

GENERATIVE AI REPORT

AI IMPLEMENTATION INSIGHTS 2024



SEPTEMBER 2024

CONTENTS

INTRODUCTION	02
1. Overview of AI Trends Across Industries	03
2. Methodology and Scope of Report	03
SECTION 1: Industry-Specific AI Implementations	04
1. Delivery and Mobility Sector	04
• Picnic: Create Personalized Shopping Lists	04
• Uber: Personalize Out-of-App Communications	07
• Careem: Mitigate Fraud Risks with Pre-Authorization	10
• Foodpanda: Balance Demand and Supply for Food Delivery	14
• Lyft: Provide Accurate ETAs	17
• Doordash: Extract Product Attributes from Unstructured Data	21
2. E-commerce and Retail Sector	25
• Etsy: Predictive Recommendations	25
• Mercado Libre: LLM-Based Use Cases	28
• Walmart: Handling Imbalances in Text Data	33
3. Tech Sector	37
• Algolia: Tailoring Content for Online Visitors	37
• GitLab: Testing Quality of AI-Generated Outputs	41
• Dropbox: Prevent Token Attacks on LLMs	43
• Feedzai: Transaction Risk Scoring	48
• Amazon: Find and remove damaged products	51
4. Travel and E-commerce Sector	53
• Expedia: Personalized Travel Recommendations	53
• Netflix: Classify video content	55
SECTION 2: Cross-Industry Analysis	59
1. Common AI Techniques Employed	59
2. Challenges Faced During Implementation	60
3. AI Adoption Trends and Future Directions	61
CONCLUSION	63
1. Summary of AI Impact Across Industries	63
2. Forecast of AI Advancements in the Coming Year	63
REFERENCES	65

INTRODUCTION

In today's fast-paced world, AI is revolutionizing how companies operate and serve their customers. As part of the generative AI community, we're excited to share key insights into how businesses are using AI to innovate and stay ahead. With more people turning to AI hubs for guidance, our monthly report, AI Implementation Insights 2024, offers a closer look at how leading companies are applying these technologies—highlighting both the wins and the hurdles they face.

At GenAI.Works, we're on a mission to make AI accessible to everyone. Our growing community of over 7 million followers brings together experts and enthusiasts to share the latest breakthroughs, foster learning, and drive real-world impact. Whether it's through our news updates, education courses, hackathons, events, or startup initiatives, we're building an ecosystem where AI is for everyone.

Now's the perfect time to join us. Invest in GenAI.Works today and get up to 25% in free shares if you invest by October 20, 2024. Visit [Investment Page](#) to learn more.

1. Overview of AI Trends Across Industries

From personalized shopping recommendations in the retail sector to real-time fraud detection in financial services, AI technologies are being employed to solve industry-specific challenges.

In the delivery and mobility sector, AI is used to provide accurate ETAs and mitigate fraud risks, while in e-commerce, it powers predictive recommendations and product attribute extraction. Across industries, companies are experimenting with large language models (LLMs), machine learning (ML), and AI-powered automation to streamline operations and enhance decision-making.

2. Methodology and Scope of Report

This report is based on an in-depth analysis of AI implementations across multiple industries, highlighting real-world use cases from renowned companies such as Uber, Picnic, Walmart, and Netflix. We have examined the specific AI technologies employed, the benefits they offer, and the challenges encountered during their deployment. The report focuses on identifying common AI techniques, the challenges of AI adoption, and future trends, providing readers with a well-rounded view of the state of AI across sectors in 2024.

SECTION 1:

INDUSTRY-SPECIFIC AI IMPLEMENTATIONS

1. DELIVERY AND MOBILITY SECTOR

PICNIC: CREATE PERSONALIZED SHOPPING LISTS

Picnic, a leading online supermarket often referred to as the "modern milkman," strives to simplify the shopping experience for customers by leveraging AI and machine learning. One of the key components of this initiative is the use of recommendation systems, designed to enhance personalization at every step.

From predicting repeat purchases to recommending new items, Picnic's AI-driven solutions aim to optimize the grocery shopping process for both convenience and customer satisfaction. In this report, we'll explore how Picnic has applied AI to address the specific challenge of shopping recommendations, examine its benefits, and consider the potential impact across industries.

User 1



*Chips
amount x
brand A*



*Soda
amount y
brand B*



*Beer
amount z
brand C*

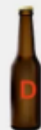
User 2



*Chips
amount x
brand A*



*Soda
amount y
brand B*



*Beer
amount z
brand D*

Solution: AI-Powered Shopping Recommendations at Picnic

Picnic's recommendation system revolves around models designed to predict customer behavior and recommend items accordingly. One such model, the Customer Article Rebuy Prediction (CARP), plays a central role in personalizing the shopping experience by predicting repeat purchases. The model's simplicity—powered by XGBoost—has proven to be highly effective in identifying patterns in customers' past buying habits. CARP helps Picnic customers build a tailored shopping list based on prior purchases, streamlining the process by surfacing frequently bought items.



However, Picnic's ambition doesn't stop at repeat purchases. The company aims to go beyond by introducing recommendations for items a customer hasn't yet bought—so-called explore recommendations. The challenge here lies in the vast assortment of 30,000 items available in Picnic's catalog, which creates a sparsity problem. Most customers buy from a limited range of products (roughly 30–60 items weekly), making it difficult to gather enough data for accurate recommendations on lesser-known items.

To address this, Picnic has experimented with several advanced AI models including Markov Chains, Recurrent Neural Networks (RNNs), and Transformer-based models. These models use sequential recommendation techniques, designed to predict what a customer might want to add to their shopping basket based on historical behavior. Despite progress, the complexity of predicting "explore" items continues to pose significant challenges.

Key Benefits & Innovations

Picnic's AI-driven shopping list model, CARP, offers several advantages that set it apart from competitors. First, its simplicity and effectiveness are a standout feature. By focusing on repeat purchase patterns, CARP enables Picnic to cater to a significant portion of customer behavior, which is predominantly repetitive. This not only improves shopping efficiency but also ensures that customers are presented with highly relevant products without having to sift through a massive catalog. Specifically, their next-basket adaptation of sequential models represents an innovative step in predicting multiple products for a customer's future orders. This sets Picnic apart by focusing not on a single-item recommendation but on predicting an entire basket, thereby enhancing the user experience.

Challenges and Risks

Despite its success, there are limitations and challenges to Picnic's AI solution. The CARP model, while excellent for repeat purchases, cannot handle the exploration of new products—a crucial feature for encouraging variety and discovery in customers' shopping behavior. Scalability also presents a potential issue. As Picnic continues to grow, the sheer volume of data and the need to maintain quick, personalized recommendations may put pressure on the infrastructure.

Conclusion

In our opinion, Picnic's approach to leveraging AI for shopping list recommendations is both practical and forward-thinking. The CARP model demonstrates how even simple machine learning models can offer substantial benefits when applied in the right context. By streamlining the grocery shopping experience, Picnic has positioned itself as a leader in AI-driven retail personalization.

Overall, Picnic's application of AI showcases the power of tailored solutions in enhancing user experience, efficiency, and satisfaction. The company's focus on building customer-centric models not only improves shopping convenience but could inspire future AI innovations across a variety of industries.

UBER: PERSONALIZE OUT-OF-APP COMMUNICATIONS

Uber, known for its innovative use of technology, has applied artificial intelligence (AI) to personalize out-of-app (OOA) communications, such as emails, SMS, and push notifications. The primary goal of this system is to recommend relevant restaurants, promotions, and services to users based on their preferences and location.

However, delivering these personalized recommendations without the benefit of real-time user data presents significant challenges. Uber's AI-driven solution solves these problems using machine learning (ML) to create a scalable and effective recommendation system.

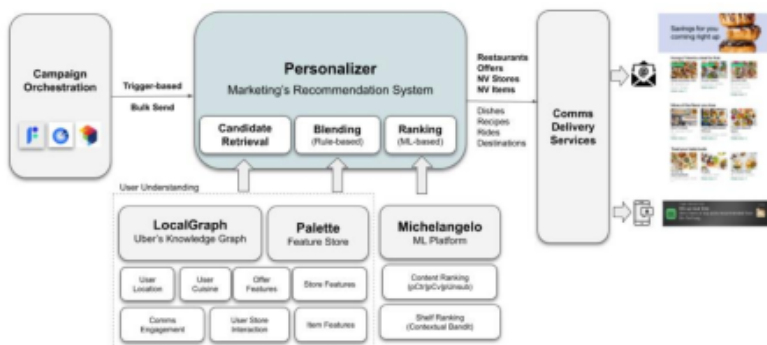
Solution

The main challenge with Uber's OOA system is the lack of immediate user context—unlike in-app recommendations, which rely on real-time data like the user's current location. To address this, Uber developed a machine learning model that predicts where a user is most likely to

place their next order. This prediction is based on historical data from both Uber Eats and Uber rides, ensuring recommendations remain relevant even when user context is missing.

To further improve relevance, the system adapts to user behavior changes, such as when users travel to new areas. For example, if a user frequently travels or changes locations, the system updates the restaurant suggestions to match their current location.

How | An extensible recommendation platform that aggregates user's lifetime & realtime signals from across Uber and optimizes for various marketing objectives.



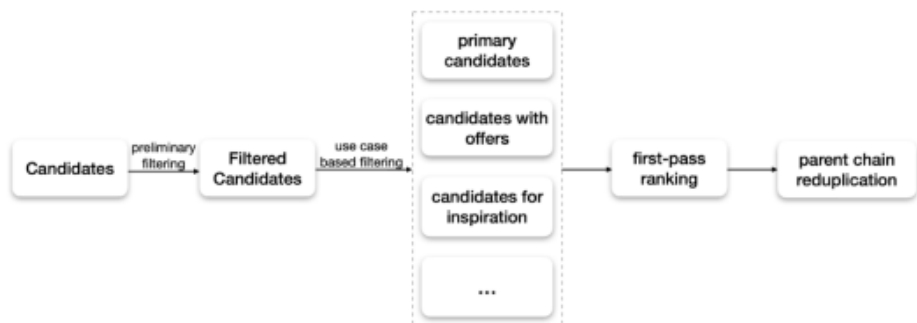
The solution handles a large volume of data—sending out over 4 billion personalized messages across different regions and services. To keep this process efficient, Uber uses techniques like optimizing algorithms and using compact data storage methods, which help reduce both storage and computational costs.

Key Benefits & Innovations

Uber's personalized OOA system stands out due to its innovative approach in handling user context. Traditional recommendation systems rely on real-time data, but Uber's system can predict user behavior even without immediate context.

This is done by analyzing long-term data from a user's previous interactions with Uber, allowing the system to make accurate predictions and provide timely recommendations.

The system's ability to scale globally is another impressive feature. Managing billions of personalized messages while keeping the process cost-effective shows how Uber has optimized its system for performance.



One key innovation is the use of data "embeddings," which compress large amounts of user information into smaller, more manageable sets of features. This method allows Uber to store and process data efficiently, reducing infrastructure costs while still delivering highly personalized content.

Furthermore, the system's flexibility is a major strength. Uber can tailor recommendations based on specific marketing strategies, giving local teams the ability to tweak recommendations for their regions. This flexibility ensures that both user preferences and business objectives are met in every communication.

Challenges and Risks

Despite its impressive capabilities, Uber's recommendation system faces certain challenges. One issue is the potential for inaccuracy

when predicting user behavior. Since the system relies on historical data, sudden changes in user habits—like a person taking a trip—may lead to irrelevant recommendations. Uber has addressed this to some extent with real-time updates, but there could still be scenarios where recommendations miss the mark.

