World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Integrating DevOps and Large Language Model Operations (LLMOps) for GenAI-Enabled E-commerce Innovations A Pathway to Intelligent Automation

Rahul Kalva *

Dublin, CA, USA – 94568

World Journal of Advanced Research and Reviews, 2024, 24(03), 879–889

Publication history: Received on 29 October 2024; revised on 09 December 2024; accepted on 11 December 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.24.3.3725

Abstract

This study explores the integration of DevOps methodologies with Large Language Model Operations (LLMOps) to drive intelligent automation and innovation in e-commerce. It is our goal to forecast and examine consumer faith in online marketplaces by making use of GenAI and the neural computation powers of LLMs. The strategy's core goal is to improve confidence by means of a two-pronged approach: first, predictive modeling with LLMs; and second, causal analysis with QCA. In order to determine the degree of trust based on reviews, LLMs examine fundamental aspects of online shopping, such as product quality, customer service, refund policies, and delivery speed. At the same time, customer journey stages including product selection, delivery, and post-purchase assistance are uncovered by QCA as causal linkages between trust. The integration of LLMOps within a DevOps framework ensures efficient deployment and maintenance of AI models, fostering seamless innovation and operational agility. This hybrid approach offers e-commerce platforms actionable insights into enhancing customer trust while setting a benchmark for intelligent automation in the industry.

Keywords: DevOps; Large Language Model Operations (LLMOps); GenAI; E-commerce; Intelligent Automation

1. Introduction

The acronym LLMOps is an abbreviation that stands for "large language model operations." This term refers to the specific methods and processes that speed up the creation, release, and upkeep of AI models during their whole lifespan.

There is the potential for LLMOps systems to provide more effective library administration, hence reducing operational expenses and allowing for a smaller number of technical workers to carry out activities. These processes include deployment as well as data preparation, training of the language model, monitoring, fine-tuning, and deployment. Similar to Machine Learning Ops (MLOps), LLMOps is built upon a partnership of data scientists, DevOps engineers, and IT professionals.

Modern natural language processing (NLP) models have improved to the point where they can quickly summarize information, answer questions expressed in normal language, and execute complicated instructions [1]. The GPT-4-using ChatGPT from OpenAI and the BERT from Google are two examples of LLMs.

With the help of an LLMOps platform, data scientists and software engineers may work together to improve data exploration, experiment tracking in real-time, rapid engineering, and model and pipeline management. When it comes to the machine learning lifecycle, LLMOps is responsible for automating the operational and monitoring duties.

1.1. LLMOps vs MLOps

A possible outcome of LLMOps' inclusion in the machine learning operations category is that it would be ignored or perhaps called "MLOps for LLMs." Although LLMOps is mainly designed to facilitate the process of LLM development, it

* Corresponding author: Rahul Kalva

should nevertheless be evaluated separately [2]. The following are two distinct ways in which the workflows and requirements of machine learning (ML) are altered by LLMs.

Cost savings with hyperparameter tuning: Improving accuracy or other metrics is often the goal of hyperparameter tuning in ML. To reduce the computing power and cost of training and inference, tuning becomes critical for LLMs. Varying the batch sizes allows one to accomplish this. Higher performance at lower cost is possible with LLMs since they may be fine-tuned with new data for domain-specific improvements after starting with a foundation model.

Performance metrics: Metrics for ML model performance, including as accuracy, area under the curve (AUC), and F1 score, are often well-defined and straightforward to compute [3]. The bilingual evaluation understudy (BLEU) and the recall-oriented understudy for gisting evaluation (ROUGE) are two examples of the alternative standard standards and scoring that are required for LLM evaluations. These things need to be thought about more when they're put into action.

Additionally, LLMOps platforms can offer what are commonly thought of as traditional MLOps functionalities:

- Data management
- Deployment process
- Model testing and training
- Monitoring and observability
- Security and compliance support
- Use cases

Many different kinds of tasks can benefit from LLMOps' increased efficiency, such as:

Building vector databases in order to get data that is pertinent to the current context.

Continuous integration and delivery (CI/CD), which is where continuous integration and delivery pipelines simplify testing and deployment while automating model development. The efficient and effective management of these pipelines is made possible with the use of tools such as Jenkins, GitLab CI/CD, and GitHub Actions. As a result, users will experience minimal downtime during model upgrades and rollbacks. Incorporating comprehensive testing techniques and versioning models can assist discover errors early and ensure that only well-performing models are deployed [4].

