

Article

A Large Language Model and Qualitative Comparative Analysis-Based Study of Trust in E-Commerce

Laleh Davoodi  and József Mezei 

Faculty of Social Sciences, Business and Economics, and Law, Åbo Akademi University, 20500 Turku, Finland; laleh.davoodi@abo.fi

* Correspondence: jozsef.mezei@abo.fi

Abstract: The primary goal of this study is to predict and analyze customer trust in e-commerce by leveraging neural computation within large language models (LLMs) alongside configurational approaches. We employ LLMs to predict trust levels based on customer reviews, applying artificial intelligence to analyze key aspects of the e-commerce experience, such as customer service, refund processes, item quality, and shipping. To extend beyond predictive performance, we integrate Qualitative Comparative Analysis (QCA) to identify the causal relationships between trust and various stages of the customer journey, including selection, delivery, and post-purchase support (recovery). This dual approach not only showcases the power of neural computation in predicting trust outcomes but also provides a deeper understanding of how specific configurations of customer experience elements contribute to either positive or negative trust. By combining machine learning techniques and QCA, this study contributes to the application of LLMs and configurational approaches, offering novel insights into the drivers of trust in e-commerce.

Keywords: large language models; e-commerce; customer feedback; trust; QCA



Citation: Davoodi, L.; Mezei, J. A Large Language Model and Qualitative Comparative Analysis-Based Study of Trust in E-Commerce. *Appl. Sci.* **2024**, *14*, 10069. <https://doi.org/10.3390/app142110069>

Academic Editors: João M. F. Rodrigues, Pedro J. S. Cardoso and Marta Chinnici

Received: 11 October 2024

Revised: 28 October 2024

Accepted: 30 October 2024

Published: 4 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the increasing shift to online shopping, understanding customer trust has become a critical focus for e-commerce businesses aiming to retain customers and enhance their experiences. Trust is a key factor in shaping customer decisions, and it is influenced by various stages of the customer journey. Customer feedback plays a critical role in the e-commerce industry, offering a great source of information about customer experiences, preferences, and pain points. By systematically analyzing this feedback with various machine learning models [1], businesses can identify trends, improve customer satisfaction, and make data-driven decisions to foster growth and innovation.

However, analyzing the vast amount of customer feedback efficiently is a significant challenge, necessitating advanced computational methods. Neural computation, through the use of large language models (LLMs), offers a sophisticated approach to this challenge, enabling the automation of feedback analysis and providing predictive insights into trust levels [2]. These models, rooted in deep learning architectures [3], are capable of understanding and predicting sentiments within customer reviews, unlocking new possibilities for analyzing trust at scale. This study investigates the use of LLMs to predict customer trust and understands its relationship with different stages of the customer journey in e-commerce by analyzing user feedback.

While neural computation has demonstrated substantial power in predictive performance, capturing trust's nuanced dynamics across the e-commerce customer journey requires other approaches. Existing predictive models, though effective, do not always provide actionable insights regarding the interactions between various stages of the journey and their cumulative impact on customer trust. To bridge this gap, we incorporate Qualitative Comparative Analysis (QCA) [4], which allows us to explore the configurations of

customer experience elements that lead to positive or negative trust outcomes. By making use of the predictions of LLM-based sentiment analysis as the input for a QCA analysis, we are able to shift from purely predictive performance to a deeper understanding of how different conditions—such as app experience, delivery quality, and post-purchase support—interact to shape trust outcomes [5].

Therefore, to build upon the limited research on the relationship between the online customer journey and customer trust, we aim to address the following research question (RQ): “How can we utilize advanced ML models to analyze online customer reviews and understand the impact of different aspects on customer trust?” We answer this question by fine-tuning LLMs on a dataset consisting of more than 900 online reviews from e-commerce platforms, and we conduct configurational modeling with QCA.

By fine-tuning advanced LLMs, we aim to accurately identify key aspects of the shopping experience and their corresponding sentiments. For this purpose, we collected and manually annotated customer reviews from Trustpilot (<https://www.trustpilot.com/>, accessed on 8 July 2021), focusing on five leading e-commerce platforms. This dataset was used to train and evaluate two LLMs: TinyPixel/Llama-2-7B-bf16-sharded and Gemma. By offering a qualitative analysis of the outcome of these models in combination with QCA analysis, we aim to determine the impact of different aspects on customer trust and to identify potential improvements in e-commerce services based on these insights. The results highlight the effectiveness of LLMs in understanding and predicting customer trust, providing actionable business insights from raw textual data.

This paper is structured as follows. Section 2 provides a literature review, covering customer trust in e-commerce and the application of LLMs in the business domain. Section 3 details the research methodology, including data acquisition, preparation, and annotation, language models, and QCA. Section 4 presents the main results of the LLM and QCA analysis. Section 5 offers a discussion, elaborating on the findings, and finally, Section 6 concludes this research.

2. Literature Review

In this section, we will provide an overview of the academic literature relevant to our research. We begin by discussing e-commerce and customer trust, highlighting the key factors that influence customer trust, as these will be essential for analyzing online reviews. Additionally, we will delve into recent applications of LLMs in e-commerce and business, showcasing how these advanced models are being utilized in contemporary research and practice.

2.1. Customer Trust in E-Commerce

E-commerce, short for electronic commerce, involves the trading of goods and services through electronic means. It encompasses various types of transactions, including B2B (business-to-business), B2C (business-to-consumer), and C2C (consumer-to-consumer), which facilitate the development of consumer preferences and markets by establishing electronic infrastructures to address future challenges [6].

For an online business to become a truly successful medium for consumer spending, its technologies must match the benefits provided by traditional high-street retail formats. These benefits can be, for example, immediate delivery, credit facilities, various payment methods, product displays, personal assistance in selecting goods, return services, and warranties [7]. Despite the numerous benefits of e-commerce, companies have encountered specific challenges. Ensuring customer satisfaction and building trust in e-commerce operations has proven to be a complex task, but it is crucial for the long-term growth of businesses [8,9].

Ref. [10] describes trust in e-commerce as a set of beliefs held by an online consumer concerning certain attributes of the seller, such as integrity, benevolence, and competence. These beliefs influence the consumer’s willingness to engage in online transactions [10]. Their study highlights that trust in e-commerce is not a single, uniform concept but consists

of multiple dimensions. It demonstrates that consumers evaluate web vendors based on specific attributes rather than general impressions. Even after a single interaction, consumers can separately assess a vendor's honesty and benevolence based on their competence. This indicates that consumers may quickly form beliefs about a vendor's honesty and goodwill without necessarily considering them competent enough to earn their business, or vice versa.

According to [11], trust is also a multifaceted concept, with key dimensions of competence, integrity, and benevolence. Competence involves a company's ability to fulfill promises, integrity reflects consistent and honest behavior, and benevolence indicates prioritizing customers' interests. These dimensions, while distinct, are interconnected and together build overall consumer trust in business-to-consumer relationships.

Consumers often hesitate to engage with online vendors due to uncertainties about vendor behavior and the risk of personal information theft. Trust is essential in mitigating these concerns, as it helps consumers feel secure in sharing information, making purchases, and acting on vendor advice, which is critical for the widespread adoption of e-commerce [10].

In addition to trust, there are several factors influencing online consumer shopping behaviors. For instance, price has become a significant factor in purchasing decisions. According to [12], respondents considered the lower prices offered compared to conventional stores and the added benefit of no transportation costs.

Another important factor is customer service. Customer service encompasses a series of activities aimed at enhancing customer satisfaction, including pre-transaction, transaction, and post-transaction. Previous studies indicate that enterprises that build excellent customer relationships can increase their profits by more than 60% [13–16].

Ease of returns and exchanges, identified as a crucial yet unsatisfactory factor, is cited by 42% of online shoppers as needing improvement. Given its importance and the clear demand for better service, retailers should prioritize enhancing this aspect [17]. An instant refund offers an efficient process for online returns by issuing the refund before the seller receives the returned product, resulting in a faster refund speed compared to conventional methods. This enhanced speed can significantly improve the consumer's return experience [18].

Moreover, consumers express a desire for improvements in the variety of brands and products and online tracking capabilities, despite already high overall satisfaction with these factors [17].