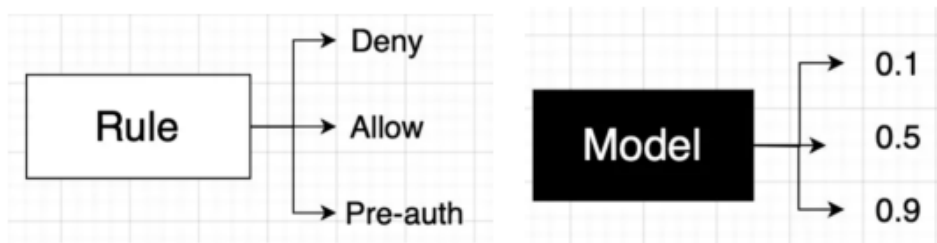
Another significant challenge is cost management. The system's complexity, combined with the need for real-time data processing and massive scalability, can lead to high operational costs. Even though Uber has implemented optimizations to control these expenses, running AI systems at this scale requires continuous investment in computing power, data storage, and infrastructure.

Conclusion

Uber's personalized out-of-app recommendation system showcases the power of AI in marketing at a global scale. Its ability to predict user behavior without real-time data is a remarkable innovation that addresses key challenges in personalized communication. By blending machine learning insights with flexible business strategies, Uber has created a system that not only delivers relevant recommendations but does so in a scalable, cost-efficient manner.

CAREEM: MITIGATE FRAUD RISKS WITH PRE-AUTHORIZATION

Careem, the leading ride-hailing and digital services platform in the Middle East, has implemented an AI-driven solution to address a growing challenge – fraud prevention. With over 1 billion rides completed, Careem's Everything App offers a range of digital services that has inevitably attracted the attention of fraudsters.



In response, Careem transitioned from traditional rule-based pre-authorization mechanisms to a sophisticated machine learning (ML) model. This shift has improved both security and user experience, marking a significant milestone in Careem's approach to fraud prevention.

Solution: Transition to Machine Learning for Pre-Authorization

Careem's previous method for fraud prevention was a rule-based system that applied pre-authorization holds on customer credit cards to prevent fraud. This approach, while effective to an extent, had limitations. Pre-authorizing all transactions risked alienating customers who viewed the process as intrusive or inconvenient. Additionally, fraudsters began reverse-engineering some of the rules, reducing the effectiveness of these defenses over time.

Recognizing these challenges, Careem adopted a more flexible machine learning model for pre-authorization. Unlike rule-based systems, which rely on strict conditions, ML models use data to predict potential fraud by analyzing patterns in real-time. This shift enables a more adaptive approach to identifying fraud, allowing Careem to fine-tune its fraud detection system and provide more nuanced outcomes such as approvals, denials, or holds on transactions.



In response, Careem transitioned from traditional rule-based pre-authorization mechanisms to a sophisticated machine learning (ML) model. This shift has improved both security and user experience, marking a significant milestone in Careem's approach to fraud prevention.

Key Benefits & Innovations

The adoption of machine learning for fraud prevention has provided Careem with several advantages. First, the use of machine learning models, specifically CatBoost, has improved precision by 60%, significantly reducing the number of false positives and enhancing the user experience. This has also led to a decrease in customer complaints, as legitimate transactions are less likely to be wrongly flagged.

Another key benefit is the adaptability of the ML models. Unlike fixed rules, machine learning models adjust based on new data, making them more effective at handling evolving fraud tactics. Additionally, the model's real-time responsiveness is critical in ride-hailing services where transactions need to be processed quickly to maintain a smooth

user experience. Careem's ability to manage large datasets effectively, while focusing only on the most relevant recent data, ensures that the models remain efficient without becoming overloaded with outdated information.

This innovation sets Careem apart from competitors who may still rely on static rules that can quickly become obsolete. By leveraging AI to refine fraud prevention in real-time, Careem has established itself as a leader in adopting cutting-edge technology to enhance both security and customer satisfaction.

Challenges and Risks

Despite the clear advantages of Careem's machine learning approach, there are inherent challenges and potential risks that must be acknowledged.

First, model performance validation is a complex task, especially when other rules and overrides affect transaction outcomes. For instance, if the model predicts a fraudulent transaction, but an override denies it, it becomes difficult to assess the model's true accuracy. This introduces a layer of complexity in determining how well the ML system is performing.

Furthermore, there are ongoing concerns related to scalability. As the platform continues to grow, maintaining real-time responsiveness while processing massive amounts of data could present future bottlenecks.

The constant evolution of fraud tactics also demands continuous retraining and updating of the models, which could introduce delays or errors if not handled efficiently.

Conclusion

Careem's transition from a rule-based fraud prevention system to a machine learning model marks a significant advancement in both security and customer experience. By embracing AI, Careem has managed to increase fraud detection precision, adapt to evolving fraud tactics, and maintain the seamless user experience that customers expect from a digital platform.

In our opinion, Careem's AI strategy provides a valuable framework for other industries, setting a precedent for how machine learning can be used to balance security and user satisfaction. This solution not only strengthens Careem's defenses but also offers a glimpse into the future of fraud prevention across sectors.

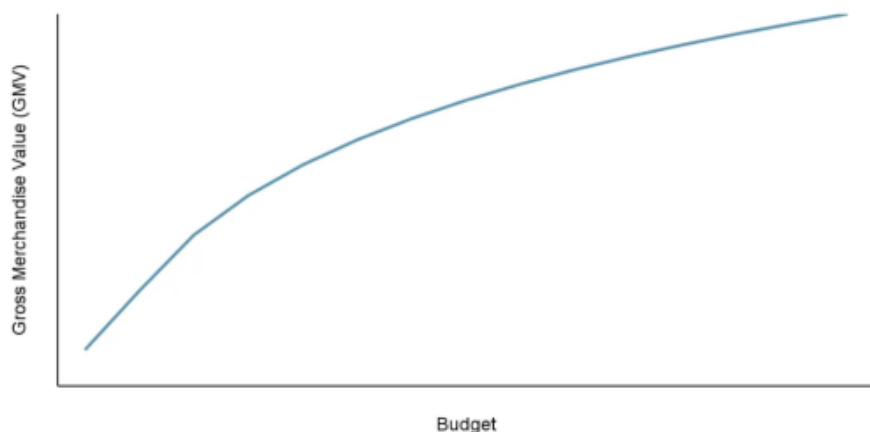
FOODPANDA: BALANCE DEMAND AND SUPPLY FOR FOOD DELIVERY

Foodpanda, a global leader in food delivery, recently introduced an innovative AI-driven initiative aimed at optimizing the balance between customer demand and rider supply. This initiative, named the "Marketplace Balance Project," is a forward-thinking solution that leverages machine learning to manage budgeting for incentives and rider payments effectively.

By analyzing historical data and trends, the company seeks to predict customer orders and rider availability more accurately, thereby preventing imbalances that could negatively affect service quality or financial performance. The core of this project is the creation of a recommendation tool designed to help decision-makers allocate resources efficiently to meet the company's growth and profitability objectives.

GMV Projections by Budget

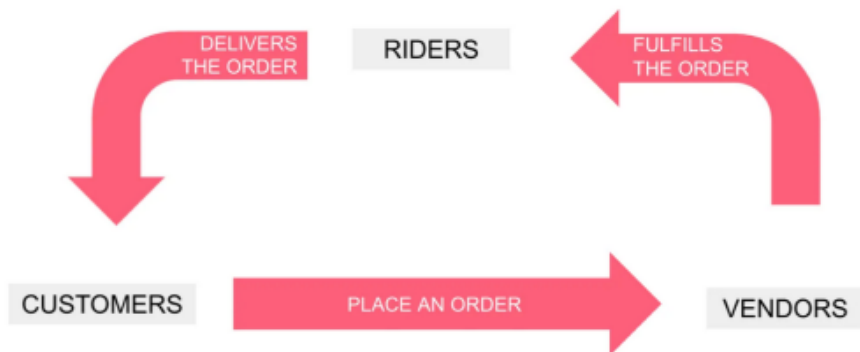
Logarithmic Growth: Improvements come very quickly, but gains decrease over time



Solution

At the heart of the Marketplace Balance Project is an advanced machine learning model designed to predict the relationship between key variables like customer acquisition costs (CARC) and rider payment incentives. The goal is to create a scenario where customer demand and rider supply are perfectly aligned, ensuring that the platform can handle a high volume of orders without overextending its resources.

The recommendation tool Foodpanda developed allows decision-makers to simulate different budget allocations for CARC and rider incentives, forecasting their impact on metrics like Gross Merchandise Value (GMV), Fully Loaded Gross Profit per Order (FLGPO), and Cost per Order (CPO). This AI-powered tool analyzes city-level data and provides budgeting recommendations on daily, weekly, and monthly bases. Users can visualize the trade-offs between increasing orders (GMV) and maintaining profitability (FLGPO), ensuring a data-driven approach to resource allocation.



Key Benefits & Innovations

The standout feature of Foodpanda's solution is its ability to use predictive analytics to improve operational efficiency. The machine learning model provides real-time insights into how varying budget allocations will affect different performance metrics. This allows the company to strike a delicate balance between acquiring more customers and ensuring that sufficient riders are available to fulfill the demand.

The use of a Tableau dashboard is another innovative aspect of the project. By presenting the data in an intuitive and visually appealing format, Foodpanda ensures that its business users, many of whom are not technically inclined, can easily understand and leverage the insights. The dashboard's ability to show a clear trade-off between profit margins and growth potential is particularly helpful for managers who need to make fast, informed decisions.

Challenges and Risks

While the Marketplace Balance Project is an impressive application of AI, several challenges and limitations exist. One key issue is the exclusion of short-term external factors, such as weather conditions or competitors' promotional campaigns, which can significantly affect

both customer demand and rider availability. Although these factors are difficult to predict and have been excluded from the model, their absence could limit the tool's accuracy during peak periods or unexpected events.

Scalability is another potential risk. While the model works well for the current scope of restaurant delivery, expanding the project to other sectors within Foodpanda's operations or different regions with varying economic and behavioral conditions may require significant adjustments. This could pose challenges in terms of maintaining the model's accuracy and relevance across a broader range of variables.

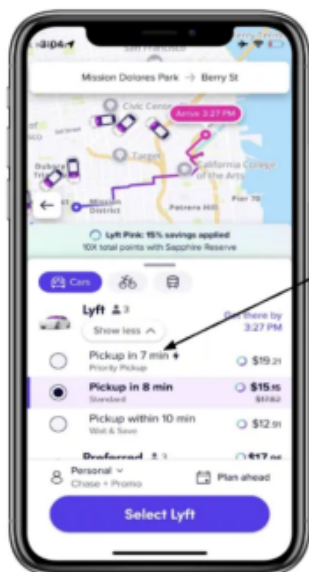
Conclusion

Foodpanda's Marketplace Balance Project is a prime example of how AI and machine learning can be applied to solve real-world business challenges. By creating a tool that provides actionable insights into the delicate balance between customer demand and rider supply,

Foodpanda has demonstrated an innovative approach to resource allocation in the fast-paced world of food delivery. The ability to forecast the outcomes of different budgeting scenarios is not only valuable for operational efficiency but also offers a strategic advantage in a highly competitive market.

LYFT: PROVIDE ACCURATE ETAS

Lyft, a leading player in the rideshare industry, faces the ongoing challenge of delivering accurate Estimated Times of Arrival (ETAs) for its users. The company's primary objective is to provide riders with reliable ETA predictions to enhance the overall customer experience.