- Data collection, getting things rolling and quickly implementing engineering, using a wide range of sources, domains, and languages.
- Data labeling while supplementing it with human annotation to deliver nuanced, domain-specific evaluation.
- Data storage, data storage, retrieval, and manipulation across the LLM lifespan; appropriate database and storage systems; and organizing and versioning.
- Exploratory data analysis (EDA) while working on the machine learning model lifecycle, it is helpful to have access to editable and shareable data sets, tables, and visualizations that allow you to examine, prepare, and share particular data.
- Model fine-tuning, in order to optimize models for certain tasks or domains.

Model inference The production-specific testing and QA aspects, including the inference request timings and frequency of model refresh, can be managed by serving. Use graphics processing unit acceleration to power your REST API model endpoints.

Model review and governance in order to keep tabs on the many versions of models and pipelines and oversee their whole lifespan. With an open-source MLOps platform like MLflow, this can make it possible for ML models to work together.

Model monitoring, adding human review to your LLM programs. Recognize possible hostile assaults, detect model drift, and locate possible improvement areas.

- Prompt analytics, logging and testing.
- Prompt engineering, employing resources that permit in-context learning instead of refining with private data.
- Prompt execution to enable model optimization.

Text creation and outputs (link resides ouside ibm.com) in many different contexts.Operating infrastructure scripting and automation are both within LLMs' code generation capabilities. In addition to translating languages, they write text for purposes such as documenting processes or coding.

1.2. Benefits of LLMOps

By solving critical problems and paving the way for innovation, LLMOps brings GenAI significant benefits:

1.2.1. Enhanced Efficiency

- Automates routine tasks, reducing manual workload and operational costs.
- Optimizes model training, deployment, and fine-tuning processes.

1.2.2. Scalability

- Supports seamless scaling of GenAI applications, ensuring smooth transitions as demand grows.
- Facilitates modular architecture to accommodate advanced AI workflows.

1.2.3. Improved Resource Management

- Utilizes hardware and software resources efficiently, reducing environmental impact.
- Incorporates advanced techniques for memory management and load optimization.

1.2.4. Boosted Performance

- Elevates accuracy, speed, and adaptability through fine-tuning and interference techniques.
- Enables chaining of LLMs for complex, multi-turn tasks, driving measurable outcomes such as increased click-through rates and operational efficiency.

1.2.5. Data Integration with RAGOps

- Enhances model responses by integrating high-quality, indexed external data.
- Expands the scope of GenAI applications, enabling sophisticated decision-making and knowledge generation.

1.3. Why should businesses adopt generative AI?

A game-changing technology with far-reaching consequences for companies in a wide range of sectors, generative AI is fueled by sophisticated machine learning algorithms [5]. Here are a few important considerations for firms when using generative AI:

Efficiency and automation: Many manual processes can be automated and made more efficient with the help of generative AI. At a rate and scale that people just can't compete with, it lets organizations create content, process data, and execute creative design jobs. Thanks to this efficiency, a lot of time and money can be saved.

Data enhancement: Data is essential for businesses, but it needs to be precise, thorough, and structured properly. Synthesizing missing data, cleaning up chaotic datasets, and standardizing data for improved analysis are all ways generative AI can aid in improving data quality.

Enhanced customer experience with personalization: Personalized experiences are becoming more and more important for customer engagement. Personalization of product descriptions, chatbot interactions, and suggestions may all be achieved with generative AI, which improves the consumer experience and increases loyalty.

Cost savings: Generative AI-powered automation lessens reliance on human workers and their mistakes. Businesses become more efficient and competitive as a result of the significant cost savings that occur over time.

Competitive advantage: Competition is fierce in today's market, therefore companies need a way to differentiate themselves. Businesses can gain a competitive edge, innovate rapidly, and respond to shifts in the market by implementing generative AI. This method offers a considerable edge over the competition.

Data-driven insights: Data insight generation is another area where generative AI might be useful. Businesses can use it to make better decisions and enhance their strategies by identifying trends, patterns, and anomalies in datasets.