E-payment plays a crucial role in modern business transactions due to its timeliness, efficiency, and convenience. It encompasses various electronic payment methods, including debit cards, mobile banking, credit cards, and online banking. Studies have identified factors influencing the use and adoption of e-payment systems, highlighting that an inefficient e-payment system can impede the success of e-commerce development [19–21].

Shipping also plays a significant role in customer satisfaction and trust. Customers are increasingly seeking higher-quality in physical distribution services, with a particular emphasis on faster delivery times. Feeling "time-starved", many customers experience anxiety and dissatisfaction when faced with lengthy delivery periods [22,23].

Online environment factors, including website quality, trust, and empathy, significantly influence customer purchase intention. According to Fazli et al., usability, website design, information quality, trust, and empathy are positively associated with online purchase intention. These findings highlight the importance of enhancing website quality, trust, and empathy to boost online purchase intention [24].

Packaging involves designing and producing a product's container or wrapper, encompassing both design and manufacturing activities. Its primary function is to preserve product quality through protection. Product packaging design captures consumer interest and helps with brand recall. Therefore, to ensure continued consumer engagement, it is crucial to maintain an appealing packaging design [25].

Another important factor influencing customer satisfaction and trust is product quality. According to [26], product quality significantly enhances customer satisfaction.

Finally, product descriptions, product photos, and ratings and reviews all positively influence purchase intention in e-commerce. High-quality product descriptions, appealing photos, and positive ratings and reviews enhance customers' intention to purchase [27].

2.2. LLMs in Business Domain

Neural computations and machine learning techniques have revolutionized the way businesses analyze and understand customer behavior in e-commerce [28]. These advanced methods, powered by deep learning and neural networks, enable the processing of vast amounts of customer data, such as reviews, transaction histories, and clickstream data, to uncover patterns that were previously difficult to detect [29]. One of the most significant developments in this area is the use of Large Language Models (LLMs), which leverage neural computation to interpret and predict customer sentiments, preferences, and trust levels or predict the relevance of information based on textual data [30]. These models, trained on massive datasets, can comprehend the subtleties of human language, allowing businesses to gain nuanced insights into customer satisfaction, decision-making, and overall experience. As e-commerce continues to grow, the ability to analyze customer feedback at scale is critical for creating personalized recommendations, improving customer service, and driving loyalty. LLMs such as GPT and BERT offer a unique advantage by not only identifying customer needs but also predicting future behavior based on past interactions. Given their capacity to handle complex language patterns and adapt to diverse contexts, LLMs have immense potential to transform customer behavior analysis in e-commerce, providing businesses with powerful tools to enhance engagement and trust.

LLMs are increasingly being used in the business and e-commerce domains for various applications. In the literature, LLMs are leveraged for their advanced Natural Language Processing (NLP) capabilities, which allow for sophisticated analysis and automation of text-related tasks. LLM can be used to automate the customer service process. For example, LLMs are employed to create chatbots and virtual assistants that can understand and respond to customer inquiries with high accuracy, providing 24/7 support [31]. Additionally, they are used to automate responses to frequently asked questions, reducing the workload on human agents [32,33].

Another application of LLMs in business is the analysis of customer feedback, reviews, and social media posts to gauge public sentiment toward products and brands. By leveraging these advanced models, businesses can gain insights into customer opinions, detect emerging trends, and make well-informed decisions based on the data collected [29,34].

Moreover, LLMs enhance recommendation systems by analyzing user behavior, preferences, and feedback to suggest relevant products [35,36]. They provide more accurate and personalized recommendations compared to traditional methods.

Additionally, LLMs can improve the accuracy and relevance of search results on e-commerce platforms by understanding the context and intent behind user queries [37]. They enhance the overall search experience for users. To the best of our knowledge, this study is the first to explore the capability of LLMs in predicting trust sentiment and its associated aspects within the e-commerce domain.

3. Materials and Methods

In this section, we detail the data acquisition and processing methods. We also provide a comprehensive explanation of how we built and trained two of the most powerful LLM models. Additionally, we present the performance results of these models and introduce the best-performing model. In order to complement the analysis with LLMs, we perform configurational analysis, and in this section, we present the basics of QCA with the results of necessity analysis.

3.1. Data Gathering and Annotation

As stated above, with the advancement of internet technologies, customers have been sharing a vast amount of data online, expressing their views about various products and

services. Companies can leverage these data to enhance their products or services, thereby improving customer satisfaction. To understand customer experiences, preferences, and pain points, we collected and manually annotated 3500 customer feedback from Trustpilot, a leading web platform for sharing user feedback. The data span from 2013 to 2021 and cover five major e-commerce platforms, as shown in Figure 1: Zalando, Wish, Sheinside, Boozt, and Nelly.

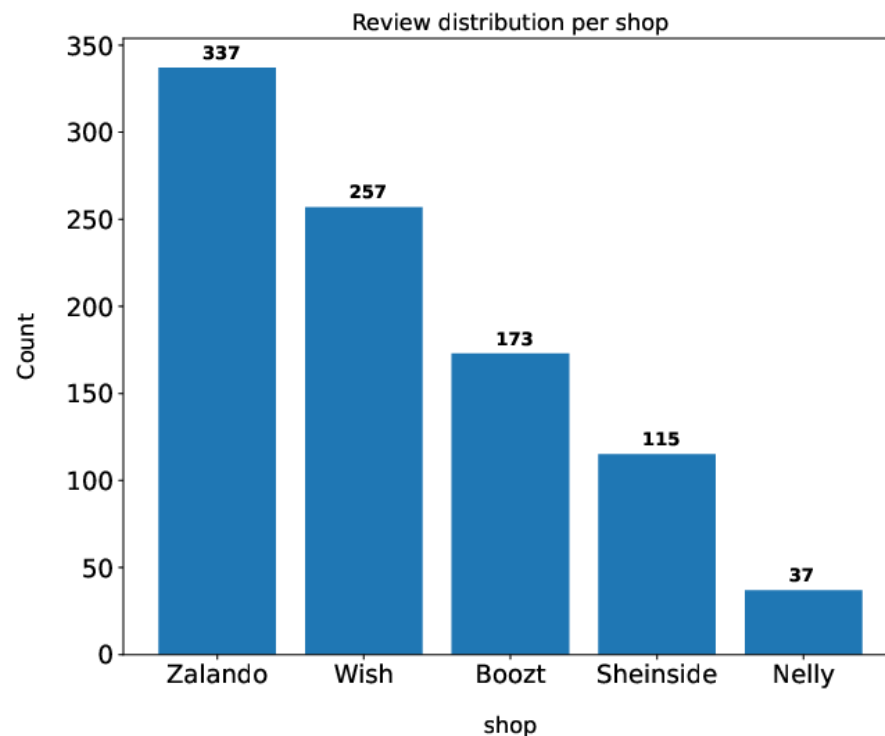


Figure 1. Distribution of reviews per shop.

Prior to data annotation, we performed several steps to detect frequent sets of the aspects mentioned in the text, as our goal was to extract both previously identified aspects from the literature and aspects unique to our corpus. First, we applied part-of-speech tagging to all the reviews. Next, we constructed five bi-term combination types, as suggested by [38], (Noun, Adjective), (Noun, Verb), (Noun, Adverb), (Verb, Adjective), and (Verb, Adverb). Finally, we identified the aspects by counting the most frequent bi-term combinations. Before starting the manual annotation for the entire dataset, we manually annotated 300 reviews to cross-check the quality of the selected aspects.

Finally, we employed two independent annotators (the authors of this paper) to manually annotate our dataset, splitting the data into seven similar parts to ensure reliability. Each annotator worked independently on their assigned parts without access to the other's work until completion. After annotating each part, we held discussions to resolve any disagreements and reach a consensus. If a consensus could not be reached, we removed those data points to maintain the integrity and reliability of the dataset. We annotated the data based on whether the sentiments toward the identified aspects were positive or negative. Additionally, we labeled the overall sentiments of the text as positive, negative, or mixed. As presented by [39], the final aspects used in our study are shown in Table 1. Each aspect–review pair was evaluated and annotated according to the presence/absence of the aspect and the corresponding sentiment value. We assigned one of the three following categories for each aspect–review pair: 'positive', 'negative', or 'not mentioned'. This reflects how the data were planned to be processed and analyzed in the study. As an example, we can consider the following review: *“Wish is great. Finds some very interesting stuff as well as great finds. Just wish delivery did not take so long. Being small business owner I wait on supplies so long and in some occasions did not get them at all. Refunds are hassle-free Wish had a*

great policy just sometimes I do not want a refund I want the products. Other that Wish is great. I am still a customer and fan. Thank you Wish". In this case, the aspects Trust, Refund_process, and Product_availability were assigned the value 'positive', the aspect Shipping was assigned the value 'negative', and all the other aspects were assigned the value 'not mentioned'. We assigned the numerical values 2, 1, and -1 for the three categories.