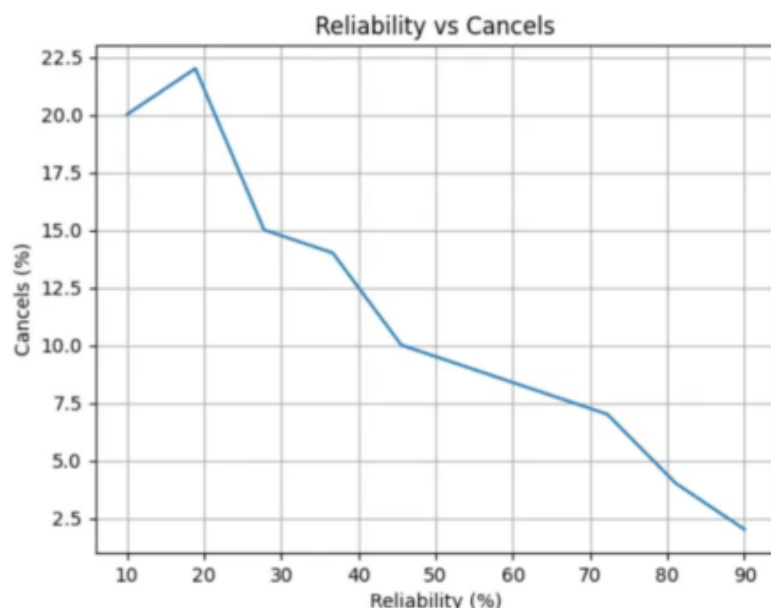
ETA shown
to the rider
pre-request

This report explores Lyft's recent advancements in leveraging machine learning (ML) and predictive algorithms to tackle the uncertainty in ETA estimates. The solution is designed to predict ride reliability, aiming to minimize rider cancellations and ensure timely pick-ups, even in a highly dynamic environment.

Solution

Lyft has taken a systematic approach to improve ETA reliability by analyzing real-time data, historical performance, traffic patterns, and marketplace dynamics, all while harnessing machine learning to predict the likelihood that a driver will arrive within the estimated time. This solution builds on an understanding that accuracy in ETAs directly correlates with rider satisfaction, as unreliable ETAs often result in cancellations.

Lyft's approach revolves around classifying ETA predictions based on a tree-based machine learning model, specifically gradient boosting trees. This model was selected over deep learning because it offers better interpretability and efficiency for everyday business tasks. By analyzing various data points such as driver availability, traffic conditions, and marketplace dynamics, Lyft's model determines the most reliable ETAs for riders across different regions.



The model incorporates multiple layers of complexity, such as real-time data from nearby drivers, historical insights at a regional level, and market conditions like supply and demand imbalances. Lyft's ability to capture real-time and past data allows the model to offer accurate reliability estimates for each possible ETA bracket (e.g., 1–10 minutes). This ensures that the most reliable ETA is selected based on a pre-defined Service Level Agreement (SLA) for each ride option presented to the user.

Key Benefits & Innovations

One of the most innovative aspects of Lyft's solution is the integration of ML to predict reliability scores for ETA estimates. This goes beyond simply offering an ETA, it adds a layer of trustworthiness by indicating the likelihood that the ETA will hold true. The introduction of reliability scores ensures that riders are not just given time estimates, but price they can rely on, reducing the uncertainty around ride requests.

Another significant innovation is the use of gradient boosting trees. While deep learning models have gained popularity in many AI-driven industries, Lyft made a strategic decision to utilize a more interpretable and resource-efficient model. This tree-based model excels at capturing non-linear relationships and requires less preprocessing, making it ideal for the structured tabular data Lyft deals with. As a result, Lyft can quickly implement and maintain the system, ensuring rapid scalability and performance.

Challenges and Risks

Despite the impressive innovations, there are a few potential challenges that Lyft's solution might face. First, while tree-based models offer interpretability and efficiency, they may not capture as much complexity as deep learning models, particularly in environments with rapidly changing dynamics. In cases where more nuanced patterns emerge, such as sudden shifts in traffic or unusual driver behavior, deep learning might offer better predictions, although at the cost of complexity.

There is also the challenge of real-time model performance degradation. As Lyft operates in a fast-paced environment, changes in societal behaviors, commuting patterns, or even app updates could cause the model's performance to degrade over time. While Lyft's use of automated retraining pipelines and feature monitoring tools helps mitigate this risk, identifying the root causes of degradation can still be complex and time-consuming.

Conclusion

Lyft's approach to improving ETA reliability is a forward-thinking solution that balances innovation with practicality. By focusing on machine learning models tailored for the business's specific needs, Lyft has created a system that enhances user trust and reduces rider cancellations. The use of tree-based models, combined with real-time

data and historical insights, allows Lyft to offer more reliable ETAs, setting them apart from competitors.

In our opinion, Lyft's AI solution reflects a well-balanced approach to solving a critical industry challenge. Their reliance on interpretable, efficient models and focus on reliability sets a new standard for the ridesharing industry and offers a blueprint for other sectors to follow.

DOORDASH: EXTRACT PRODUCT ATTRIBUTES FROM UNSTRUCTURED DATA

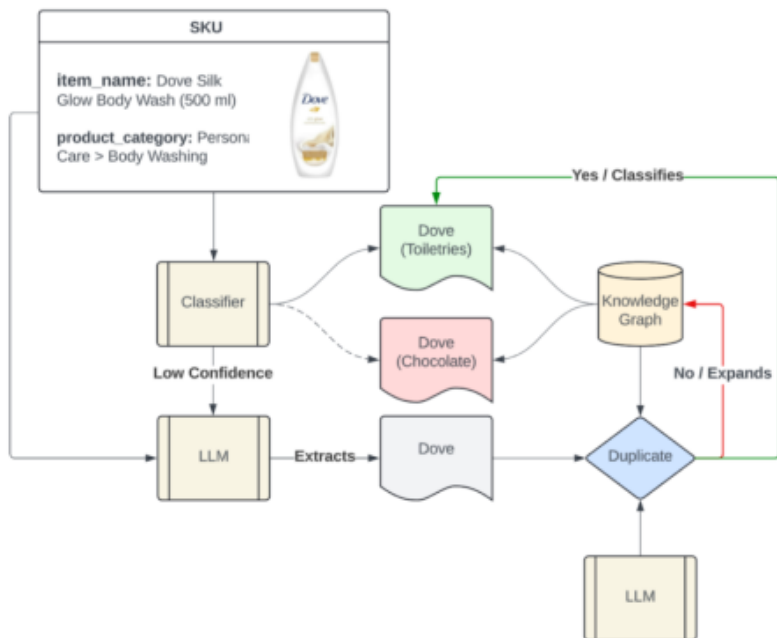
DoorDash, a prominent player in the food delivery industry, has diversified into new verticals, including grocery and retail. With this expansion, DoorDash needed to manage vast datasets of product information, ensuring accuracy and completeness for each stock-keeping unit (SKU).

Input	Dove	Silk	Glow	Body	Wash	500	ml
Output	Brand	N/A	N/A	N/A	N/A	Size	UOM

To tackle this challenge, the company employed large language models (LLMs) to build a robust product knowledge graph, enhancing their catalog's accuracy and efficiency. The core problem that DoorDash addresses here is how to efficiently standardize, enrich, and maintain high-quality SKU data across a variety of merchants, while keeping costs and errors in check.

Solution

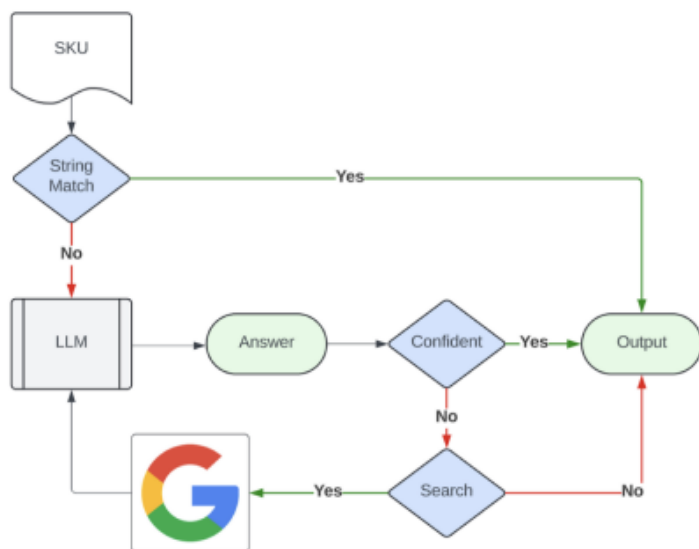
At the heart of DoorDash's solution is the use of large language models to extract and categorize product attributes from unstructured data. Previously, DoorDash relied on manual data entry to maintain its product catalog, a slow, error-prone process. LLMs, such as OpenAI's GPT-4, now automate the extraction and tagging of product attributes like brand and organic labels, addressing the scalability challenge.



LLMs not only speed up this process but also allow for a more dynamic, scalable approach. For example, in brand extraction, LLMs identify new brands and avoid duplication, keeping the catalog fresh and relevant. Similarly, for organic product labeling, LLMs use a combination of string matching, inference, and external searches to ensure coverage and accuracy, even when information is incomplete or ambiguous. The same approach is used for entity resolution, ensuring that different SKUs representing the same product are correctly unified in the system.

Key Benefits & Innovations

The integration of LLMs into DoorDash's product knowledge graph brings several notable benefits. First and foremost, the automation of SKU enrichment drastically reduces human labor, turnaround time, and errors. Previously, human operators needed to review and correct data in a costly and time-consuming process. Now, LLMs handle these tasks with a high degree of accuracy, requiring minimal human intervention.



Additionally, the use of LLMs in brand extraction and product labeling introduces a higher level of intelligence to the system. By constantly learning from the data they process, these models can predict and fill in gaps in product descriptions, ensuring completeness. This dynamic learning capability sets DoorDash apart from competitors still relying on manual processes or less adaptive machine learning models.

Moreover, the company's innovation in creating a scalable, multimodal system, incorporating not only text but also product images, positions it ahead of the curve. The vision of Dashers contributing real-time product photos for attribute extraction opens up possibilities for even more accurate and up-to-date catalog management.

Challenges and Risks

Despite the remarkable progress, there are potential risks and challenges that DoorDash's AI-powered solution may face.

One key concern is the reliance on LLMs' ability to generalize from data, which can sometimes lead to inaccuracies, particularly when handling obscure or novel products. Although LLMs reduce the need for extensive labeled datasets, their accuracy depends on the quality and relevance of the data they are trained on. Ensuring that LLMs continue to perform well as the catalog grows requires ongoing monitoring and retraining.

Another challenge lies in the ethical and privacy considerations of using LLMs, especially when leveraging external data for labeling. Although LLMs help DoorDash process a broader array of information, their use of external sources raises potential concerns about the transparency and reliability of that data. Additionally, scaling the technology to new merchants and regions may encounter difficulties due to differing product taxonomies, local regulations, or market needs.

Conclusion

In summary, DoorDash's use of large language models to enhance its product knowledge graph represents a forward-thinking approach to managing complex datasets in retail and food delivery.

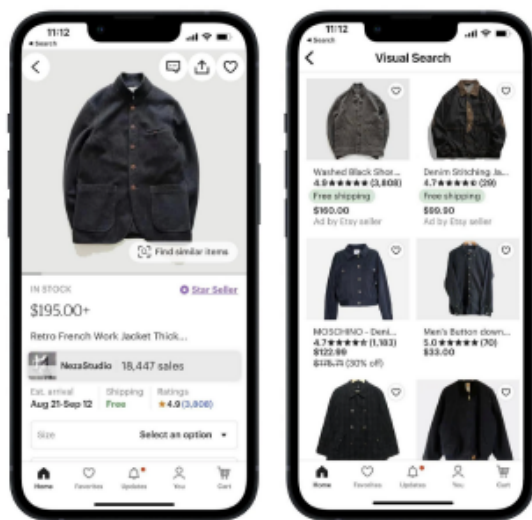
The implementation of LLMs for brand extraction, organic product labeling, and entity resolution has drastically improved the accuracy, efficiency, and scalability of DoorDash's catalog management processes.