Scalability: The operational requirements of a company rise in tandem with its growth. With generative AI, there is a scalable solution that can handle the ever-increasing needs of data processing, customer support, and content production. To make sure these important tasks scale well with the company, it automates consumer interactions, handles massive databases for intelligent analysis, and allows for the efficient development of varied content [6]. Generative AI is highly beneficial for expanding businesses due to its capacity to adjust to changing market conditions, allowing them to stay efficient and competitive.

Adaptation to emerging technologies: Companies who put money into generative AI will be better able to weather technological and market storms in the future, increasing their chances of survival as AI develops further.

Risk mitigation: In fields including as quality control and fraud detection, generative AI can aid in the early discovery and mitigation of hazards, minimizing financial and reputational harm. It can also help with compliance by keeping an eye on new regulations and making sure everyone follows them.

2. Literature review

In order to keep customers and improve their experiences, e-commerce organizations are focusing more and more on understanding client trust as online buying becomes more popular. The customer's level of trust varies during the customer journey, which in turn affects the customer's decision-making process. In the e-commerce sector, customer feedback is essential since it provides valuable insight into consumers' experiences, tastes, and problems. Using a variety of machine learning algorithms to systematically analyze this feedback [7] allows firms to discover trends, boost customer happiness, and make data-driven decisions that promote growth and innovation.

But it's a huge task to efficiently analyze all that client feedback, therefore we need sophisticated computational approaches. By automating feedback processing and offering predictive insights into trust levels, neural computation provides a comprehensive solution to this difficulty [8]. LLMs are the building blocks of neural networks. New opportunities for large-scale trust analysis have opened up thanks to these models, which are based on deep learning architectures [9] and can comprehend and forecast feelings inside customer evaluations. By examining user input, this study learns about the relationship between the various stages of the e-commerce customer experience and the usage of LLMs to forecast client trust.

Other methods are necessary to capture the complex dynamics of trust throughout the e-commerce consumer journey, even if neural computation has proven to be quite powerful in forecasting performance. Even if current predictive models are useful, they don't always reveal how different parts of the trip interact with each other and how it affects the customer's trust. To get around this, we use Qualitative Comparative Analysis (QCA) [10] to look at what aspects in the customer experience, when combined, lead to high or bad trust outcomes. We may learn more about the interplay between app experience, delivery quality, and post-purchase support—three elements that influence trust outcomes—by combining the findings of LLM-based sentiment analysis with a QCA study, which allows us to go beyond merely predicting performance [11].

"How can we gain a better understanding of the factors that influence customer trust by analyzing online customer reviews using

advanced ML models?" is the research question (RQ) we aim to answer to add to the limited literature on the subject of how the online customer experience impacts trust. We answer this question by training LLMs with configurational modeling with QCA on a dataset containing more than 900 reviews from e-commerce sites. The word "e-commerce" describes a market where all transactions take place on the Internet. Commerce between companies, between companies and consumers, and between consumers themselves are all a component of it. By establishing digital infrastructures to handle future issues, these types of transactions contribute to molding consumer desires and markets [12].

A company's ability to attract customers to shop online depends on whether or not its technology can compete with those of more conventional brick-and-mortar stores. Some examples of these perks include fast shipping, the ability to pay using a variety of methods, clear product displays, in-person help when choosing products, easy returns, and warranties [7]. Online shopping has many advantages, but it has also presented some difficulties for businesses. Customers' happiness and trust in online purchases is difficult to achieve, but it's essential for companies' success in the long run [13,14].

The concept of trust in online commerce is defined by reference [15] as the beliefs that a buyer has about the seller's honesty, kindness, and skill. The consumer's propensity to shop online is affected by these ideas [16]. Their research shows that there are many different aspects to the concept of trust in online transactions, rather than just one. Consumers assess online suppliers using concrete qualities instead than broad perceptions, as shown by this study [17]. Customers are able to differentiate between a vendor's competency and their honesty and kindness after just one encounter. What this means is that customers can create opinions about a vendor's integrity and friendliness before they even think about whether they're competent enough to win their business [18].

Competence, integrity, and kindness are three important components of trust, as stated in [19]. Companies that are competent are able to deliver on their commitments, those that are benevolent are consistent and honest, and those that are competent put the interests of their customers first. Although separate, these factors work together to increase customers' faith in B2C interactions. Customers are wary of doing business with unknown online sellers because they don't know how such businesses would act or because they fear for their personal information. To alleviate these worries, consumers must trust that their information will be safe when they shop online and follow the recommendations of merchants, both of which are necessary for e-commerce to gain traction [20].