Table 1. List of aspects categories and definitions.

	Aspect	Definition
1	Shipping	Quality, cost, and timeline of delivery
2	Trust	Customers' overall sentiment about the store
3	Item quality	Quality of the products
4	Customer service	Quality of customers' interactions with the shop's representatives
5	Pricing	Price level, including discounts and campaigns
6	Product features	Quality of product images, descriptions, and size guides provided on the website
7	Refund process	Speed and quality of handling refund issues
8	Return process	Speed, convenience, and cost of the return process
9	App experience	User experience with the store's website
10	Delivered product status	Condition of delivered products, e.g., broken
11	Information	Availability and accuracy of information, including misleading ads
12	Packaging	Attractiveness and quality of packaging
13	Payment	Quality and diversity of the financial transaction process
14	Product availability	Variety of products or brands offered

After performing the manual annotation, for this study, we selected a final dataset of 919 customer feedback entries, where the trust aspect, along with its sentiment, were mentioned in the text. Figure 2 displays the sentiment distribution for various aspects of the final dataset, sorted by the total count of sentiments. The aspects are categorized by their positive (green) and negative (blue) mentions. "Shipping" had the highest number of mentions, with a significant portion being positive. "Item quality" and "Customer service" also had high counts, showing a mix of positive and negative sentiments. Other aspects like "Refund process", "Product availability", and "Product features" had fewer mentions but still showed a notable distribution of sentiments. Aspects like "Packaging" and "Payment" had the fewest mentions, indicating they were less frequently discussed in the reviews. The overall distribution suggests that certain aspects like shipping, item quality, and customer service are critical points of feedback for customers.

To find the correlation between the trust aspect and other aspects, we transformed our annotated dataset by creating new columns for each aspect with "_positive" and "_negative" suffixes. Subsequently, we filled those columns with 1 or 0 based on whether the original column had a positive (2) or negative (1) value. This created new columns for each aspect in the format 'aspect_positive' and 'aspect_negative', marking the occurrence with '1' if the value is positive or negative, respectively, and '0' otherwise.

As presented in Tables 2 and 3 and Figure 3, positive trust is strongly influenced by positive experiences in shipping, item quality, and pricing. Negative trust is most affected by negative experiences in customer service, shipping, and refund processes. This insight can help prioritize improvements in these areas to enhance or protect trust levels.

The quantitative analysis of the dataset reveals several key observations that emerge regarding the relationship between various aspects and the sentiment of trust. Positive experiences with shipping, item quality, and customer service are highly correlated with positive trust sentiments. For instance, efficient and timely shipping (327 occurrences), high-quality items (214 occurrences), and good customer service (104 occurrences) significantly bolster trust. Conversely, negative experiences in these areas substantially diminish trust. Negative shipping experiences (192 occurrences), poor customer service (162 occurrences),

and issues with the refund process (113 occurrences) are strongly associated with negative trust sentiments. Additionally, aspects such as product features, return processes, and product availability also play crucial roles. Positive product features and smooth return processes foster trust, while negative experiences in these areas contribute to distrust. These insights underscore the importance of improving customer service, shipping, item quality, and the refund/return process to enhance overall customer trust and satisfaction.

Table 2. Strongest positive trust indicators.

Aspect	Correlation Value
Shipping_positive	0.58
Item_quality_positive	0.42
Pricing_positive	0.35
Product_availability_positive	0.29
Customer_service_positive	0.27

Table 3. Strongest negative trust indicators.

Aspect	Correlation Value
Customer_service_negative	0.58
Shipping_negative	0.53
Refund_process_negative	0.46
Delivered_product_status_negative	0.35
Item_quality_negative	0.31

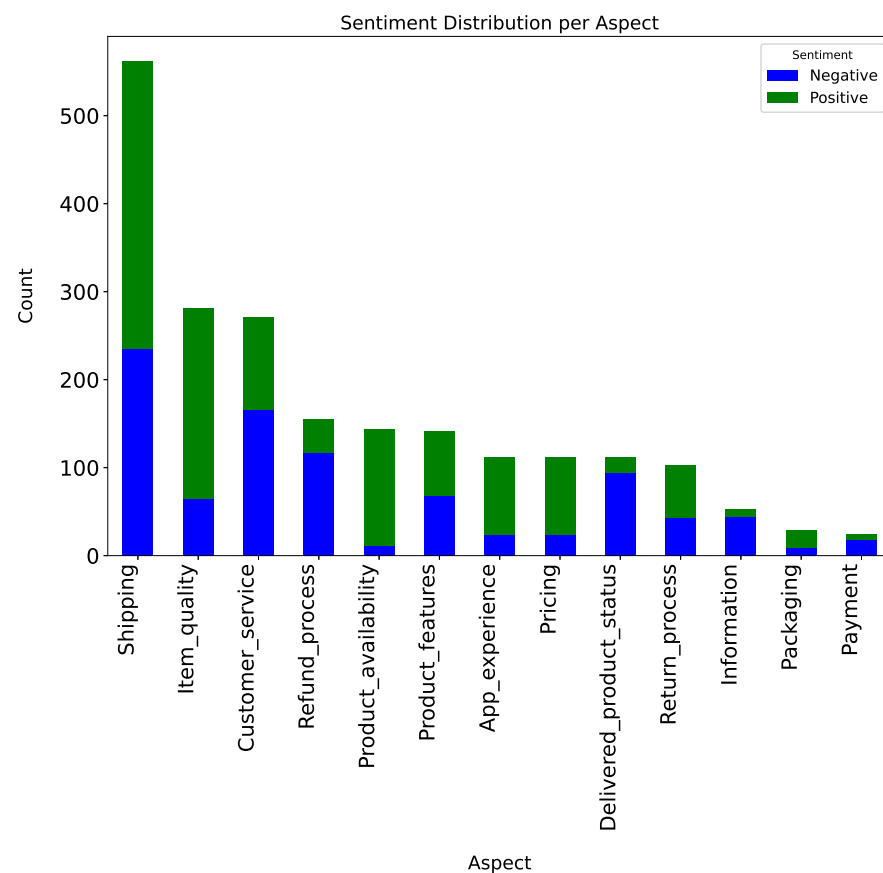


Figure 2. Distribution of sentiments per aspect.