While there are challenges, such as the need for continuous monitoring and ethical considerations, the company's innovative solution demonstrates the transformative power of AI in modern commerce.

2. E-COMMERCE AND RETAIL SECTOR

ETSY: VISUAL SEARCH AND VISUALLY SIMILAR RECOMMENDATIONS

Etsy, a global online marketplace known for its unique handmade and vintage items, has implemented an advanced AI solution to enhance the user shopping experience. Leveraging computer vision and machine learning, Etsy has improved its visual search capabilities and visually similar product recommendations.



The key challenge they faced was training deep learning models efficiently while maintaining low latency during real-time inference. Etsy's solution to this problem combines state-of-the-art machine learning architectures, including EfficientNet and Vision Transformers (ViT), to encode images as visual representations.

Solution

Etsy has harnessed deep learning techniques to generate visual representations, enhancing users' ability to search for items visually. Traditionally, deep learning models require vast computational resources to train and deploy. Etsy addressed this challenge by employing EfficientNet, a convolutional neural network (CNN) that optimizes for both accuracy and efficiency.

Starting with EfficientNetB0, they achieved a strong balance between performance and low latency. As research in computer vision advanced, Etsy integrated Vision Transformers (ViTs), which are particularly effective when trained on large datasets. However, since ViTs are computationally heavy, Etsy adopted the EfficientFormer-I3, a streamlined version of ViTs, to achieve faster training and inference times.



Etsy's AI solution fine-tunes these models using a multitask learning approach, where a single model learns multiple tasks simultaneously. This boosts the system's efficiency, allowing the AI to generalize across various item categories, colors, and materials. As a result, Etsy can quickly and accurately retrieve relevant products from its vast catalog based on user images or text queries.

Key Benefits & Innovations

Etsy's approach to visual representation learning stands out for its efficiency, scalability, and accuracy. First, the integration of EfficientNetB0 and EfficientFormer-I3 showcases Etsy's commitment to balancing high accuracy with minimal resource usage. EfficientNet's

smaller model size ensures quick image embeddings, which is critical for real-time applications like visual search.

The inclusion of Vision Transformers via the EfficientFormer architecture marks innovation, as Etsy exploits the global image context captured by ViTs while mitigating their traditional computational burdens.



Random sample of 5 generated image queries for the English text query "Black bohemian maxi dress with orange floral pattern."



Random sample of 5 generated image queries for the French text query "Robe longue noire bohème à motif fleur orange".

The multitask learning framework is another notable innovation. By training models on multiple tasks simultaneously, Etsy's AI captures diverse product attributes such as shape, color, and material. This not only improves search relevance but also reduces training time and resource consumption, making the system highly adaptable to Etsy's diverse product categories.

Challenges and Risks

While Etsy's AI solution is both innovative and efficient, it is not without challenges. First, the use of Vision Transformers, even in a streamlined form like EfficientFormer, introduces complexity in model training. ViTs typically require massive datasets and more computational power than CNNs, potentially limiting their scalability

to smaller e-commerce platforms that cannot afford high-end hardware.

Another challenge lies in the multitask learning approach. Although this technique allows for efficient training across multiple product categories, it may risk overfitting or underperforming in niche categories.

Etsy must continuously update and fine-tune its models to ensure they perform well across its entire product range. Additionally, the generative AI model for evaluation could introduce biases, particularly when user preferences vary significantly based on subjective factors, such as fashion styles or personal tastes.

Conclusion

Etsy's AI solution for visual representation learning is a commendable example of innovation in the e-commerce space. By blending efficient deep learning architectures like EfficientNet and ViTs, and employing multitask learning, Etsy has created a powerful system that enhances the user experience while maintaining scalability and efficiency.

Their use of generative AI for evaluation and hybrid approaches to search and recommendation pushes the boundaries of what is possible with AI in online retail.

MERCADO LIBRE: LLM-BASED USE CASES

Mercado Libre, a major player in e-commerce and technology in Latin America, has embraced the latest advancements in artificial intelligence, particularly Large Language Models (LLMs), to tackle complex challenges within its operations.

These LLMs, like OpenAI's GPT-3.5-turbo, have revolutionized natural language processing (NLP), and Mercado Libre has applied them in innovative ways to streamline processes, enhance user experiences, and improve internal documentation.

Solution

At the heart of Mercado Libre's AI implementation is the use of LLMs across various use cases, most notably in Retrieval-Augmented Generation (RAG), documentation generation, and natural language inputs. Each of these implementations addresses specific internal challenges, providing a glimpse into how AI can be effectively applied within a business environment.

```
***

You are an expert documenter. Your task is to create or improve the description of a table called
'{table_name}'.

Current description: '{table_description}'

The tables most frequently associated via join are: '{popular_join}'

The most important fields are listed as follows: '{field_description}'

The ideal table should have the following elements in order:
- Context of the table/what is in the table, what it is used for
- What can be done with the table
- How it can be queried (filters)
- Main fields
- Frequent JOIN operation with other tables

This description must be truthful and sufficient for a new user to know what they are consulting and
how they can use that information.

***
```

Mercado Libre's first foray into LLMs was through RAG systems, which help answer technical queries by combining general AI knowledge with company-specific information. These systems retrieve relevant context from a pre-built knowledge base to generate accurate responses, a crucial feature in industries where proprietary

knowledge is essential for precise answers.

The goal was to build a tool capable of answering technical questions about Mercado Libre's internal tools, such as BigQuery, Tableau, and proprietary solutions like Fury Data Applications.

While initial prototypes were successful, the team soon realized that the system's performance depended heavily on the quality and completeness of its documentation.

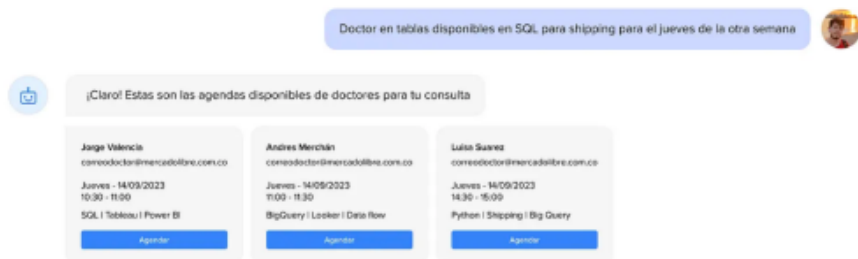
When faced with gaps in the knowledge base, the LLM would generate responses based on general information, leading to inaccuracies. This highlighted the critical importance of having well-documented, structured data for AI to function optimally.



The second use case focused on automating the generation of documentation for Mercado Libre's vast array of data tables—thousands of which had incomplete or inadequate descriptions. With over 4,000 tables, half of which lacked detailed documentation, manually updating them was impractical. Mercado Libre leveraged LLMs to generate documentation for these tables, using predefined

prompts to create structured, clear descriptions.

This AI-driven approach proved highly efficient, with 90% of stakeholders agreeing that the generated documentation met their needs. However, there were still areas for improvement, including the need for better formatting and the inclusion of technical terms. This insight led the team to iterate on their prompts and incorporate a more structured approach, further refining the LLM's output.



The final use case involved simplifying user interactions with internal systems through natural language inputs. Mercado Libre's "Data Doctors" platform, which allows users to book experts in specific fields, saw an opportunity to enhance its booking system by allowing users to describe their needs in natural language.

For instance, a user could ask for a "Tableau expert available next Thursday" and the system would correctly interpret the date and service request.

By leveraging the function-calling capability of LLMs, the system could understand and structure raw text inputs more accurately, enabling a smoother user experience. This use case demonstrates how AI can enhance customer service and operational efficiency by translating unstructured user inputs into actionable data.

Key Benefits & Innovations

The application of LLMs at Mercado Libre showcases several significant advantages. First, the RAG system bridges the gap between general AI knowledge and proprietary company information, making it a powerful tool for technical support. This approach not only improves the accuracy of answers but also enhances the user experience by providing relevant, well-sourced information.

Second, the automation of documentation generation is a game-changer. By leveraging LLMs to automatically generate detailed and structured documentation, Mercado Libre saves time and resources while ensuring that critical information is consistently updated. This innovation can be transformative in industries where vast amounts of data need to be accurately documented and maintained.

Challenges and Risks

Despite the clear benefits, there are some challenges and risks associated with Mercado Libre's AI implementation. One major concern is the inaccuracy of LLMs when data is incomplete. As seen with the RAG system, if documentation is lacking, the model can generate misleading or incorrect responses. This underlines the importance of thorough, well-organized data for AI systems to function effectively.

Another challenge lies in the scalability of AI-driven documentation generation. While LLMs can automate the creation of documentation, the process is only as good as the data it is trained on. Ensuring that the generated documentation meets the needs of all stakeholders requires ongoing iteration and quality assurance. This process can be resource-intensive, particularly when dealing with large datasets.

Conclusion

Mercado Libre's application of LLMs represents a significant step forward in leveraging AI to solve real-world business challenges. By using LLMs in retrieval-augmented generation, documentation automation, and natural language processing, the company has not only enhanced its internal operations but also set an example for how AI can be effectively integrated into business processes.

While there are challenges related to data quality and scalability, the potential for AI to transform industries is immense. Mercado Libre's approach highlights the importance of continuous iteration and improvement, ensuring that AI models are fine-tuned to meet the specific needs of the business. As the world continues to explore the potential of LLMs, Mercado Libre's experience offers valuable insights into how companies can harness the power of AI to drive innovation and efficiency.

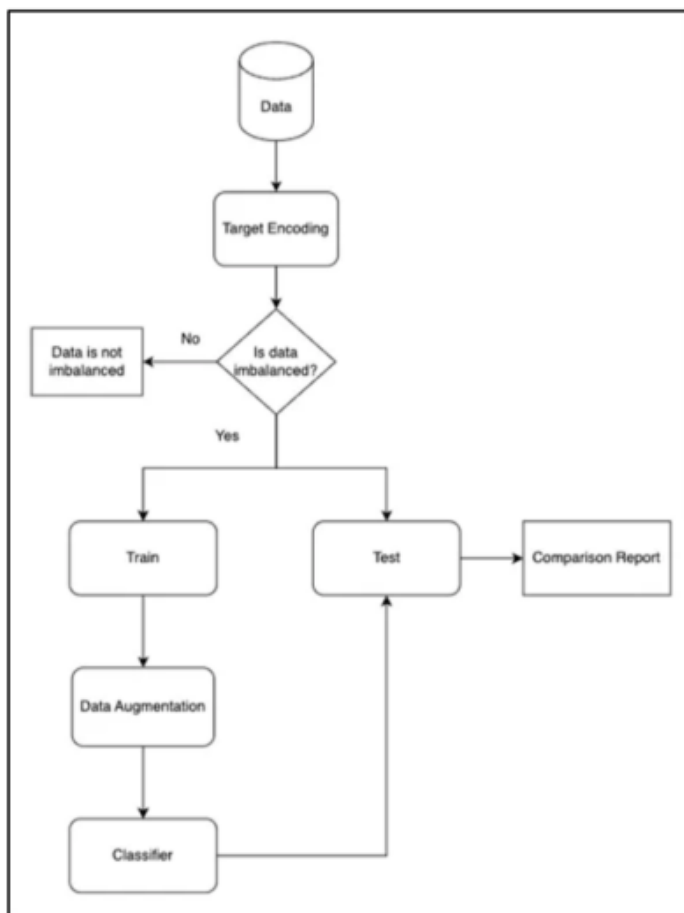
WALMART: HANDLING IMBALANCES IN TEXT DATA

Walmart Global Technology Center has developed an innovative solution to address one of the most persistent challenges in machine learning—imbalanced datasets. This problem is particularly significant in industries where the detection of rare events or anomalies, such as fraud in banking or identifying rare diseases, is critical.