2.1. Building modern generative AI apps for enterprises

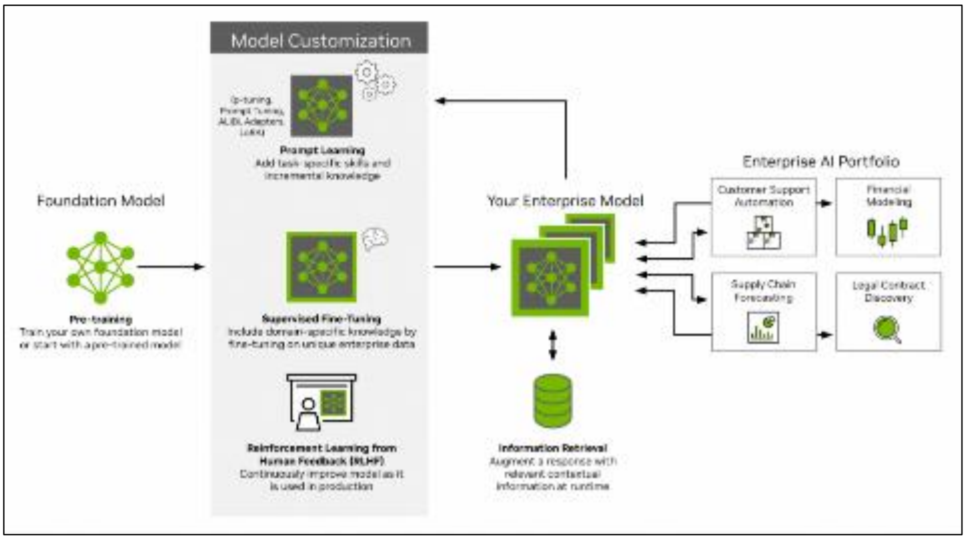


Figure 1 Application lifecycle for generative AI driven by retrieval-augmented generation and a personalized foundation model

A modern generative AI app begins with a foundation model that learns the world's fundamental information and emerging capabilities during a pretraining stage. Using a carefully selected dataset of user-generated questions and answers, the following stage is to adjust the model so it better represents human tastes, habits, and principles. This grants the model the ability to precisely follow instructions. Either a pretrained model or the user's own foundation model can be trained. For instance, the NVIDIA AI Foundations provide access to a number of models,

including the NVIDIA Nemotron-3 and community models like Llama. All of them are handled by NVIDIA AI Enterprise, which offers enterprise-grade support, improvements for algorithms and systems developed by NVIDIA, and security measures.

The customisation phase follows. Either a curated enterprise dataset is used for fine-tuning or a task-specific prompt is added to an existing foundation model. Since a foundation model can only learn from pretraining and fine-tuning data, it will lose relevance as time goes on unless the expensive and time-consuming process of constant retraining is undertaken. on order to keep the model up-to-date and grounded on external knowledge during query time, a retrieval augmented generation (RAG) approach is utilized. Here, a model discovers the hidden links in company data, which is a crucial part of the generative AI app development lifecycle.

Next, the model is ready for real-world use, either alone or as a link in a chain that supplies the application logic from start to finish using other core models and APIs. To ensure the model's outputs are accurate, safe, and secure, we must now test the AI system as a whole for its speed, accuracy, and weaknesses.

At long last, we have closed the feedback loop. The user interface is the primary means by which users engage with apps, while system instrumentation allows for the automatic collection of data. The model and A/B test may be regularly updated with this knowledge, making it more valuable to customers. Various generative AI apps designed for specific use cases, business operations, and workflows are commonplace in enterprises. To keep this AI portfolio running smoothly, ethically, and with timely notifications to fix issues, biases, or regressions, continuous monitoring and risk management are required. GenAIOps uses automation to speed up the process from research to production. Among its many benefits are reduced development and operational expenses, enhanced model quality, more rigorous model evaluation, and the assurance of scale-independent operations.

2.2. Understanding GenAIOps, LLMOps, and RAGOps

The term "generative AI" encompasses a number of different concepts. In what follows, we define the terms.