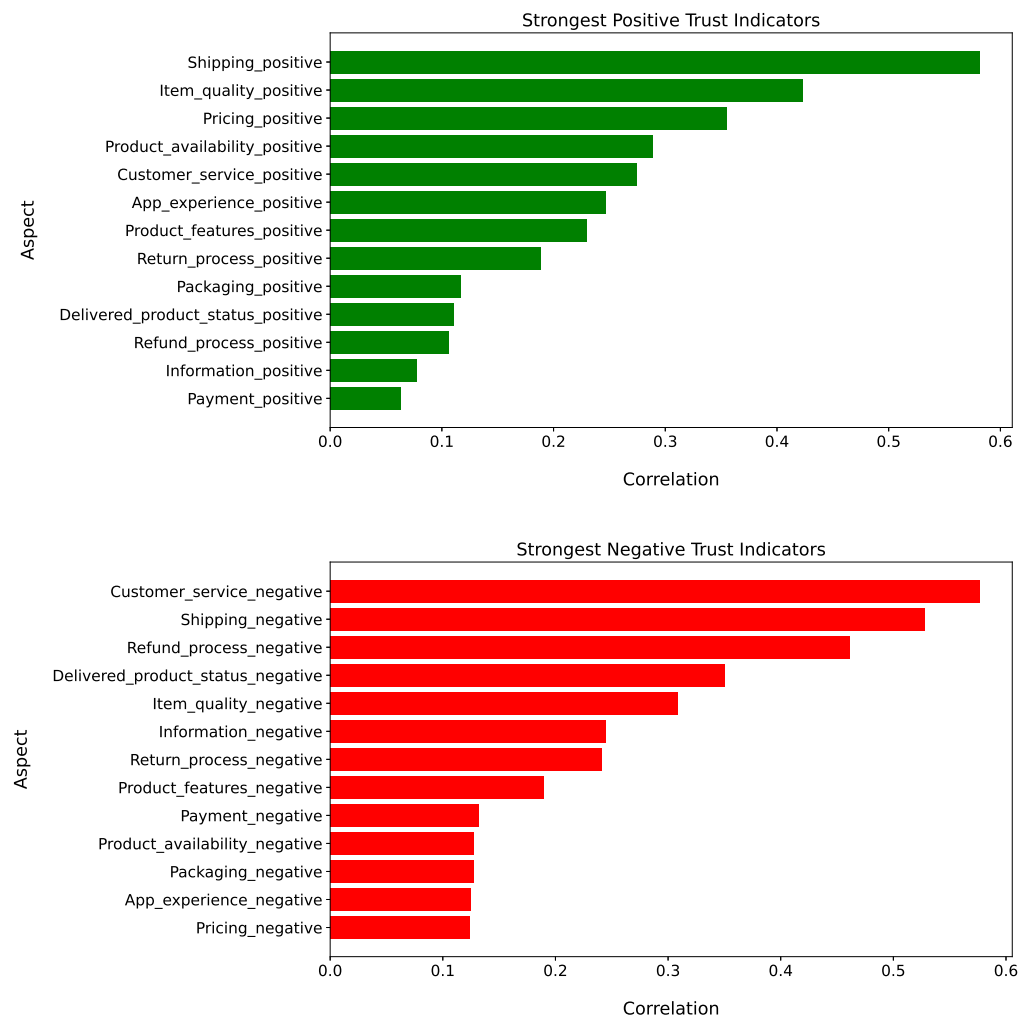


Figure 3. Correlation table.

3.2. Model Building

To test our dataset with LLM models, we divided it into training (819) and validation (100) sets. Given the small size of the dataset, we allocated 100 data points for the validation set and error analysis. We employed two different LLM models for this study: TinyPixel/Llama-2-7B-bf16-sharded (<https://huggingface.co/TinyPixel/Llama-2-7B-bf16-sharded>, accessed on 14 June 2024) and google/gemma-7b (<https://huggingface.co/google/gemma-7b>, accessed on 14 June 2024), using the same training and validation sets for both models. The prompt used for fine-tuning of both models was as follows:

Instruction: Take customer feedback as input and identify key aspects discussed in the text that show customer feelings toward the company and their corresponding sentiments.
 ### Input: More Shein! I have only tried Shein once but very pleased with the products and also the delivery was surprisingly fast and well packaged. Will definitely be ordering again.

Response: Trust: Positive, Shipping: Positive, Packaging: Positive, Item_quality: Positive.

For both models, we utilized Google Colab Pro and T4GPU for model building and testing. Fine-tuning of both models involved several techniques and configurations to optimize performance and efficiency. We utilized several key modules of the Python programming language in the process. One of the most important modules is the transformers library from Hugging Face. This library was instrumental in enabling the fine-tuning and deployment of the Llama model (and other large language models) for sentiment analysis on customer feedback. Specifically, the transformers module allowed us to efficiently manage the model architecture, training, and inference processes, as we will describe later.

Additionally, the Pandas library was heavily used for data processing and analysis, aiding in the organization and manipulation of the dataset used in this research.

First, we employed Parameter-Efficient Fine-Tuning (PEFT) techniques. PEFT allows for efficient model tuning by updating fewer parameters, which is particularly useful when resources are limited. Instead of fine-tuning the entire pre-trained model, PEFT involves updating only a small number of additional or selected parameters. This approach has been shown to achieve a performance comparable to full model fine-tuning while modifying or adding only a tiny fraction of the model's parameters [40–42].

To manage model size and improve inference speed, we applied quantization, specifically low-bit quantization, where the model weights were reduced to 4-bit integers, which significantly reduces the model size while maintaining acceptable performance levels. Ref. [43] demonstrated that quantizing weights to 4-bit integers can achieve substantial memory savings and faster inference times without significant loss in model accuracy. Mixed precision training was another technique used; this method leverages both 32-bit and 16-bit floating-point representations during training, which accelerates the training process and reduces memory usage without substantially affecting model accuracy [44].

Additionally, for capturing essential modifications to the pre-trained weights, we used Low-Rank Adaptation (LoRA). LoRA freezes the pre-trained model weights and injects trainable rank decomposition matrices into each layer of the transformer architecture, significantly reducing the number of trainable parameters for downstream tasks [40]. The parameters for LoRA were set using `lora-r 16` for the rank of adaptation matrices, `lora_alpha 32` for the scaling factor, and `lora_dropout 0.05` to apply dropout, preventing overfitting by randomly setting a portion of the LoRA updates to zero during training.

Several other training configurations were employed to optimize the fine-tuning process. These included setting the learning rate, batch size, and number of epochs. The training also utilized a cosine annealing schedule for adjusting the learning rate, which helps in achieving better convergence. These comprehensive configurations and techniques ensured efficient and effective fine-tuning of the LLMs, enabling accurate predictions of customer trust and associated aspects in e-commerce.

3.3. Configurational Analysis with Qualitative Comparative Analysis

Traditional studies examining the relationship between customer trust and its antecedents in e-commerce often rely heavily on various forms of regression analysis. However, as highlighted by [45], these techniques face three significant limitations by relying on the assumptions of symmetric and linear interactions among variables. To overcome these limitations, researchers have increasingly turned to methodologies that allow for more nuanced configurations of variables. One such approach that has gained prominence is Qualitative Comparative Analysis (QCA) and its various extensions, such as Fuzzy-set QCA (FsQCA) [4]. QCA is widely used for identifying causal patterns and relationships between an outcome of interest, such as customer trust, and its potential antecedents. This method excels in situations where the relationships between variables are complex, and it enables the exploration of multiple causal paths to the same outcome.

In this study, the outcome of interest is trust towards the e-commerce platform, and to complement the described LLMs, we aim to understand what combination of aspects and corresponding sentiments results in positive or negative trust. Similarly to [39], instead of focusing on individual aspects in this part of the analysis, the goal is to understand the impact of the stages of the customer journey experienced with the e-commerce platform. One can specify three distinct stages in this process: (i) selection, (ii) delivery, and (iii) recovery. The first stage, selection, includes information related to the customers' interaction with the platform until making a purchase decision: app experience, information, product availability, product features, pricing, and payment. The second stage, delivery, focuses on the shipping process and the experience with the product, as expressed by the following attributes: shipping, delivered product status, packaging, and item quality. The third stage, recovery, captures any possible interactions with the company

after the product is received, as measured by the aspects of refund process, return process and customer service.

For each stage and review pair, two variables can be defined [39]: (i) one capturing whether there is any positive sentiment mentioned towards the stage, and (ii) another capturing whether there is any negative sentiment mentioned towards the stage. For example, 'Selection Positive' will take on the value 1 if at least one of the six related aspects is mentioned positively in the review, and 0 otherwise. In this study, according to this variable specifications, we utilize the basic, crisp version of QCA, as all the values in the data are either 0 or 1.

As the initial step of the analysis, in the necessity analysis, we assessed whether any single condition was required for the outcome to occur. Following that, we identified combinations of the three stages and their negations that were sufficient to achieve either positive or negative trust. The modeling was conducted using various R packages, and we followed the recommended steps and procedures [45].