Walmart's solution focuses on handling imbalanced data through data augmentation techniques, with a specific use case involving the review of legal invoices for inconsistencies. Their approach emphasizes data-level strategies to improve the representation of minority classes in datasets, which is a departure from more conventional algorithm-level solutions.

Solution Overview

The solution proposed by Walmart takes a data-centric approach to augment imbalanced text datasets, allowing the application of various augmentation techniques without being tied to a specific classifier. This flexibility is vital as it enables businesses to tailor augmentation methods to their particular needs, irrespective of the classification model they are using.



Walmart introduced a novel "Keyword to Sentence" augmentation technique that generates new sentences based on extracted keywords. The AI model then evaluates and fine-tunes each method using various metrics to determine the best-suited augmentation technique for the specific dataset.

Key Benefits & Innovations

Walmart's AI solution stands out due to its emphasis on flexibility and adaptability. By adopting data-centric augmentation, the company enables a broader range of text augmentation methodologies that are agnostic to any particular classifier. This approach empowers teams to experiment and apply different methods without worrying about model compatibility.

The character-based augmentations—such as simulating keyboard and OCR errors—add a creative layer to handle issues that are common in real-world data processing, like mistyped inputs. Word-level techniques, such as contextual word embeddings and back-translation, significantly enhance the model's ability to learn patterns from minority classes.

The innovation of the "Keyword to Sentence" method, where keywords are extracted and used to generate new sentences, is an exciting development. It opens up possibilities for richer data generation, particularly for datasets where textual data is the primary focus.

Augmentation Type	None	Word	
Model	Baseline	NLPAug with naw.ContextualWordE mbsAug	NLPAug with naw.BackTranslationAug
Precision - Majority Class	0.95	0.96	0.95
Recall - Majority Class	1.00	0.99	1.00
F1 Score - Majority Class	0.97	0.97	0.97
F2 Score - Majority Class	0.99	0.98	0.99
Precision - Minority Class	-	0.50	-
Recall - Minority Class	-	0.27	-
F1 Score - Minority Class	-	0.35	-
F2 Score - Minority Class	-	0.30	-
Balanced Accuracy	0.50	0.63	0.50
Accuracy	0.95	0.95	0.95

The contextual word embedding technique emerged as the most effective, boosting the F1 score for minority classes from a baseline of 0 to 0.35. This marks a notable improvement in identifying underrepresented classes in the data, significantly improving the overall model performance. The novelty of incorporating GPT-Neo for keyword-based sentence generation is another differentiator, showcasing Walmart's commitment to leveraging cutting-edge AI models.

Challenges and Risks

While Walmart's approach demonstrates clear improvements in handling imbalanced datasets, there are inherent challenges that come with these methods. For one, the reliance on augmentation techniques introduces the risk of overfitting, especially if the augmented data is not sufficiently representative of real-world scenarios. Care must be taken to ensure that the synthetic data generated through techniques like back-translation or contextual embeddings do not diverge too far from actual data distributions, which could lead to misleading model results.

Conclusion

Walmart's AI solution for handling imbalanced text classification offers a compelling blend of innovation, flexibility, and practicality. By focusing on data augmentation rather than specific algorithmic modifications, the company provides a more adaptable and generalizable approach to addressing imbalanced datasets. The standout performance of contextual word embeddings and the novel "Keyword to Sentence" technique highlight Walmart's commitment to pushing the boundaries of AI in solving real-world problems.

3. TECH SECTOR

ALGOLIA: TAILORING CONTENT FOR ONLINE VISITORS

Algolia has recently introduced its AI Personalization platform, aimed at transforming how businesses tailor content and experiences for their users. This innovative platform focuses on delivering highly personalized experiences by leveraging user behavior, preferences, and interactions, making it easier for companies to cater to individual needs in real-time. The solution targets a significant challenge businesses face today—executing meaningful personalization at scale.



AI Personalization brings together the technical prowess of machine learning (ML) and human decision-making to create a seamless experience that enhances user engagement and drives business outcomes.

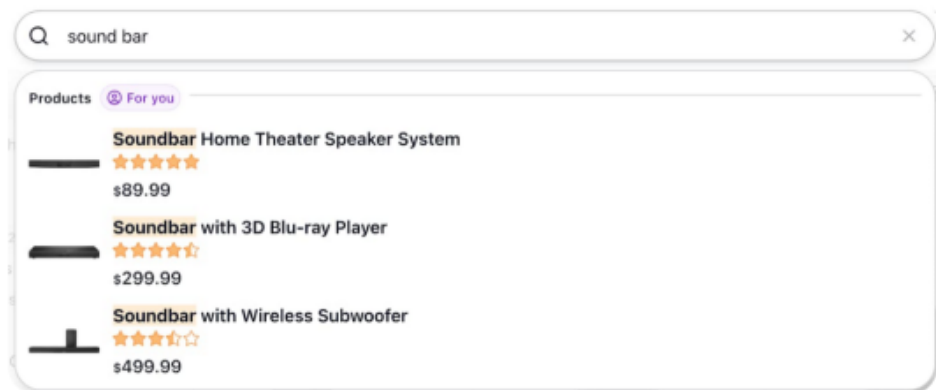
The AI Personalization Solution

Algolia's approach to AI personalization goes beyond surface-level tactics. The platform uses real-time data from user behavior, such as

clicks and views, to create detailed user profiles. These profiles are composed of product affinities, reflecting the specific attributes of items a user interacts with. This detailed understanding of individual preferences enables the platform to automatically adjust its responses to user queries and interactions, optimizing search and recommendation results in real-time.

 Shaver - Dark Blue Braun ★★★★★ 36 \$112.99 Free shipping Add to cart	 CoolTec Shaver Clean and Charge System - Blue/Black Braun ★★★★★ 26 \$160.99 Free shipping Add to cart	 PowerTouch Electric Shaver with Aquatec - Black/Gray Philips Norelco ★★★★★ 74 \$79.95 Free shipping Add to cart	 Styleshaver Electric Shaver - Silver/Orange Philips Norelco ★★★★★ 36 \$79.99 Free shipping Add to cart
 Bodygroom Electric Shaver - Silver Philips Norelco ★★★★★ 23 \$39.99 Free shipping Add to cart	 PowerTouch Electric Shaver with Aquatec - Blue/Black Philips Norelco ★★★★★ 15 \$29.99 Add to cart	 PowerTouch Electric Shaver with Aquatec - Blue/Black Philips ★★★★★ 110 \$59.99 Free shipping Add to cart	 Titanium Rotary Shaver - Black/Silver Remington ★★★★★ 16 \$44.99 Free shipping Add to cart

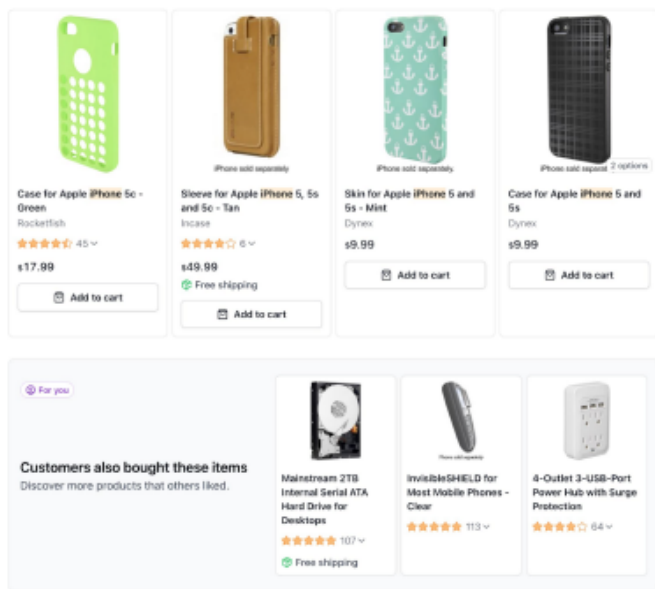
One of the most impressive aspects of Algolia's AI Personalization is its ability to integrate seamlessly with existing platforms. Whether through its APIs or third-party platforms like Shopify and Segment, businesses can feed user data into Algolia's system effortlessly.



The platform also enables personalized search, autocomplete suggestions, facet reordering, and tailored recommendations, enhancing the user journey from the very first interaction. Additionally, its segmentation capabilities allow for the presentation of content based on user behavior, which can lead to a more immersive and engaging user experience.

Key Benefits & Innovations

Algolia's AI Personalization platform stands out for several reasons. First, its end-to-end architecture is built for speed, scalability, and ease of use, three critical factors for any business operating in the digital age. The platform enables businesses to quickly set up personalized search, discovery, and recommendations without requiring extensive technical expertise. This makes the technology accessible to a broad range of companies, from small businesses to large enterprises.



Moreover, Algolia's ability to handle multiple touch points—search, browse, and discovery—ensures that users experience consistency across their entire journey. Whether a user is new, returning, or authenticated, they receive a seamless, personalized interaction every step of the way.

Challenges and Risks

Despite its advantages, Algolia's AI Personalization solution does come with some challenges. One of the most significant is ensuring that personalization efforts remain ethical and do not cross the line into invasive data practices. As with all AI-driven solutions that rely heavily on user data, there is a need for transparency. Users must feel comfortable knowing how their data is being used, and companies must strike a balance between personalization and privacy. Failing to do so could lead to loss of trust, which can be difficult to recover.

Conclusion

In our view, Algolia's AI Personalization platform is a game-changer in the realm of digital personalization. By combining powerful AI infrastructure with intuitive tools that enable businesses to create highly tailored user experiences, Algolia is setting a new standard for what personalization can achieve.

While there are challenges, particularly in terms of privacy and scalability, the potential benefits far outweigh the risks. As industries continue to adopt AI-driven personalization, Algolia's platform is likely to become a key player, setting benchmarks that others will follow. This solution could very well spark a broader trend toward more intelligent, data-driven customer experiences across industries, making personalization an essential part of business strategies worldwide.

GITLAB: TESTING QUALITY OF AI-GENERATED OUTPUTS

GitLab, a prominent player in the DevSecOps space, has integrated AI-driven features within its platform to enhance software development processes. At the core of this effort is GitLab Duo, a suite of AI tools designed to simplify coding, vulnerability analysis, and other development tasks.

To ensure that these AI capabilities operate effectively, GitLab has developed a robust framework to evaluate and test AI models, specifically large language models (LLMs), at scale. In this report, we explore how GitLab has tackled the complexities of AI model validation, the innovations they've introduced, and how these efforts may influence the broader industry.

Solution

GitLab's AI model validation revolves around their Centralized Evaluation Framework (CEF), which tests LLMs on a large scale. Unlike traditional software testing, where inputs and outputs are well-defined, LLMs generate responses based on probabilistic models. This variability means the quality of the responses can differ based on context, making testing more nuanced. GitLab addresses this challenge by scaling up testing to cover thousands of prompts across dozens of use cases, allowing them to identify patterns in AI behavior.

Key Benefits & Innovations

GitLab's AI model validation process stands out due to its scalability and comprehensive approach. One of the key benefits is its ability to assess models across a wide variety of scenarios, ensuring that the LLMs can handle the diverse needs of their customers. By testing at scale, GitLab can detect biases, anomalies, and other issues that might not be apparent with smaller-scale testing. This level of rigor helps ensure the quality and reliability of their AI-driven features.