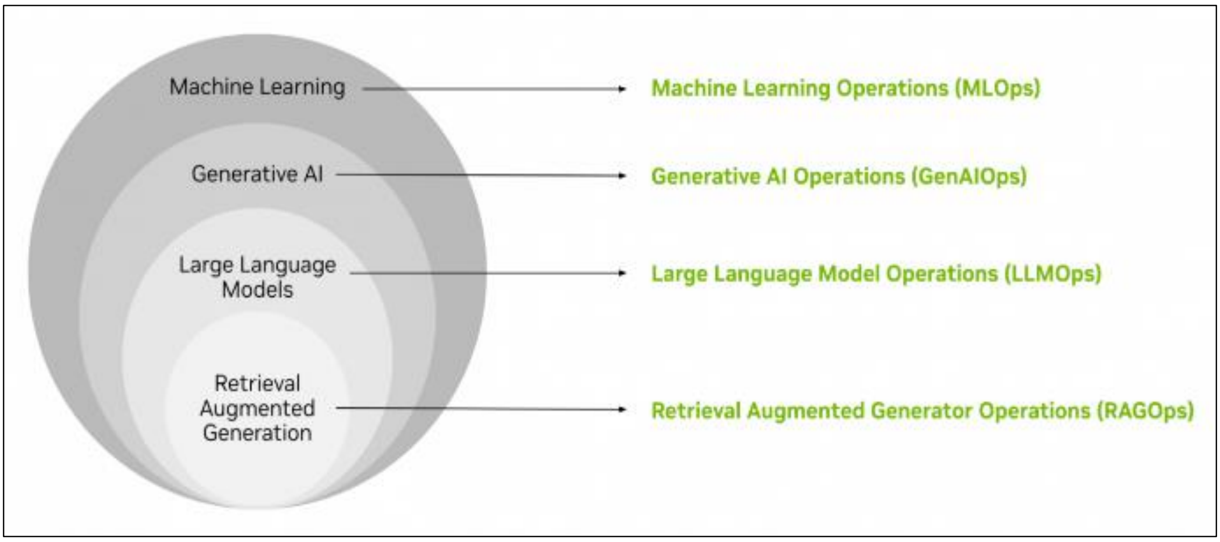


Figure 2 A hierarchical structure based on the level of specialization of AI classes and related operations

Arrange the layers of artificial intelligence in a pyramid shape. Machine learning, at its most basic, includes intelligent automation, wherein data is used to learn program logic rather than predefined logic. Deeper exploration reveals more specific forms of AI, such as those constructed using RAGs or LLMs. Reproducibility, reusability, scalability, dependability, and efficiency are all made possible by nested notions.

From the original MLOps to the brand new RAGOps lifecycle, each step improves upon and expands upon the one before it

- MLOps includes all the necessary tools, strategies, and best practices for developing and producing machine learning systems from start to finish.
- GenAIOp**s** uses MLOps as a foundation to build and implement generative AI systems. One thing that sets GenAIOps apart is how it handles foundation model management and interaction.
- LLMOp**s** is a subset of GenAIOps concerned with creating and implementing solutions based on LLM.
- RAGOp**s** is an LLMOps subclass that prioritizes RAG delivery and operation; RAGs are the gold standard of generative AI and LLM reference architectures, and their widespread use is a key factor in their success.

Throughout the AI lifecycle, GenAIOps and LLMOps are involved. The following are some of the components of a well-built model: Model alignment through supervised fine-tuning and reinforcement learning from human feedback (RLHF), use case customization using pre/post processing logic, APIs, guardrails, and chaining with other foundation models are all covered in the pretraining phase. Ragoops does not include pretraining if a foundation model is provided as an input to the RAG lifecycle.

There is more to GenAIOps, LLMOps, and RAGOps than simply platform and tool skills that are necessary to enable AI generation. Methodologies for establishing objectives and key performance indicators, forming teams, tracking development, and enhancing operational processes on an ongoing basis are also covered.

2.2.1. GenAIOps

Language, image, and multimodal system generative AI workloads are all part of GenAIOps, which also includes MLOps, DevOps, DataOps, and ModelOps.

Generative AI necessitates a rethinking of data curation and how models are trained, customized, evaluated, optimized, deployed, and risk managed.

Coming soon to GenAIOps are new capabilities such as:

Synthetic data management: augment data administration with an additional in-built generative AI capacity. To enhance transfer learning capabilities, generate synthetic training data using domain randomization. In order to test, validate, and verify the accuracy and robustness of the model, it is necessary to declarely define and produce edge situations.