The results of the necessity analysis, presented in Tables A1 and A2, reveal several important insights regarding the conditions influencing trust in e-commerce platforms. First, positive evaluation regarding the second stage, delivery, emerges as the most consistent and relevant factor for fostering positive trust. The high consistency and relevance indicate that it is a key driver of customer confidence. Second, positive perception in the first stage, selection, while less consistent than delivery, is still highly relevant for building positive trust, suggesting that factors like app experience and product availability play an important role, though not as universally. Third, positive evaluation of the third stage, recovery, encompassing refund and customer service processes, shows lower consistency, indicating that positive trust can still be achieved, even when these post-purchase processes are not consistently strong. In contrast, for negative trust, negative evaluation of delivery is highly relevant and consistent, highlighting that poor delivery experiences significantly contribute to negative trust perceptions. Finally, negative recovery also shows strong relevance, although we can note that while poor post-purchase processes are crucial to generating negative trust, they are not always present in every case.

The results of the necessity analysis clearly indicate the need for conducting a sufficiency analysis and examining variable configurations rather than focusing solely on individual variables. The different polarities of the three stages exhibit high relevance but low consistency, suggesting they do not function as standalone necessary conditions for trust. This inconsistency highlights the likelihood that these factors interact to influence the outcome. Additionally, trust formation in e-commerce is a complex process, where factors like delivery, decision, and devotion may work together rather than in isolation. For instance, while positive delivery has high consistency, positive trust is likely contingent on favorable interactions with selection-related variables such as app experience and product availability. By considering how different combinations of factors drive positive or negative trust, we can better understand the diverse ways in which customers experience e-commerce platforms. Therefore, sufficiency analysis is crucial to capture these complex causal relationships that are otherwise missed when examining variables individually.

4. Results

In this section, we will present the main results of the modeling process. To provide an understanding of trust, we present the results of LLM models in terms of performance metrics and a qualitative analysis based on several example representative reviews. To further understand the impact of different aspects combined into the three stages of the customer journey, the results of the sufficiency analysis with QCA complement the LLM models.

4.1. Evaluating the Performance of LLMs

We applied five widely used evaluation metrics to quantify the performance of each model: accuracy, precision, recall, F1 and specificity scores. These metrics are essential

for evaluating the performance of machine learning models, particularly in classification tasks [46]. The evaluation metrics are calculated based on the elements of the confusion matrix: (i) true positive (TP), correctly predicting a positive outcome, (ii) true negative (TN), correctly predicting a negative outcome, (iii) false positive (FP), incorrectly predicting a positive outcome, and (iv) false negative (FN), incorrectly predicting a negative outcome. Precision, defined as the ratio of TP predictions to the sum of TP and FN predictions, measures the accuracy of positive predictions and is crucial when the cost of FPs is high. Recall is the ratio of TP predictions to the sum of TP and FN predictions. It assesses the model's ability to identify all relevant instances, and it is vital when missing positive cases is costly. The F1 score, the harmonic mean of precision and recall, provides a balanced measure of a model's performance, especially when the dataset is imbalanced. Accuracy is the proportion of correct predictions (both TPs and TNs) out of all predictions. Specificity measures the proportion of TNs out of all actual negatives. Specificity is useful for understanding how well the model identifies the absence of a certain aspect. Together, these metrics provide a comprehensive evaluation of model performance, addressing different aspects of classification accuracy and reliability.

In this study, we fine-tuned and trained the two LLMs on our manually annotated dataset. After training, we evaluated the performance of the models using a separate validation set, which was also used for error analysis. The validation set consists of manually annotated true responses (aspects present in the review) and the corresponding responses generated by the models (aspects predicted to be in the review). The elements of the confusion matrix are calculated in this case by comparing the set of actual aspects (True_response) and predicted aspects (Predicted_response) for each sample. Specifically, the True_response and Predicted_response columns are first split into sets of aspects for each sample. The union of all possible aspects across the dataset is then obtained to ensure that all relevant aspects are considered in the evaluation. For each aspect, TPs are counted as the instances where the aspect appears in both the actual and predicted sets. FPs are cases where the aspect appears in the predicted set but not in the actual set, indicating an incorrect prediction. FNs occur when the aspect is present in the actual set but missing from the predicted set, representing an aspect the model failed to identify. Finally, TNs are calculated as the instances where the aspect is absent from both the actual and predicted sets, indicating correct identification of an irrelevant aspect. This comparison across all samples allows for calculating precision, recall, F1 score, accuracy, and specificity for each aspect by aggregating these counts.

The results for the two models are presented in Tables 4 and 5. The Gemma model exhibited high performance in several aspects, particularly in predicting positive sentiments for shipping (precision: 0.89, recall: 0.97, F1 score: 0.93) and trust (precision: 0.98, recall: 0.97, F1 score: 0.98). It also performed well in predicting negative sentiments for Customer_service (precision: 0.77, recall: 0.94, F1 score: 0.85).

The Llama model, on the other hand, showed consistently high performance across various aspects. It excelled in predicting positive sentiments for Refund_process (precision: 1.00, recall: 1.00, F1 score: 1.00) and trust (precision: 0.98, recall: 1.00, F1 score: 0.99). Additionally, it performed strongly in negative sentiments for Refund_process (precision: 1.00, recall: 0.92, F1 score: 0.96) and Customer_service (precision: 0.84, recall: 0.89, F1 score: 0.86).

As presented in Table 6, the Llama model performed better overall. The F1 score in this table is calculated as the weighted average of F1 scores for individual aspects. It demonstrated higher precision, recall, and F1 scores in critical aspects such as Refund_process and trust and maintained more consistent performance across various metrics. The Gemma model, while strong in some areas, showed variability and struggled with certain aspects. Thus, the Llama model is considered the better performer for this task.

Table 4. Gemma model performance for various aspects with precision, recall, F1 score, accuracy, and specificity.

Index	Accuracy	Precision	Recall	F1 Score	Specificity
Return_process: Positive	0.97	0.75	0.60	0.67	0.99
Return_process: Negative	0.98	1.00	0.33	0.50	1.00
Pricing: Positive	0.95	0.84	0.89	0.86	0.96
Pricing: Negative	0.93	0.14	0.50	0.22	0.94
Refund_process: Positive	0.99	1.00	0.75	0.86	1.00
Refund_process: Negative	0.97	0.92	0.85	0.88	0.99
Shipping: Positive	0.95	0.89	0.97	0.93	0.94
Shipping: Negative	0.93	0.86	0.94	0.90	0.93
Product_features: Positive	0.92	0.63	0.50	0.56	0.97
Product_features: Negative	0.97	0.50	0.67	0.57	0.98
Product_availability: Positive	0.98	0.92	0.92	0.92	0.99
Item_quality: Positive	0.94	0.83	0.91	0.87	0.95
Item_quality: Negative	0.95	0.67	0.57	0.62	0.98
Customer_service: Positive	0.98	0.88	0.88	0.88	0.99
Customer_service: Negative	0.94	0.77	0.94	0.85	0.94
Trust: Positive	0.97	0.98	0.97	0.98	0.97
Trust: Negative	0.97	0.95	0.97	0.96	0.97

Table 5. Llama Model performance for various aspects with precision, recall, F1 score, accuracy, and specificity.

Index	Accuracy	Precision	Recall	F1 Score	Specificity
Return_process: Positive	0.99	0.83	1.00	0.91	0.99
Return_process: Negative	0.97	0.50	0.33	0.40	0.99
Pricing: Positive	0.96	0.85	0.94	0.89	0.96
Pricing: Negative	0.94	0.00	0.00	0.00	0.96
Refund_process: Positive	1.00	1.00	1.00	1.00	1.00
Refund_process: Negative	0.99	1.00	0.92	0.96	1.00
Shipping: Positive	0.94	0.97	0.85	0.91	0.98
Shipping: Negative	0.93	0.93	0.84	0.89	0.97
Product_features: Positive	0.94	0.75	0.60	0.67	0.98
Product_features: Negative	0.99	0.75	1.00	0.86	0.99
Product_availability: Positive	0.96	0.80	0.92	0.86	0.97
Product_availability: Negative	0.99	0.67	1.00	0.80	0.99
Item_quality: Positive	0.91	0.76	0.86	0.81	0.92
Item_quality: Negative	0.97	0.83	0.71	0.77	0.99
Customer_service: Positive	0.98	0.80	1.00	0.89	0.98
Customer_service: Negative	0.95	0.84	0.89	0.86	0.96
Trust: Positive	0.99	0.98	1.00	0.99	0.97
Trust: Negative	0.99	1.00	0.97	0.99	1.00

Table 6. Average F1 score of the Gemma and Llama models.