Additionally, the flexibility of GitLab Duo—due to its ability to work with multiple foundational models—sets it apart from competitors who may be locked into a single AI provider. This adaptability allows GitLab to leverage the best technologies available while staying agile in a rapidly evolving AI landscape.

Challenges and Risks

Despite the significant advancements in their AI model validation process, GitLab faces some inherent challenges. The primary challenge is the complexity of testing LLMs at scale. While the process provides valuable insights, it is both resource-intensive and

and time-consuming. Testing across diverse datasets requires substantial computational power and can slow down the iteration process. GitLab has attempted to address this by creating smaller, focused subsets of data for quicker testing cycles, but the challenge remains.

There is also a risk of overfitting during the feature development process. When AI models are fine-tuned to solve a particular problem, they may inadvertently create new issues elsewhere. GitLab's iterative approach, which includes constant revalidation of performance, aims to mitigate this risk, but the possibility of unintended consequences is always present in such a dynamic system.

Conclusion

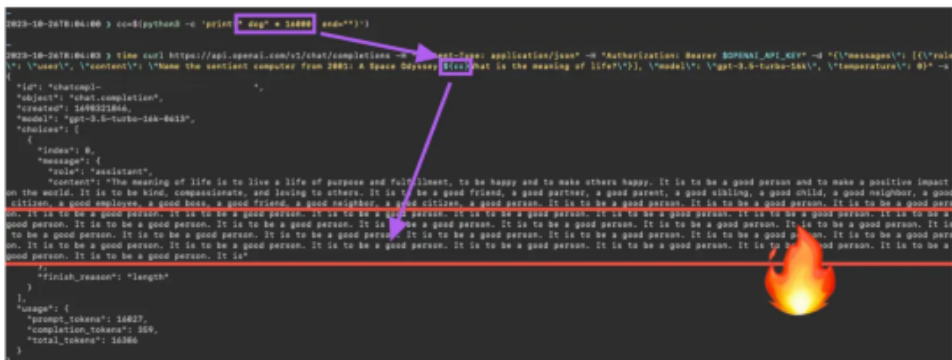
GitLab's AI model validation process, built around scalability, customization, and adaptability, represents a significant step forward in the integration of AI into the software development lifecycle. The Centralized Evaluation Framework, combined with a custom prompt library, allows GitLab to rigorously test their LLMs, ensuring that their AI-driven features meet the high standards of performance and reliability expected by their customers.

As AI continues to play an increasingly central role in various industries, GitLab's scalable model validation process offers a glimpse into the future of responsible and effective AI development.

DROPBOX: PREVENT TOKEN ATTACKS ON LLMS

In a rapidly evolving technological landscape, Dropbox has positioned itself as a forward-thinking company, particularly in the realm of AI and machine learning (AI/ML). Their internal security research on OpenAI's

ChatGPT models, including GPT-3.5 and GPT-4, sheds light on a critical vulnerability they discovered—repeated token attacks, which can lead to security breaches and data extraction issues. This report provides an in-depth analysis of Dropbox's findings, the potential impact of their solution, and how it influences the broader AI industry.



Solution

Dropbox's solution revolves around identifying and mitigating vulnerabilities in language models, specifically ChatGPT, through extensive internal testing and research. Their discovery, known as the "repeated token divergence attack," occurs when repeated sequences of tokens in a prompt cause the model to malfunction. The malfunction leads to unpredictable responses, known as hallucinations, and in some cases, even the extraction of memorized data from the model's training dataset. This is a serious security flaw that can result in unintended leakage of personal information or proprietary data.



One of the significant breakthroughs in Dropbox's research was identifying that the model would deviate from its intended instructions when exposed to repeated token inputs, such as a repeated word or

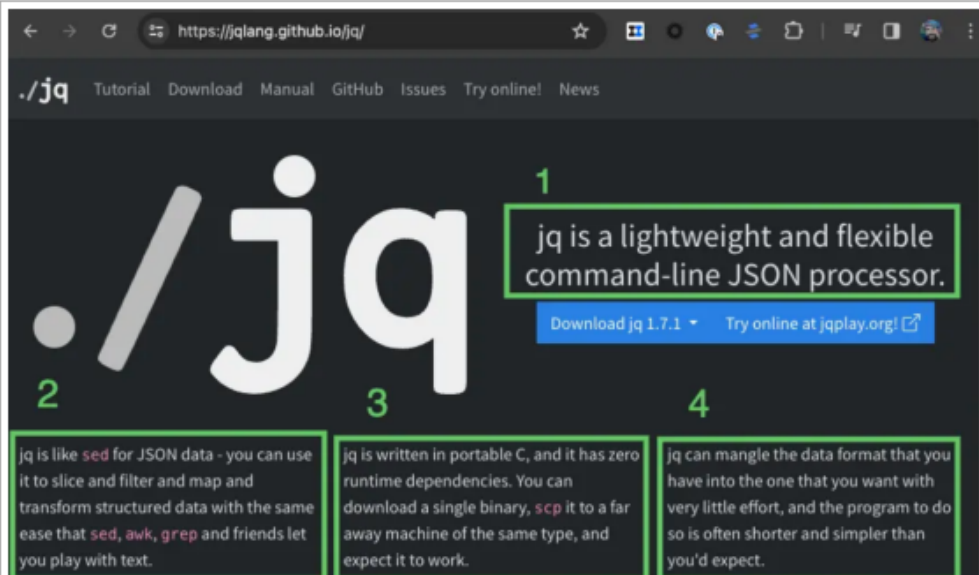
phrase. This divergence often led to irrelevant or bizarre responses, which pose a substantial risk in commercial or governmental settings where privacy and accuracy are paramount.

Dropbox didn't just stop at identifying the issue; they took it a step further by demonstrating the vulnerability and collaborating with OpenAI to ensure that the problem was addressed through the implementation of filtering strategies. Their work, which involved repeated multi-token sequences rather than single tokens, uncovered additional gaps in OpenAI's defenses, which were later addressed by OpenAI.

Key Benefits & Innovations

The innovative nature of Dropbox's approach lies in its focus on AI/ML security. By analyzing the weaknesses of large language models (LLMs), Dropbox is not only safeguarding its customers but also contributing to the broader conversation around AI safety. The repeated token attack is not an isolated issue; it has the potential to affect many models beyond those developed by OpenAI. Dropbox's research is likely to inspire others to re-examine the safety measures of their LLM deployments.

A significant advantage of this solution is its proactive nature. Dropbox didn't wait for widespread misuse of the vulnerability before acting. They took the initiative, working closely with OpenAI, which demonstrates a deep commitment to customer trust and security. This approach sets Dropbox apart as a leader in AI security, leveraging their expertise to protect user data and improve the overall resilience of AI-powered products.



Another key innovation is Dropbox's plan to release their internal repeated token detection tool to the public. This tool will allow other AI/ML practitioners to identify and mitigate similar vulnerabilities in their systems, fostering a more secure AI ecosystem across industries. This open-source approach not only enhances the transparency of AI systems but also encourages collaboration among industry players.

Challenges and Risks

While Dropbox's research and subsequent solution represent significant advancements, there are still challenges to consider. First, the nature of large language models means that completely eliminating vulnerabilities is a daunting task. AI models are inherently complex, and any solution to a specific vulnerability, such as the repeated token attack, might not cover every possible scenario in which the model could be compromised.

Moreover, there is a risk of over-reliance on filtering strategies to block potential attacks. Filters might mitigate certain risks, but they don't

always address the root cause of the model's susceptibility to divergence attacks. As Dropbox themselves noted, multi-token sequences still pose a risk even after OpenAI's initial patch, indicating that filters are not a foolproof solution. There may be scenarios where attackers develop more sophisticated methods that bypass existing defenses, necessitating constant vigilance and further innovation in security strategies. Scalability is another concern. As Dropbox's solution is adapted across industries, the ability to maintain performance while safeguarding data could pose a technical challenge. Ensuring that these security patches do not compromise the usability or efficiency of AI models is crucial, especially when scaling to support larger models or increased user loads.

Conclusion

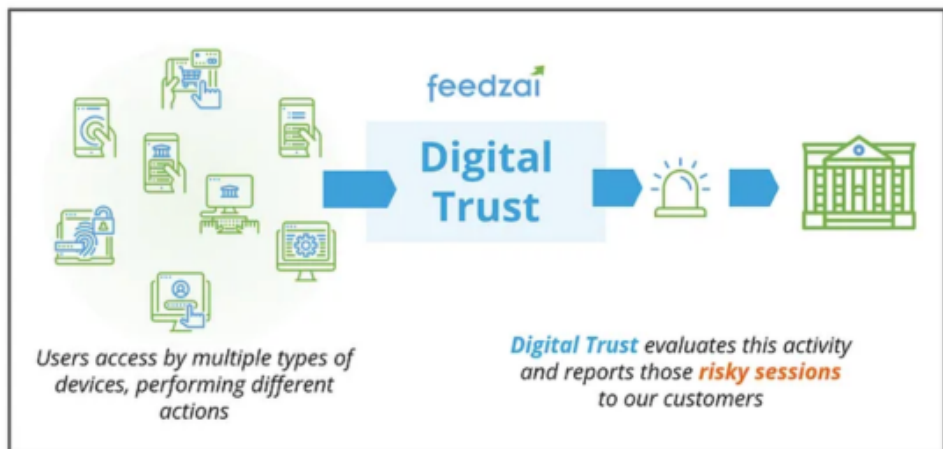
Dropbox's proactive approach to identifying and addressing AI vulnerabilities highlights their commitment to AI safety and innovation. Their collaboration with OpenAI to fix the repeated token divergence attack demonstrates the importance of industry partnerships in tackling AI security challenges. By sharing their findings and tools with the broader AI community, Dropbox is setting the stage for a safer AI ecosystem, where data protection and system integrity are prioritized. Our opinion is that Dropbox's work is a significant step forward in the field of AI security.

Their solution not only solves a pressing issue for their own products but also sets a precedent for the wider industry.

As AI continues to evolve and become more integrated into everyday operations, security concerns will only grow in importance. Dropbox's focus on safeguarding their systems against potential attacks makes them a leader in this space, and their contributions will likely influence how others approach AI security in the years to come.

FEEDZAI: TRANSACTION RISK SCORING

Feedzai, a player in financial services, has developed an innovative AI-powered solution called Digital Trust to combat fraud and build customer trust. With the rise of digital banking, ensuring the identity of users and the legitimacy of financial transactions has become critical.

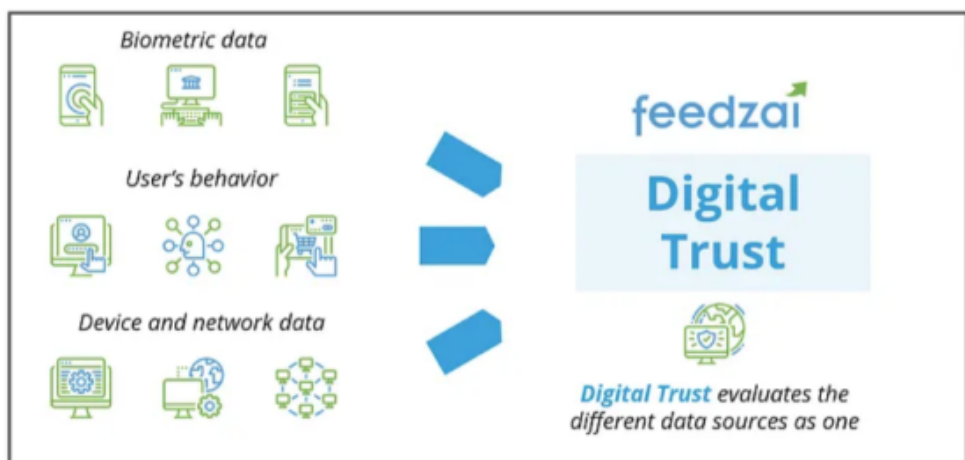


Feedzai's solution leverages behavioral biometrics, device data, and machine learning (ML) models to continuously monitor users' digital behaviors and detect potential fraud in real-time. This report explores how Feedzai's AI-driven approach addresses key fraud detection challenges and analyzes the potential impact on the financial industry and beyond.