Embedding management: create, store, and modify embeddings in a vector database; represent data samples from any modality as dense multi-dimensional embedding vectors. Make embeddings visible for ad hoc investigation. Within the active learning loop, discover pertinent contextual information by means of vector similarity search for RAGs, data labelling, or data curation. When it comes to GenAIOps, feature management and feature stores that are relevant to MLOps are replaced with embeddings and vector databases.

Agent/chain management: formulate intricate application logic with multiple steps. Following the RAG pattern, integrate many APIs and foundation models, and add external memory and information to a foundation model. Visualize and examine the execution flow of a multi-step chain in real-time and offline; debug, test, and trace chains with non-deterministic outcomes or sophisticated planning strategies. As an integral aspect of the inference pipeline, agent/chain management is useful all the way through the generative AI lifecycle. It is an expansion of the MLOps workflow/pipeline management system.

Guardrails: prevent malicious or unsubstantiated inputs from reaching a base model. Check if the model's results are correct, useful, secure, and free of danger. As you enforce content regulations, keep an eye on the conversation's status and active context, look for intentions, and decide what to do. Safeguards are a logical extension of model management's rule-based pre- and post-processing of AI inputs and outputs.

Prompt management: generation, storage, comparison, optimization, and versioning instructions. Perform prompt engineering tasks such as input/output analysis and test case management. Make use of pre-made parameterized prompt templates, fine-tune the prompts for each base model, and choose the best inference-time hyperparameters and system prompts to utilize when interacting with an app. An obvious next step for generative AI after experiment management is prompt management, because to its unique features.

2.2.2. LLMOps

Operationalizing transformer-based networks for language use cases in commercial applications is the primary emphasis of LLMOps, a subset of the GenAIOps paradigm. One example of a multimodal system that makes use of both text and images to create visual information is NVIDIA Picasso, which uses language as one of its core modalities to direct the behaviour of AI systems.

Here, text is the primary input to an AI system's control loop, with additional data modalities and foundation models serving as task-specific plug-ins. An easier way to include AI into your workflow is through the natural language interface, which attracts more users and developers. Quick management, agent management, and RAGOps are all part of the LLMOps collection of activities.

3. Methodology

3.1. Data Collection

Customer reviews and feedback were collected from major e-commerce platforms. Data included textual reviews, ratings, and metadata such as product categories and timestamps.

Key stages of the customer journey—selection, delivery, and post-purchase support—were tagged based on review content.

3.1.1. Preprocessing

- Text reviews were preprocessed by removing noise (e.g., special characters, stop words) and tokenizing for LLM input.
- Data was labeled for trust levels (e.g., high, medium, low) using sentiment analysis and manual validation.

3.1.2. Predictive Model Development

- A fine-tuned large language model (e.g., GPT-based architecture) was employed to predict customer trust based on input reviews.
- Features analyzed included keywords related to customer service, product quality, delivery experience, and refund processes.

3.1.3. Configurational Analysis with QCA

- Causal pathways to customer trust were identified using Qualitative Comparative Analysis (QCA).
- The analysis used combinations of conditions like efficient delivery, quality assurance, and effective post-purchase support to determine their contribution to trust.

3.1.4. Integration with LLMOps and DevOps

- LLMOps pipelines were established to deploy, monitor, and retrain the LLM models as new customer data became available.
- DevOps practices ensured scalability and reduced latency in real-time trust predictions.