Model	Average F1 Score
Llama	0.88
Gemma	0.80

In the presented results tables, we include precision, recall, and F1 scores for various aspects alongside the overall trust score. The reason for showcasing these additional aspects is to provide a comprehensive understanding of how specific components influence the final trust sentiment. By analyzing the precision, recall, and F1 scores of individual aspects, such as return process, pricing, and shipping, we can identify which factors contribute to the positive or negative perception of trust in e-commerce transactions. This detailed breakdown allows us to pinpoint the key elements that drive customer trust and

understand the interplay between different aspects and the overall trust score. Hence, presenting these scores not only validates the reliability of the trust prediction but also elucidates the underlying reasons for the trust sentiment, offering actionable insights for improving e-commerce platforms.

Table 7 presents a few examples of the completions generated by the best model.

Table 7. Predictions generated by the Llama model.

Review Number	Customer Feedback
1	I did not love the dress, but great customer service, I will recommend them.
2	I hate them, where is my order
3	very hard to navigate. I will delete the app
4	good packaging but fake products. Stay away
5	free and fast delivery. I like these guys
6	I love the dress, great customer service, I will recommend them.
Review Number	Model's Prediction
1	Trust: Positive, Customer_service: Positive, Item_quality: Negative
2	Trust: Negative, Shipping: Negative
3	Trust: Negative, App_experience: Negative
4	Trust: Negative, Packaging: Positive, Item_quality: Negative
5	Trust: Positive, Shipping: Positive
6	Trust: Positive, Customer_service: Positive, Item_quality: Positive

4.2. Configurations from QCA Analysis

In the QCA analysis of understanding the antecedents of positive trust, three distinct configurations were identified (Table 8). In the table, ● and ○ indicate the presence and absence of the condition, respectively.

Solution 1 highlights the importance of positive perception of delivery as a sufficient condition to generate trust. Similarly, according to solution 2, positive evaluation of the selection stage can play the same role. Both of these solutions exhibit a high consistency of over 0.95, with coverage values of 0.73 and 0.28, respectively. Solution 1 is the most dominant configuration, and the unique coverage of 0.22 suggests that this solution alone explains 22% of the cases where positive trust is observed, indicating that a favorable delivery process is a critical driver of trust in e-commerce. In contrast, solution 3 shows that recovery alone is not a sufficient condition to achieve trust unless it is accompanied by a positive evaluation of the delivery stage. Overall, these findings reveal multiple pathways to achieving positive trust in e-commerce underscoring the complexity of the phenomenon. Furthermore, delivery plays a prominent role in two of the solutions, highlighting the role of this stage of the customer journey in contributing to trust formation.

Table 8. Configurations for positive trust.

Conditions	Solution 1	Solution 2	Solution 3
Selection: Positive		●	
Selection: Negative			
Delivery: Positive	●		
Delivery: Negative			○
Recovery: Positive			●
Recovery: Negative			
Consistency	0.995	0.952	0.988
Coverage	0.732	0.279	0.570
Unique coverage	0.223	0.067	0.112

The QCA analysis for conditions leading to negative trust identified two configurations that are sufficient to explain the outcome (Table 9).

Solution 1 highlights the importance of the recovery stage, as negative experiences with any post-purchase processes such as returns, refunds, and customer service are sufficient to foster negative trust. The configuration shows a very high consistency of 0.95, and a high coverage of 0.65. Solution 2 demonstrates that the combination of the absence of positive perception of selection and recovery, combined with negative evaluation of delivery, is another pathway leading to negative trust. In addition to high consistency, the unique coverage of 0.27 shows the high number of cases that are explained exclusively by this configuration. This shows that the combination of a poor delivery experience without a positive compensation in other stages of the customer journey process can play a significant role in damaging trust.

Table 9. Configurations for negative trust.

Conditions	Solution 1	Solution 2
Selection: Positive		○
Selection: Negative		
Delivery: Positive		
Delivery: Negative		●
Recovery: Positive		○
Recovery: Negative	●	
Consistency	0.950	0.921
Coverage	0.648	0.739
Unique coverage	0.178	0.269

5. Discussion

In this article, we explored how a combination of neural computation through LLMs and qualitative comparative analysis could be used to predict and explain customer trust in e-commerce. LLMs such as TinyPixel/Llama-2-7B-bf16-sharded and Gemma were fine-tuned to analyze customer reviews from leading e-commerce platforms, accurately predicting sentiments related to key aspects such as shipping, customer service, and product quality. The predictive performance of the LLMs was strong, particularly in terms of detecting trust signals within textual feedback, showcasing the potential of neural computation in processing and interpreting large-scale, unstructured data.

However, while LLMs performed well in predictive tasks, they do not inherently offer an understanding of the causal relationships between various aspects of the customer journey and the resulting trust outcomes. To address this, we complemented it with QCA, which allowed us to uncover configurations of factors across three key stages of the customer journey: selection, delivery, and recovery. Through QCA, we were able to capture not only which individual factors mattered but also how combinations of these factors influenced either positive or negative trust.

5.1. Examples of Misclassifications by LLMs

An interesting task is to characterize, if possible, situations when LLMs typically make mistakes in predicting the presence and sentiment of aspects. For the best-performing model, we observed that errors often occur when the model predicts additional categories that are not present in the true response. For example, there are cases where the true response includes “Trust: Positive” and “Shipping: Positive”, but the predicted response adds an unrelated category like “App_experience: Positive”.

TinyPixel could identify aspects that were mistakenly missed during manual annotation. For example, in the following instance, the aspect of Product_availability was missing in the manual annotation but was correctly identified and marked as negative by the model:

“Long wait times with no communication or certainty. I ordered on the 26th April it is now 1st May and my order still shows as processing awaiting picking. In reality this is only 5 days but

reading other reviews I am worried that I will not be receiving my items. I spent nearly 250 on this order and they have already taken this money out of my account. Terrible communication and I am now worried that many of my items will not be available. First and last time using SHEIN."

Moreover, some of the misclassifications by the TinyPixel model involved the introduction of new aspects. For example, the pricing aspect was wrongly identified as Sale_pricing. However, the associated sentiment was correctly identified as positive by the model.

"I am a regular customer and I absolutely love Zalando love the sales on brands delivery is quick. Absolute A!"

Another mistake is related to having infrequent aspects in the training sample, such as information. For example, in the following instance, the model identified the information aspect instead of Product_features. This kind of misclassification error can be addressed by introducing more training data.

"It's great. Wish is good. You just have to read full descriptions so you do not get disappointed."

Finally, we observed that this approach was successful, as the power of LLM models can turn raw text into actionable business insights. It helps businesses identify customer pain points and prevent potential issues. These experiences show that any customer feedback can be transformed into labeled data with sentiments, reducing the manual annotation workload.

5.2. Contributions to Academia and Practice

The main academic contribution of this study lies in utilizing LLMs and QCA to enhance the understanding of customer trust in e-commerce. Traditional research on trust in e-commerce has often relied on regression models, which are limited in capturing the complexity and non-linear interactions among different stages of the customer journey. By incorporating LLMs, we introduce a powerful machine learning approach that can process large volumes of unstructured data, specifically customer reviews, to accurately predict trust-related outcomes. This represents a methodological advancement by utilizing neural computation to model and predict trust sentiment from natural language, offering richer insights into customer behavior than conventional structured data methods like surveys.

A key academic contribution of this study is the synergy between predictive performance (LLMs) and explanatory analysis (QCA). While LLMs excel in predicting trust based on textual data, QCA provides a tool to uncover the configurations of factors—such as selection, delivery, and recovery stages—that combine to influence trust. This combination of methods bridges the gap between predictive analytics and causal explanation, enabling a more comprehensive understanding of the factors that drive trust in e-commerce.

This study provides valuable practical insights for e-commerce businesses, particularly in how they can leverage advanced tools to predict and understand customer trust. First, the ability of LLMs to process large volumes of unstructured text data, such as customer reviews, offers businesses a highly scalable solution for understanding customer sentiment in real time. This provides a real-time pulse on customer trust levels and enables businesses to respond proactively to emerging trust issues before they become widespread.