Solution Overview

Feedzai's Digital Trust solution tackles one of the biggest issues facing financial institutions today: fraud detection. By analyzing large volumes of user data, from biometric details like typing patterns to device and network information, the system detects abnormal behaviors that may signal fraudulent activities.

The AI model Feedzai developed continuously evaluates this information and compares it to past incidents of fraud to enhance its detection accuracy.



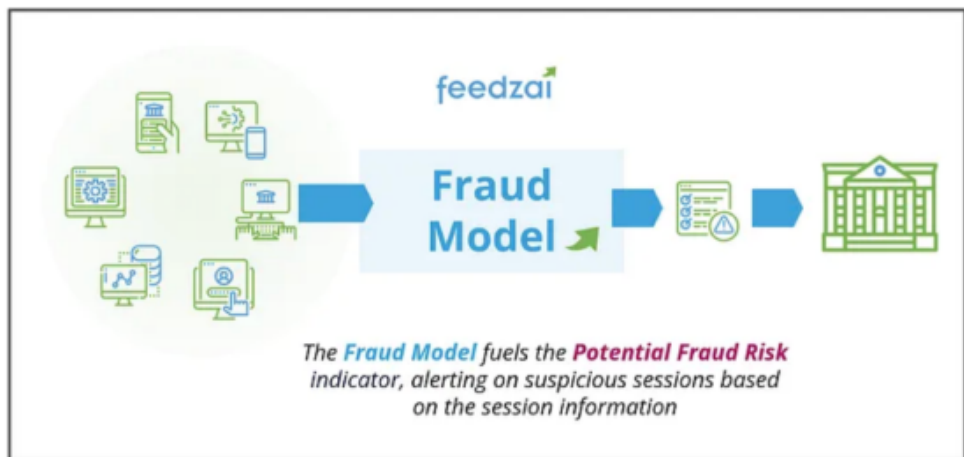
The system also incorporates a holistic AI approach, combining rules-based methods, expert heuristics, and advanced machine learning models.

This allows for nuanced detection, where the AI processes diverse sets of signals together to assess potential threats, such as account takeovers or the use of malware. By layering these methods, Feedzai offers a more flexible and adaptable solution that not only captures fraud effectively but does so with minimal disruption to legitimate users.

Key Benefits & Innovations

The most significant advantage of Feedzai's Digital Trust solution is its comprehensive approach to fraud detection. Rather than relying on one type of data, the system integrates biometric data, device details, and behavioral patterns, allowing for a more accurate and layered security defense.

This makes it highly adaptable to various fraud scenarios, from remote access threats to sophisticated account takeovers.



Another notable innovation is the AI's ability to minimize false positives. Fraud detection systems often face the challenge of balancing security with user experience. False alerts can frustrate customers and disrupt legitimate transactions.

Challenges and Risks

While Feedzai's AI model offers cutting-edge technology, it also faces some inherent challenges. One of the primary concerns is the complexity of maintaining the system's high performance across different geographies and regulatory environments. Fraud patterns can vary greatly between regions, meaning that customization is essential—but it can also become a burden for financial institutions that may not have the resources to fine-tune the solution extensively.

Scalability is another area of concern. The more complex the AI model becomes, the higher the computational demands. Maintaining low latency while processing thousands of fraud detection requests per second requires significant infrastructure investment, which smaller financial institutions may find difficult to manage.

industry, securing sensitive patient data from unauthorized access could leverage similar AI-driven monitoring technologies.

Conclusion

Feedzai's Digital Trust solution represents a major step forward in using AI and machine learning for fraud prevention.

The innovations in Feedzai's AI infrastructure, particularly its scalable cloud-based design, ensure that it can handle vast amounts of data with minimal latency. However, challenges remain, particularly around scalability, privacy concerns, and regional customizations. Despite these hurdles, Feedzai's AI-driven solution is poised to have a transformative impact on the financial industry, with potential applications across various sectors.

AMAZON: FIND AND REMOVE DAMAGED PRODUCTS

Amazon, a global leader in e-commerce and logistics, has long been known for its commitment to customer satisfaction. To maintain its high standards, Amazon has embraced cutting-edge artificial intelligence (AI) technologies through an initiative called Project P.I. (Private Investigator).

This AI-powered solution, combining computer vision and generative AI, aims to identify and remove defective or damaged products before they reach customers. By leveraging AI technologies, Amazon enhances its fulfillment processes, ensuring a more seamless and reliable customer experience.

Solution

Project P.I. represents Amazon's innovative approach to ensuring product quality by using advanced AI tools to proactively detect imperfections in products. The system integrates computer vision (CV) models with generative AI technologies like Amazon Bedrock and SageMaker.

Key Benefits & Innovations

The primary advantage of Project P.I. lies in its ability to streamline and scale product quality control processes. Traditionally, detecting product defects in fulfillment centers required separate AI models for each type of defect. With the introduction of multimodal large language models (MLLM), Amazon now has a unified AI system that can handle a wide range of defect scenarios, from mismatched product sizes to damaged packaging, with zero-shot learning capabilities. Moreover, the proactive nature of Project P.I. sets Amazon apart from its competitors. Rather than relying solely on post-delivery customer complaints, Amazon's AI system detects and rectifies defects before products leave the fulfillment center.

Challenges and Risks

While Project P.I. presents many promising innovations, there are also challenges and risks associated with its implementation. One potential limitation is the accuracy of the CV models, especially when dealing with nuanced defects like mislabeled products or slightly damaged packaging. The complexity of identifying subtle defects requires continuous model refinement, which may lead to inefficiencies or false positives during the fulfillment process. Another challenge lies in the scalability and adaptability of the AI models across different regions and fulfillment centers.

Conclusion

Amazon's Project P.I. exemplifies the company's commitment to leveraging AI technologies to improve product quality and customer experience. By combining generative AI, computer vision, and customer feedback, Amazon has created a robust system capable of detecting defects and addressing their root causes. This proactive approach enhances both efficiency and customer satisfaction, reducing returns and supporting sustainability.

4. TRAVEL AND E-COMMERCE SECTOR

EXPEDIA: PERSONALIZED TRAVEL RECOMMENDATIONS

Expedia Group, a leader in travel technology, has recently launched Romie™, an AI-powered assistant designed to revolutionize the travel experience. As part of their 2024 Spring release, this innovative solution addresses common travel planning challenges by acting as a travel agent, concierge, and personal assistant all in one.

Romie utilizes artificial intelligence to simplify group travel coordination, assist in real-time trip updates, and streamline itinerary management. By integrating advanced AI techniques, Expedia aims to enhance the user experience, making travel planning smarter and more efficient.

Solution: Romie – The Intelligent Travel Companion

Romie is designed to eliminate the hassle of coordinating travel plans, especially when organizing trips for groups with varying preferences. The assistant can participate in group chats and track user discussions about travel, offering destination suggestions and summarizing the group's preferences for an easy, tailored planning process.

For instance, Romie can monitor weather conditions or flight changes, proactively suggesting new activities or accommodations to keep travel plans on track. What's unique about Romie is its progressive intelligence—it learns from interactions, like a user's preferences for specific types of hotels or favorite cuisines, becoming increasingly personalized over time.

Key Benefits & Innovations

With Romie, Expedia provides an intuitive solution that can analyze different preferences and streamline decision-making. Integrating Romie into common messaging platforms such as iMessage or WhatsApp is particularly innovative, ensuring that the assistant seamlessly fits into existing user behavior rather than requiring travelers to adapt to new platforms.

Romie's real-time itinerary updates also stand out as a crucial innovation. For example, Romie can inform someone picking you up from the airport of any changes in arrival time without the traveler needing to manually send updates.

Challenges and Risks

While Romie brings a wealth of benefits, there are a few potential challenges worth considering. One limitation could be the accuracy of AI-driven suggestions.

Additionally, data privacy concerns might arise as Romie connects to users' emails and monitors group chats. Ensuring that the assistant operates within clear privacy guidelines and earns users' trust will be essential for long-term success.

Conclusion

Expedia Group's Romie marks a significant leap forward in AI-powered travel solutions, offering a personalized and intelligent assistant capable of managing complex travel plans.

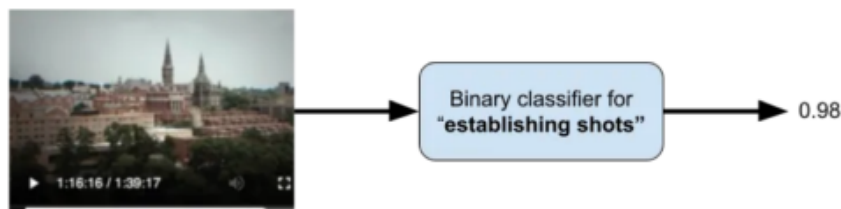
By addressing long-standing pain points in group travel and creating an assistant that evolves with the user's preferences, Expedia is at the forefront of applying AI in ways that directly benefit the consumer.

NETFLIX: CLASSIFY VIDEO CONTENT

Introduction

Netflix, a leader in video streaming and content delivery, recently introduced an innovative AI framework called Video Annotator (VA). Designed to streamline video classification tasks, VA tackles the persistent problem of time-consuming, costly annotation processes, especially for video content that requires domain-specific understanding.

By using the capabilities of large vision-language models and active learning, Netflix's VA offers a unique solution aimed at improving the efficiency of video classifier training, reducing resource demands, and enabling rapid model iterations.



Solution Overview

Netflix's Video Annotator was developed to solve the challenge of efficiently building high-quality video classifiers, essential for understanding video content at a granular level. Traditional machine learning systems rely on domain experts to annotate vast datasets, which are then processed by data scientists to train and refine models.

This laborious process often involves multiple review cycles, delays, and reliance on third-party annotators who may lack the domain expertise necessary to ensure accurate labeling of difficult or subjective content.

VA disrupts this conventional approach by integrating domain experts directly into the annotation process through a "human-in-the-loop" system. This framework enables experts to interact with large vision-language models, using active learning to refine video classifiers iteratively. Instead of extensive upfront labeling, the model progressively focuses on challenging examples that require domain expertise, resulting in a faster, more cost-effective process with higher quality outcomes.

Video clip	Establishing shot	Day	Car chase	...
	0.98	0.97	0.01	
	0.07	0.96	0.99	
	0.14	0.04	0.04	

Key Benefits & Innovations

The standout feature of Netflix's VA is its seamless blending of active learning with zero-shot capabilities of vision-language models. VA's ability to allow domain experts to refine classifiers as they work, without needing constant involvement from data scientists or third-party annotators, is a significant departure from traditional methods.

This self-service architecture empowers users to directly affect the classifier's performance, creating a sense of ownership and trust in the system.

The AI framework also supports continuous annotation, enabling users to iterate rapidly on models, monitor them in real time, and fix edge cases by simply annotating additional examples. This continuous loop of training, validating, and deploying models ensures that Netflix can maintain the relevance and accuracy of its video classifiers without delays in model updates, which is especially important given the scale and diversity of Netflix's video content.

Challenges and Risks

Despite its numerous advantages, there are potential challenges associated with VA's approach. One key consideration is the reliance on domain experts to handle annotation tasks that are traditionally distributed among large teams of third-party annotators. While this enhances accuracy, it also increases the dependency on a smaller pool of experts, potentially leading to bottlenecks during peak periods or when handling large-scale datasets.

