing can significantly enhance the quality and accuracy of product data. By categorizing products efficiently and extracting relevant details from product information and customer reviews, we can tailor descriptions to target audiences and improve overall user experience.

This approach ensures the creation of high-quality, data-driven reports, enhancing business decision-making and operational efficiency. Ultimately, the combination of advanced AI technologies and efficient batch processing techniques provides a scalable solution for managing large volumes of product data. This approach not only improves data accuracy and searchability but also positions businesses to better meet the evolving needs of their customers.

## 7. REFERENCES

- [1] Shanmugapriya, S & Subramani, Pavithra. (2024). ARTIFICIAL INTELLIGENCE AND E-COMMERCE.
- [2] Abid Haleem, Mohd Javaid, Mohd Asim Qadri, Ravi Pratap Singh, Rajiv Suman, Artificial intelligence (AI) applications for marketing: A literature-based study, International Journal of Intelligent Networks, Volume 3, 2022, Pages 119-132, ISSN 2666-6030, <https://doi.org/10.1016/j.ijin.2022.08.005>.
- [3] Fedorko, Richard & Král', Štefan & Bačík, Radovan. (2022). Artificial Intelligence in E-commerce: A Literature Review. 10.1007/978-981-16-9113-3\_50.
- [4] Serbin, Vitalii. (2023). Artificial intelligence as a tool for creating dynamic product descriptions in e-commerce. Marketing and Digital Technologies. 7. 79-94. 10.15276/mdt.7.4.2023.5.
- [5] Abu Daqar, Mohannad & Smoudy, Ahmad. (2019). The Role of Artificial Intelligence on Enhancing Customer Experience. International Review of Management and Marketing. 9. 22-31. 10.32479/irmm.8166.
- [6] Ahmed, Shezan. (2023). The Role of Artificial Intelligence in enhancing Customer experience in E-commerce.
- [7] Tiutiu, Miriam & Dabija, Dan-Cristian. (2023). Improving Customer Experience Using Artificial Intelligence in Online Retail. Proceedings of the International Conference on Business Excellence. 17. 1139-1147. 10.2478/picbe-2023-0102.