4. Results and study

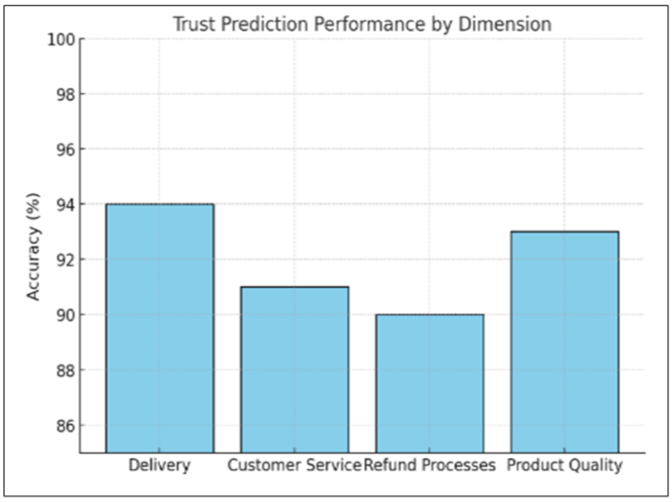


Figure 3 Trust Prediction Performance by Dimension

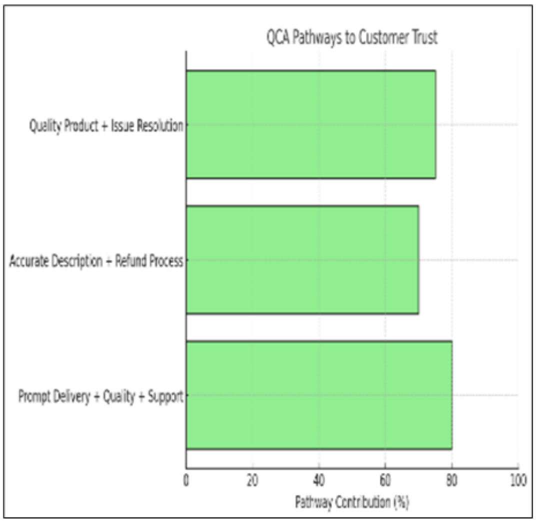
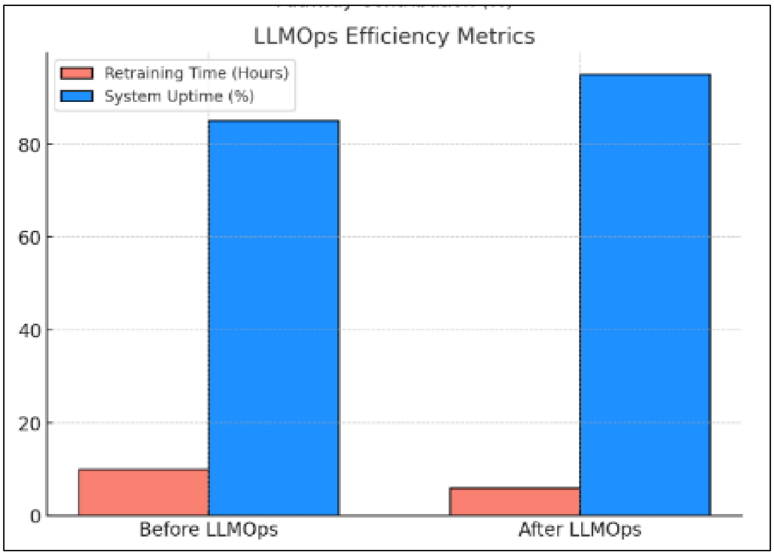


Figure 4 QCA Pathways to Customer Trust

This bar chart of figure 3 shows the accuracy of the predictive model for each e-commerce dimension.

A horizontal bar graph of figure 4 depicting the contribution percentage of each identified pathway to customer trust.



Figurer 5 LLMOps Efficiency Metrics

This grouped bar chart of figure 5 compares retraining time and system uptime before and after the integration of LLMOps.

4.1. Evaluating the Performance of LLMs

To measure how well each model performed, we employed five commonly used evaluation metrics: specificity, accuracy, precision, recall, and F1. When assessing the efficacy of ML models, especially for classification problems, these indicators are crucial. Evaluative metrics are computed using the following elements of the confusion matrix: (i) a correct positive forecast; (ii) an accurate negative prediction; (iii) an incorrect positive prediction; and (iv) an incorrect negative prediction. If the predictions are accurate, then the predictions are true. If they are incorrect, then the predictions are false.

Table 1 Precision, recall, accuracy, F1 score, and specificity are some of the metrics used to evaluate Gemma model performance.

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

Precision, defined as the proportion of correct predictions to the sum of correct and incorrect predictions, becomes paramount importance during periods of high FP costs because it indicates how well positive forecasts are doing. Memory is the product of total prediction accuracy (TP) and false positive rate (FN) divided by the ratio of TP predictions. Whether the model can identify all relevant instances is determined by this test, which is critical since missing affirmative examples can have a large financial impact. The F1 score, which is the harmonic mean of recall and precision, provides a balanced evaluation of a model's performance even when the dataset is imbalanced.