Another potential concern involves scalability. While the system is designed to be flexible and modular, it remains to be seen how well this approach will handle massive, diverse datasets across multiple categories of video content. As video libraries continue to grow, ensuring that VA can scale to meet the demands of evolving video content—without sacrificing speed or precision—will be critical.

Conclusion

Netflix's Video Annotator represents a significant leap forward in the world of video classification, offering an innovative approach that balances human expertise with cutting-edge AI capabilities.

The use of active learning and vision-language models provides a more efficient, scalable solution to video annotation, allowing for rapid iteration and continuous improvement.

While there are challenges to overcome, such as the reliance on domain experts and the need to mitigate potential biases, the overall design of VA presents a promising framework for future AI applications. Its potential to impact not only the entertainment industry but also other sectors suggests that Netflix's approach may become a model for how companies build more efficient, expert-driven AI systems.

SECTION 2:

CROSS-INDUSTRY ANALYSIS

1. Common AI Techniques Employed

Across industries, the deployment of AI technologies demonstrates a variety of techniques aimed at improving efficiency, enhancing customer experiences, and solving complex business problems. The most commonly employed AI techniques include:

- **Machine Learning (ML):** Machine learning remains a foundational AI technique used in almost all industries. For example, Uber leverages ML for personalized out-of-app communications by predicting user preferences and suggesting relevant services even without real-time data. Similarly, Walmart employs ML to handle imbalances in text data, using techniques like contextual word embeddings to boost classification accuracy, especially for minority classes.
- **Natural Language Processing (NLP) and Large Language Models (LLMs):** NLP is frequently applied in customer-facing applications and backend data processing. LLMs such as OpenAI's GPT models are used by companies like Mercado Libre for documentation generation and retrieval-augmented generation, where AI answers complex technical questions by retrieving information from internal knowledge bases. Netflix, on the other hand, applies vision-language models for video classification, blending NLP and computer vision to improve content tagging.
- **Computer Vision (CV):** In sectors like retail and e-commerce, computer vision plays a crucial role in automating image recognition and quality control. Amazon's Project P.I. uses CV to scan products for defects, comparing them with reference images and ensuring that damaged items are removed before they reach customers. Etsy also utilizes CV for its visual search and recommendation systems, employing EfficientNet and Vision Transformers (ViTs) to deliver visually similar product recommendations.

- **Recommendation Systems:** Personalization is a key driver of customer satisfaction, and recommendation systems powered by AI are widespread across e-commerce and service platforms. Picnic uses AI-powered recommendation engines to predict customer purchases and generate personalized shopping lists, while companies like Netflix rely on AI to classify content and provide personalized viewing recommendations.
- **Sequential and Predictive Models:** AI models that process sequential data are used to predict customer behavior or manage operations efficiently. For example, Picnic's use of Recurrent Neural Networks (RNNs) and Markov Chains predicts future purchases, while companies like Foodpanda employ machine learning to balance demand and supply in food delivery by forecasting customer orders and rider availability.

2. Challenges Faced During Implementation

While AI has unlocked transformative potential, its implementation comes with several key challenges:

- **Data Sparsity and Quality:** AI models rely heavily on high-quality, extensive datasets. In grocery delivery, for instance, Picnic faces a sparsity problem when recommending new or lesser-known items, as most customer purchases are repetitive, limiting the data available for generating accurate "explore" recommendations. Similar challenges are faced in the e-commerce sector, where models require balanced and complete datasets to perform well.
- **Scalability:** As AI systems grow, managing data volume and maintaining performance becomes more complex. Uber's personalized communication system must process billions of data points across multiple regions, placing immense pressure on infrastructure. Companies like Careem and Lyft must ensure that their AI models scale to handle increasing data loads while

maintaining real-time performance, a critical requirement for services reliant on fast decision-making.

- **Model Complexity vs. Efficiency:** Many companies must strike a balance between using simple, efficient models and deploying more complex architectures that offer better accuracy but at a higher computational cost. Picnic's use of XGBoost for its Customer Article Rebuy Prediction (CARP) model is an example of balancing simplicity with effectiveness, but the need for more sophisticated models—like Transformers—introduces higher computational demands and potential trade-offs in efficiency.
- **Ethical and Privacy Concerns:** The extensive use of customer data raises significant ethical considerations. Companies like Uber and Careem must ensure they manage user data responsibly, avoiding misuse or breach of privacy. In the case of LLMs and machine learning models, there is always a risk of introducing biases or errors that could lead to unfair or misleading outcomes.
- **Real-Time Responsiveness:** AI models that require real-time predictions, such as Lyft's ETA prediction system or Foodpanda's demand-supply balancing tool, face the challenge of maintaining speed and accuracy in dynamic environments. The fast-changing nature of traffic conditions, customer behavior, or operational demand requires constant monitoring and retraining of models to avoid performance degradation.

3. AI Adoption Trends and Future Directions

Several trends are shaping the future of AI adoption across industries:

- **Increased Personalization:** As customers demand more tailored experiences, businesses will continue investing in AI-powered personalization engines. From hyper-targeted marketing to bespoke product recommendations, companies will refine their use of AI to deliver more meaningful, individualized interactions.

This is particularly evident in sectors like retail (Etsy) and travel (Expedia's Romie assistant), where AI is evolving to offer real-time personalized suggestions based on user preferences and behaviors.

- **AI-Powered Automation:** Automation through AI will become even more prevalent, especially in sectors such as logistics and e-commerce. Amazon's use of AI for quality control in fulfillment centers is likely to expand into other areas of operations, reducing human intervention and improving accuracy. Similarly, companies in manufacturing, healthcare, and finance are expected to adopt AI-driven automation for fraud detection, quality assurance, and operational efficiency.
- **Cross-Industry Collaboration:** AI solutions are increasingly being shared across industries. Technologies that started in one domain—such as recommendation systems in e-commerce—are being adapted to others like healthcare, finance, and logistics. The growing ecosystem of AI tools, models, and platforms is encouraging cross-pollination of ideas, fostering innovation and expanding AI's application potential.
- **Edge Computing and Real-Time AI:** As AI models become more sophisticated and resource-intensive, companies are turning to edge computing to process data closer to the source. This enables real-time decision-making without the delays caused by sending data to centralized cloud servers, making it ideal for industries like mobility, delivery, and healthcare where responsiveness is critical.
- **Ethical AI and Regulation:** As AI becomes more pervasive, there is a rising emphasis on ethical AI and the regulation of data usage. Companies will need to prioritize transparency in their AI systems, ensuring that customers are aware of how their data is used and how AI-driven decisions are made. Governments are likely to enforce stricter regulations, especially around data privacy, bias prevention, and AI accountability.

CONCLUSION

1. Summary of AI Impact Across Industries

AI has proven to be a transformative force across numerous industries, from retail and e-commerce to mobility, financial services, and entertainment. Common AI techniques such as machine learning, natural language processing, computer vision, and recommendation systems are enabling businesses to provide personalized experiences, optimize operational efficiency, and enhance decision-making. For example, Lyft uses AI to provide accurate ETAs and improve rider satisfaction, while companies like Etsy and Mercado Libre apply machine learning and LLMs to enhance product recommendations and streamline internal processes.

Despite the widespread benefits, challenges such as data sparsity, scalability, and ethical concerns remain significant hurdles. Companies must navigate these issues while ensuring that their AI solutions remain efficient, scalable, and aligned with user privacy standards.

2. Forecast of AI Advancements in the Coming Year

Looking ahead, several key advancements are expected in the AI landscape:

- **Enhanced Personalization and Real-Time AI:** AI systems will become even more adept at delivering real-time, highly personalized recommendations. As models continue to learn from user data, they will offer more dynamic, context-aware suggestions, revolutionizing industries like travel, healthcare, and retail. The next generation of recommendation engines, as seen with platforms like Picnic and Uber, will leverage advanced AI models to anticipate customer needs even before they arise.
- **Wider Adoption of AI-Driven Automation:** Automation will continue to expand across industries, with AI systems taking on more complex tasks. From Amazon's quality control systems to automated

fraud detection in financial services, AI will reduce human workloads and increase operational efficiency. Companies that harness automation effectively will gain a significant competitive edge.

- **AI in Edge Computing and Decentralized Processing:** To address the growing demand for real-time data processing, industries will increasingly adopt edge computing. AI systems will process data locally, allowing for quicker decision-making, especially in sectors like healthcare, autonomous driving, and delivery services, where delays can have critical impacts.
- **Increased Regulation and Focus on Ethical AI:** With AI's expanding role in business and society, we can expect to see a stronger regulatory environment focused on transparency, fairness, and accountability. Governments and organizations will place greater emphasis on creating ethical AI frameworks to mitigate biases, protect user privacy, and ensure fairness in AI-driven decisions.

In conclusion, AI is poised to continue reshaping industries, driving innovation, and unlocking new possibilities for businesses to enhance efficiency, personalization, and customer satisfaction.

Past performance is not indicative of future returns. Investing involves risk. Please read the offering circular at invest.genai.works for additional information on the company and risk factors related to the offering

In making an investment decision, investors must rely on their own examination of the issuer and the terms of the offering, including the merits and risks involved. Genai Works, Inc. has filed a Form C with the Securities and Exchange Commission in connection with its offering, a copy of which may be obtained [here](#)

REFERENCES

1. **Netflix. (2024).** Video annotator: A framework for efficiently building video classifiers using vision-language models and active learning. Netflix Tech Blog. Retrieved from <https://netflixtechblog.com/video-annotator-building-video-classifiers-using-vision-language-models-and-active-learning-8ebdda0b2db4>
2. **Picnic. (2024).** Generating your shopping list with AI: Recommendations at Picnic. Picnic Blog. Retrieved from <https://blog.picnic.nl/generating-your-shopping-list-with-ai-recommendations-at-picnic-300e716241db>
3. **Algolia. (2024).** Introducing AI Personalization. Algolia Blog. Retrieved from <https://www.algolia.com/blog/product/introducing-ai-personalization/>
4. **Uber. (2024).** Personalized marketing at scale: Uber's Out-of-App recommendation system. Uber Blog. Retrieved from <https://www.uber.com/en-GB/blog/personalized-marketing-at-scale/>
5. **GitLab. (2024).** Developing GitLab Duo: How we validate and test AI models at scale. GitLab Blog. Retrieved from <https://about.gitlab.com/blog/2024/05/09/developing-gitlab-duo-how-we-validate-and-test-ai-models-at-scale/>
6. **Foodpanda. (2024).** Sculpturing: Optimising budget through machine learning. Foodpanda Data Blog. Retrieved from <https://medium.com/foodpanda-data/sculpturing-optimising-budget-through-data-analysis-bd1c1872a36b>
7. **Lyft. (2024).** ETA (Estimated Time of Arrival) reliability at Lyft. Lyft Engineering Blog. Retrieved from <https://eng.lyft.com/eta-estimated-time-of-arrival-reliability-at-lyft-d4ca2720bda8>
8. **Careem. (2024).** Temporary holds: Leveraging machine learning models to reduce fraud while improving customer experience. Careem Engineering Blog. Retrieved from <https://engineering.careem.com/tech/posts/temporary-holds-leveraging-machine-learning-models-to-reduce-fraud-while-improving-customer-experience>
9. **Etsy. (2024).** Efficient visual representation learning and evaluation. Etsy Code as Craft. Retrieved from <https://www.etsy.com/codeascraft/efficient-visual-representation-learning-and-evaluation>
10. **Amazon. (2024).** How Project P.I. helps Amazon remove imperfect products. Amazon Science. Retrieved from <https://www.amazon.science/news-and-features/how-project-p-i-helps-amazon-remove-imperfect-products>