In addition to predictive insights from LLMs, this study highlights the importance of configurations of customer experience elements across the selection, delivery, and recovery stages. For businesses, this means that improving trust requires a holistic view of the customer journey, rather than focusing on any single touchpoint. For example, this research shows that while an efficient delivery process is critical to fostering trust, businesses must also pay attention to pre-purchase decisions (selection) and post-purchase support (recovery). This means that a seamless and cohesive experience across these stages is key to building long-term customer trust. Therefore, businesses should aim to optimize all touchpoints, ensuring that customers have positive experiences at every stage of their journey.

6. Conclusions

In this study, we explored the application of two advanced LLMs for predicting trust in e-commerce. In the research, we employed and fine-tuned two LLMs, TinyPixel/Llama-2-7B-bf16-sharded and Gemma, to assist in processing online customer reviews. The fine-tuning of these LLMs enabled us to automate the aspect extraction and sentiment classification processes, which are integral parts of analysis. By leveraging LLMs, we were able to identify trust indicators in customer feedback, reducing the manual analysis. This process involved training the models on annotated data and using them to predict key aspects such as shipping, product quality, and customer service, along with their corresponding sentiments. For training both models, we used the Google Colab Pro environment, and training each model took approximately 2 h. Our analysis revealed that these models significantly enhanced the accuracy of trust prediction by effectively analyzing customer feedback and sentiments. To the best of our knowledge, no prior research has employed LLMs specifically for this purpose, making our approach and dataset unique. Furthermore, making use of the predictions produced by LLMs, we illustrated how to use a configurational approach to shed lights of the important stages of the customer journey process having a high impact on the formation of trust.

The combination of LLMs and QCA provided a more comprehensive framework for understanding trust in e-commerce. LLMs excelled at accurately predicting trust levels and associated sentiments from customer reviews, automating a task that would otherwise be labor-intensive and error-prone. However, QCA added an explanatory layer by revealing how different customer journey stages interacted to shape overall trust. This synergy between predictive performance and explanatory power is one of the core contributions of this study. It demonstrates that machine learning models can provide accurate predictions, but configurational methods like fsQCA are needed to explain the interactions that lead to those predictions.

For future research, we recommend training other machine learning models such as RoBERTa or Llama, or combining different models to augment training data. Performing similar experiments on larger datasets could confirm the scalability and effectiveness of the proposed approach. As manual data annotation is a very labor-intensive task, additional synthetic annotated data could be utilized to address the scarcity of manually annotated data and to improve the model's performance on infrequent aspects. By employing fine-tuned LLMs for generating high-quality synthetic data, we can efficiently expand a limited training dataset, focusing on underrepresented aspects such as payment, refund, and product availability. In a future paper, the different NLP analysis steps could be combined: a fine-tuned BERT model for aspect detection with a fine-tuned LLM-based ABSA model to handle cases where aspects are not explicitly mentioned as nouns, and Llama for sentiment analysis.

Additionally, integrating real-time user feedback to continuously refine and adapt the models could ensure sustained accuracy and relevance in a dynamic market environment. Furthermore, in future research, various interpretability techniques can be incorporated in the model-building and analysis process. These could include simple approaches, such as feature importance and attribution methods, and also more advanced techniques, such as surrogate models (LIME and SHAP). This addition can help to explain the decision-making process of our large language models (LLMs). These techniques will allow decision makers to better understand the contribution of specific aspects—such as customer service, shipping, and product quality—toward predicting trust in a brand or product. By integrating both predictive performance and interpretability, one could provide a comprehensive analysis of trust, enhancing transparency at both the global and individual prediction levels.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app142110069/s1>.

Author Contributions: Conceptualization, L.D. and J.M.; methodology, L.D. and J.M.; formal analysis, L.D.; data curation, L.D. and J.M.; writing—original draft preparation, L.D. and J.M.; writing—review and editing, L.D. and J.M.; visualization, L.D.; supervision, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Foundation For Economic Education (Liikesivistysrahasto), Jenny ja Antti Wihurin rahasto and Marcus Wallenbergin Liiketaloudellinen Tutkimussäätiö.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to fine-tune the LLMs and to perform QCA is available as Supplementary Material.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Result of QCA Necessary Condition Analysis

Table A1. Consistency and coverage values for necessary conditions for positive trust (inclN = inclusion, RoN = relevance of necessity, covN = coverage).

Condition	inclN	RoN	covN
Not selection positive	0.396	0.521	0.405
Selection positive	0.604	0.970	0.953
Not delivery positive	0.268	0.547	0.306
Delivery positive	0.732	0.996	0.995
Not recovery positive	0.721	0.329	0.547
Recovery positive	0.279	0.989	0.952
Not selection negative	0.940	0.405	0.702
Selection negative	0.060	0.863	0.219
Not delivery negative	0.900	0.820	0.875
Delivery negative	0.100	0.680	0.171
Not recovery negative	0.979	0.659	0.819
Recovery negative	0.021	0.751	0.050

Table A2. Consistency and coverage values for necessary conditions for negative trust (inclN = inclusion, RoN = relevance of necessity, covN = coverage).

Condition	inclN	RoN	covN
Not selection positive	0.951	0.615	0.595
Selection positive	0.049	0.619	0.047
Not delivery positive	0.994	0.733	0.694
Delivery positive	0.006	0.545	0.005
Not recovery positive	0.977	0.289	0.453
Recovery positive	0.023	0.825	0.048
Not selection negative	0.653	0.224	0.298
Selection negative	0.347	0.957	0.781
Not delivery negative	0.209	0.394	0.125
Delivery negative	0.791	0.911	0.829
Not recovery negative	0.352	0.299	0.181
Recovery negative	0.648	0.983	0.950

References

1. Alrayani, B.; Kalkatawi, M.; Abulkhair, M.; Abukhodair, F. From Customer's Voice to Decision-Maker Insights: Textual Analysis Framework for Arabic Reviews of Saudi Arabia's Super App. *Appl. Sci.* **2024**, *14*, 6952. [\[CrossRef\]](#)
2. Wang, Z.; Zhu, Y.; He, S.; Yan, H.; Zhu, Z. Llm for sentiment analysis in e-commerce: A deep dive into customer feedback. *Appl. Sci. Eng. J. Adv. Res.* **2024**, *3*, 8–13.
3. Samant, R.M.; Bachute, M.R.; Gite, S.; Kotecha, K. Framework for deep learning-based language models using multi-task learning in natural language understanding: A systematic literature review and future directions. *IEEE Access* **2022**, *10*, 17078–17097. [\[CrossRef\]](#)