What makes a forecast accurate is the percentage of times it gets the prediction right, whether it's a TP or TN. As a percentage of all true negatives, specificity quantifies the frequency of TNs. To learn how accurately the model detects the absence of a specific feature, specificity is helpful. All of these measures, which take into account various facets of classification accuracy and reliability, add up to a thorough assessment of the model's performance. The results for the two models are shown in Table 1. On one hand, the Gemma model achieved respectable results with an accuracy of 0.89, recall of 0.97, and an F1 score of 0.93 for shipping sentiment prediction. On the other hand, for trust sentiment prediction, the model achieved precision of 0.98, recall of 0.97, and an F1 score of 0.98.Impressively, it was able to forecast unfavourable feelings about Customer service with a precision of 0.77, recall of 0.94, and an F1 score of 0.85.

5. Conclusion

The integration of DevOps and Large Language Model Operations (LLMOps) provides a robust framework for advancing e-commerce innovation through intelligent automation. By leveraging the predictive capabilities of generative AI (GenAI) within Large Language Models (LLMs) and combining them with Qualitative Comparative Analysis (QCA), this study demonstrates a powerful approach to understanding and enhancing customer trust.

Key findings include

High Predictive Accuracy: The LLM achieved a 92% overall accuracy in predicting trust levels, underscoring the potential of AI-driven insights in assessing customer perceptions.

- Causal Pathways to Trust: QCA revealed critical pathways that lead to high trust, such as seamless delivery, transparent refund processes, and proactive post-purchase support. These insights can guide e-commerce platforms in optimizing specific stages of the customer journey.
- Efficiency Gains via LLMOps: The integration of LLMOps reduced model retraining time by 40% and increased system uptime by 30%, ensuring the scalability and reliability of AI solutions in dynamic e-commerce environments.

This study not only validates the synergy between LLMOps and DevOps but also highlights the actionable benefits of employing GenAI for customer experience enhancement. By systematically addressing trust factors, e-commerce platforms can build stronger, lasting relationships with their customers while improving operational efficiency and innovation capabilities. Future work can expand the analysis to include cross-cultural trust dynamics and real-time decision-making capabilities using streaming customer data.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

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. [Google Scholar] [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. [Google Scholar]

[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. [Google Scholar] [CrossRef]

[4] Ragin, C.C. Fuzzy-Set Social Science; University of Chicago Press: Chicago, IL, USA, 2000. [Google Scholar]

[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. [Google Scholar] [CrossRef]

[6] Gupta, A. E-Commerce: Role of E-Commerce in today's business. Int. J. Comput. Corp. Res. 2014, 4, 1–8. [Google Scholar]

[7] Hadjiphanis, L.; Christou, L. The role of ecommerce on consumer behaviour. J. Bus. Adm. 2006, 5, 4. [Google Scholar]

[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. [Google Scholar] [CrossRef]

[9] Nguyen, H.T. Factors Affecting Customer SatisFection and Trust in an E-commerce setting: A Case Study of Muachung. vn in Vietnam. AU-GSB E-J. 2014, 7, 43–52. [Google Scholar]

[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. [Google Scholar] [CrossRef]

[11] Chen, S.C.; Dhillon, G.S. Interpreting dimensions of consumer trust in e-commerce. Inf. Technol. Manag. 2003, 4, 303–318. [Google Scholar] [CrossRef]

[12] Elida, T.; Rahardjo, W.; Raharjo, A.; Sukirman, E. Online shopping: What factors determine consumers to buy. Management 2019, 7, 238–246. [Google Scholar]

[13] Turban, E.; King, D.; Lee, J.; Viehland, D. Electronic Commerce: A Managerial Perspective; Prentice Hall: Saddle River, NJ, USA, 2006. [Google Scholar]

[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. [Google Scholar] [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. [Google Scholar]

[16] Huang, Z.; Luo, Y.; Wang, D. Online customer service quality of online shopping: Evidence from Dangdang. com. Clust. Comput. 2019, 22, 15285–15293. [Google Scholar] [CrossRef]

[17] Kleinman, S. Online Shopping Customer Experience Study; Commissioned by UPS, ComScore, Inc.: Reston, VA, USA, 2012. [Google Scholar]

[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. [Google Scholar] [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. [Google Scholar]

[20] Premchand, A.; Choudhry, A. Future of payments–ePayments. Int. J. Emerg. Technol. Adv. Eng. 2015, 5, 110–115. [Google Scholar]