4. Ragin, C.C. *Fuzzy-Set Social Science*; University of Chicago Press: Chicago, IL, USA, 2000.
5. Lee, C.K.H.; Tse, Y.K.; Leung, E.K.H.; Wang, Y. Causal recipes of customer loyalty in a sharing economy: Integrating social media analytics and fsQCA. *J. Bus. Res.* **2024**, *181*, 114747. [\[CrossRef\]](#)
6. Gupta, A. E-Commerce: Role of E-Commerce in today's business. *Int. J. Comput. Corp. Res.* **2014**, *4*, 1–8.
7. Hadjiphanis, L.; Christou, L. The role of ecommerce on consumer behaviour. *J. Bus. Adm.* **2006**, *5*, 4.
8. Noori, A. Evaluating the impact of the determinants of e-commerce customer trust and satisfaction. *Int. J. Bus. Manag. Res. (IJBMR)* **2019**, *9*, 21–38. [\[CrossRef\]](#)
9. Nguyen, H.T. Factors Affecting Customer Satisfaction and Trust in an E-commerce setting: A Case Study of Muachung. vn in Vietnam. *AU-GSB E-J.* **2014**, *7*, 43–52.
10. McKnight, D.H.; Choudhury, V.; Kacmar, C. Developing and validating trust measures for e-commerce: An integrative typology. *Inf. Syst. Res.* **2002**, *13*, 334–359. [\[CrossRef\]](#)
11. Chen, S.C.; Dhillon, G.S. Interpreting dimensions of consumer trust in e-commerce. *Inf. Technol. Manag.* **2003**, *4*, 303–318. [\[CrossRef\]](#)
12. Elida, T.; Rahardjo, W.; Raharjo, A.; Sukirman, E. Online shopping: What factors determine consumers to buy. *Management* **2019**, *7*, 238–246.
13. Turban, E.; King, D.; Lee, J.; Viehland, D. *Electronic Commerce: A Managerial Perspective*; Prentice Hall: Saddle River, NJ, USA, 2006.
14. Van Riel, A.C.; Semeijn, J.; Pauwels, P. Online travel service quality: The role of pre-transaction services. *Total Qual. Manag. Bus. Excell.* **2004**, *15*, 475–493. [\[CrossRef\]](#)
15. Reichheld, F.F.; Sasser, W.E. *Zero Defections: Quality Comes to Services*; Harvard Business Publishing: Brighton, MA, USA, 1990; Volume 68, pp. 105–111.
16. Huang, Z.; Luo, Y.; Wang, D. Online customer service quality of online shopping: Evidence from Dangdang. com. *Clust. Comput.* **2019**, *22*, 15285–15293. [\[CrossRef\]](#)
17. Kleinman, S. *Online Shopping Customer Experience Study*; Commissioned by UPS, ComScore, Inc.: Reston, VA, USA, 2012.
18. Martínez-López, F.J.; Feng, C.; Li, Y.; López-López, D. Using instant refunds to improve online return experiences. *J. Retail. Consum. Serv.* **2022**, *68*, 103067. [\[CrossRef\]](#)
19. Kabir, M.A.; Saidin, S.Z.; Ahmi, A. Adoption of e-payment systems: A review of literature. In Proceedings of the International Conference on E-Commerce, Kuching, Sarawak, Malaysia, 20–22 October 2015; pp. 112–120.
20. Premchand, A.; Choudhry, A. Future of payments—ePayments. *Int. J. Emerg. Technol. Adv. Eng.* **2015**, *5*, 110–115.
21. Alzoubi, H.; Alshurideh, M.; Kurdi, B.A.; Alhyasat, K.; Ghazal, T. The effect of e-payment and online shopping on sales growth: Evidence from banking industry. *Int. J. Data Netw. Sci.* **2022**, *6*, 1369–1380. [\[CrossRef\]](#)
22. Collier, J.E.; Bienstock, C.C. *How Do Customers Judge Quality in an E-Tailer?* MIT Sloan Management Review: Cambridge, MA, USA, 2006.
23. Ma, S. Fast or free shipping options in online and Omni-channel retail? The mediating role of uncertainty on satisfaction and purchase intentions. *Int. J. Logist. Manag.* **2017**, *28*, 1099–1122. [\[CrossRef\]](#)
24. Mohd Sam, M.F.; Tahir, M.N.H. Website quality and consumer online purchase intention of air ticket. *Int. J. Basic Appl. Sci.* **2009**, *9*, 20–25.
25. Akbar, A.R.; Kalis, M.C.I.; Afifah, N.; Purmono, B.B.; Yakin, I. The Influence of Product Packaging Design and Online Customer Review on Brand Awareness and Their Impact on Online Purchase Intention. *South Asian Res. J. Bus. Manag.* **2023**, *5*, 10–18. [\[CrossRef\]](#)
26. Murni, Y.; No, J.M.S.; Kembangan, W.J. Analysis of The Influence of Trust, Easiness in Shopping, Product Quality to Customer Satisfaction to Increasiness in Shopping Consumer Loyalty Online Shopping. *Int. J. Econ. Bus. Manag. Res.* **2018**, *2*, 50–64.
27. Mokobombang, M.G.; Kusumawati, N. The Impact of Product Description, Product Photo, Rating, and Review on Purchase Intention in E-commerce. *J. Consum. Stud. Appl. Mark.* **2023**, *1*, 137–147. [\[CrossRef\]](#)
28. Policarpo, L.M.; da Silveira, D.E.; da Rosa Righi, R.; Stoffel, R.A.; da Costa, C.A.; Barbosa, J.L.V.; Scorsatto, R.; Arcot, T. Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. *Comput. Sci. Rev.* **2021**, *41*, 100414. [\[CrossRef\]](#)
29. Zhan, T.; Shi, C.; Shi, Y.; Li, H.; Lin, Y. Optimization Techniques for Sentiment Analysis Based on LLM (GPT-3). *arXiv* **2024**, arXiv:2405.09770. [\[CrossRef\]](#)
30. Soviero, B.; Kuhn, D.; Salle, A.; Moreira, V.P. ChatGPT Goes Shopping: LLMs Can Predict Relevance in eCommerce Search. In Proceedings of the European Conference on Information Retrieval, Glasgow, UK, 24–28 March 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 3–11.
31. Ethape, P.; Kane, R.; Gadekar, G.; Chimane, S. Smart Automation Using LLM. *Int. Res. J. Innov. Eng. Technol.* **2023**, *7*, 603.
32. Zhuang, Y.; Yu, Y.; Wang, K.; Sun, H.; Zhang, C. Toolqa: A dataset for llm question answering with external tools. *Adv. Neural Inf. Process. Syst.* **2024**, *36*, 50117–50143.
33. Abbasiantaeb, Z.; Yuan, Y.; Kanoulas, E.; Aliannejadi, M. Let the llms talk: Simulating human-to-human conversational qa via zero-shot llm-to-llm interactions. In Proceedings of the 17th ACM International Conference on Web Search and Data Mining, Merida, Mexico, 4–8 March 2024; pp. 8–17.
34. Miah, M.S.U.; Kabir, M.M.; Sarwar, T.B.; Safran, M.; Alfarhood, S.; Mridha, M. A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. *Sci. Rep.* **2024**, *14*, 9603. [\[CrossRef\]](#)

35. Zhao, Z.; Fan, W.; Li, J.; Liu, Y.; Mei, X.; Wang, Y.; Wen, Z.; Wang, F.; Zhao, X.; Tang, J.; et al. Recommender systems in the era of large language models (llms). *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 6889–6907. [[CrossRef](#)]
36. Xu, X.; Xu, Z.; Ling, Z.; Jin, Z.; Du, S. Emerging Synergies Between Large Language Models and Machine Learning in Ecommerce Recommendations. *arXiv* **2024**, arXiv:2403.02760.
37. Kathiriya, S.; Mullapudi, M.; Karangara, R. Optimizing ECommerce Listing: LLM Based Description and Keyword Generation from Multimodal Data. *Int. J. Sci. Res. (IJSR)* **2023**, *12*, 2123–2130.
38. Im, J.; Song, T.; Lee, Y.; Kim, J. Confirmatory aspect-based opinion mining processes. *arXiv* **2019**, arXiv:1907.12850. [[CrossRef](#)]
39. Mezei, J.; Davoodi, L.; Nikou, S. Customer Review Analysis of Online E-commerce Platforms—A Configurational Approach. In Proceedings of the 57th Hawaii International Conference on System Sciences, Hilton Hawaiian Village Waikiki Beach Resort, Honolulu, HI, USA, 3–6 January 2024; Bui, T.X., Ed.; ScholarSpace: Honolulu, HI, USA 2024; pp. 1476–1485.
40. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. Lora: Low-rank adaptation of large language models. *arXiv* **2021**, arXiv:2106.09685.
41. Lester, B.; Al-Rfou, R.; Constant, N. The power of scale for parameter-efficient prompt tuning. *arXiv* **2021**, arXiv:2104.08691.
42. Liu, H.; Tam, D.; Muqeeth, M.; Mohta, J.; Huang, T.; Bansal, M.; Raffel, C.A. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 1950–1965.
43. Jacob, B.; Kligys, S.; Chen, B.; Zhu, M.; Tang, M.; Howard, A.; Adam, H.; Kalenichenko, D. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2704–2713.
44. Micikevicius, P.; Narang, S.; Alben, J.; Diamos, G.; Elsen, E.; Garcia, D.; Ginsburg, B.; Houston, M.; Kuchaiev, O.; Venkatesh, G.; et al. Mixed precision training. *arXiv* **2017**, arXiv:1710.03740.
45. Pappas, I.O.; Woodside, A.G. Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for research practice in Information Systems and marketing. *Int. J. Inf. Manag.* **2021**, *58*, 102310. [[CrossRef](#)]
46. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.