International Journal of Advanced Artificial Intelligence Research (IJAAIR)
eISSN: 3087-4270
Volume03 Issue11 | Page No: - 49-64

Received: 09 Oct 2026, Revised: 23 Oct 2026, Accepted: 03 Nov 2026, Published: 12 Nov 2026

DOI: https://doi.org/10.55640/ijaair-v03i11-04

Building Explainable Sales Intelligence Systems: A Product-Oriented Model for Embedding AI Insights into Enterprise Sales Workflows

Aakash Nagare
Independent Researcher, USA

ABSTRACT: While Enterprise sales systems have dashboards to track revenue, customers, and buying trends, reports often require reading between the lines to determine what's good and what's bad before deciding which accounts to focus on. In this research, an explainable Sales Intelligence framework is proposed that transforms transaction-level information into analytical evidence at the customer level and converts it into conversations to support decision-making. The following is an example of the framework: the UCI Online Retail dataset consists of 541,909 online retail transactions from a UK retail company from 1 December 2010 to 9 December 2011. Data preparation removes missing customer identifiers, invalid quantity/unit price data, and expired entries. This results in 397884 transaction lines, 18532 invoices, 4338 identifiable customers, 3665 products, 37 countries, and the following actual retained revenue (facts as of 31 December 2018): £8,911,407.90. Customer revenue, product activity, average order value, purchase frequency, and recency are calculated using deterministic methods to segment and assess account health. The proposed architecture leverages numerical computation and evidence retrieval to interpret AI-generated results and connect explanations/recommendations to verified transaction indicators, with an emphasis on algorithms. The proposed architecture interprets artificial intelligence results and connects evidence/messages and/or recommendations to certified transaction indicators, with an emphasis on algorithms. The study provides empirical results from the transaction analysis process. It proposes definitions of measures to evaluate an agent's future operational application, including numerical accuracy, retrieval precision and recall, F1 score, recommendation relevance, explanation agreement, error rate, and response latency. Validate data from the enterprise CRM and sales-interaction data (recommended).

KEYWORDS: Explainable AI; Sales Intelligence; Customer-Level Analytics; Conversational AI; AI-Augmented Decision Support.

1. Introduction

Organizations have moved from spreadsheets and hand-held customer records to Customer Relationship Management (CRM), Business Intelligence (BI), predictive analytics, and Artificial Intelligence (AI), and now it's impossible to imagine a business functioning without them. Spreadsheets offered simple analytical functions but were time-consuming and required regular reporting. With CRM and BI solutions, businesses gained access to sales data through centralized data stores, dashboards, and interactive reports. Additionally, predictive analytics added more functionality by enabling customer grouping, buying pattern analysis, demand forecasting, and behavioral change identification. More recently, AI has expanded sales analysis beyond traditional reporting by providing interactive analysis and decision-making capabilities in enterprise systems. Despite these achievements, conventional sales-intelligence systems cannot offer useful interpretation at the customer level. Dashboards typically display set criteria that are dependent on other factors like revenue, purchase frequency, and purchase recency. For instance, a drop in sales can result from lower transaction amounts, fewer transactions, or shifts in which products are sold. This can require users to use several distinct reports, apply filters, and make other calculations. While transactions, or sales data, can show customers who bought what and when, they can't necessarily tell whether or when they bought for the first time, whether they also purchased something

else, which particular product they purchased, or whether a company employee made the sale. This is an opportunity for easy change offered by the AI-powered Sales intelligence, with natural-language interpretation and conversational access to analytical data. These capabilities, however, require analytical grounding and both numerical precision and "explainability." If the user cannot, or does not know how, to verify the "why" behind the recommendation, it can be hard to determine whether it is correct or applicable. If possible, separate the math behind customer indicators from the analysis of those indicators. Numerical results should come from repeatable analytical processes, and the AI element should retrieve and interpret verified evidence rather than creating customer metrics on its own.

This research proposes an explainable framework that identifies product orientation during sales and maps it to customer-level analytical evidence and customer-level sales intelligence. We used the UCI Online Retail data set, which contains transactions from an online retailer from 1 December 2010 to 9 December 2011 in the United Kingdom. The data set includes invoice number, product, quantity, transaction date, unit price, customer, and country information. These variables can be used to derive customer metrics such as customer revenue, purchase frequency, recency, average order value, and product activity. The generated indicators enable analysis of buying behavior, such as drop-off, lapsed customers, inactivity, and opportunities to win back customers. The study does not cover enterprise-wide CRM or B2B sales management capabilities. It focuses on customer intelligence, particularly transaction-based. The data set does not include CRM notes, sales calls, opportunity records, pipeline stages, or detailed histories of customer interactions. For this reason, the proposed framework is tested conceptually and analytically using evidence from transactions, not as an operationally mature, empirically proven operational AI agent. The AI element is introduced as a key driver to provide a complementary interpretation and conversational interface, which could serve as a channel for communicating analysis results and offering evidence-based sales recommendations. The key contribution is the ability to derive deterministic customer insight, leverage AI interpretation and explainability to guide conversational interactions, and do all of this within one product-centric sales intelligence solution. The architecture separates numerical calculations from language translation into customer values so that customer values can be traced back to the underlying transaction evidence. AI-generated explanations & recommendations can be viewed as decision support outputs that must be suitable for human review as part of an enterprise flow.

1.1 Aim of the Study

The aim of this study is to propose an explainable, product-oriented sales-intelligence framework that transforms transaction-level data into customer-level analytical evidence and supports evidence-grounded conversational decision-making.

1.2 Objectives of the Study

The specific objectives are:

(1) To develop deterministic customer-level analytical measures, including revenue, purchase frequency, recency, average order value and product activity, using transaction-level data. (2)To examine customer purchasing behaviour through segmentation, account-health indicators and the identification of high-value, declining, inactive and potential reactivation accounts. (3) To propose an evidence-grounded AI retrieval and conversational architecture that connects customer-level analytical results to natural-language explanations and sales recommendations. (4) To establish an evaluation framework for assessing numerical accuracy, retrieval precision, recall and F1-score, recommendation relevance, explanation agreement, error rate and response latency. (5) To examine the applicability and limitations of the proposed framework and identify requirements for future validation using enterprise CRM, sales-pipeline and customer-interaction data.

1.3 Scope of the Study

The study focuses solely on case studies of the UCI Online Retail dataset and proposes an Explainable Sales-Intelligence model framework based only on transactions. Researched factors include customer revenue, purchase frequency, customer recency, average order value, product activity, and indicators of purchasing behavior. These measures form the foundation for analyzing customer segments, customer health, and the potential to reactivate customers. The study also considers an AI-driven retrieval and conversational interface that communicates analysis results to users. Unlike the space-time-complicated

workflows of natural-language interpretation and deterministic data processing, the proposed architecture separates evidence retrieval from data interpretation. The study doesn't mention that a fully developed AI sales agent has been implemented or run. The evaluation criteria for a future implementation do not differ from those mentioned, and the empirical component covers only transaction data and customer-level analysis. The data comes from an online retail context and does not include all aspects of enterprise sales activity. In particular, it lacks many CRM interactions, sales team activity, opportunity management, contractual details, and customer sentiment. The findings reflect the analytical strengths of transaction-based customer intelligence and do not necessarily indicate the solution's performance in all enterprise selling situations.

1.4 Significance of the Study

The paper offers valuable insights on sales intelligence, providing a structured, systematic framework for annotating any sales transaction and converting it into meaningful customer-level sales intelligence. The proposed framework takes a two-stage approach that separates numerical computation from the generation of the interpretation using AI so that linkages between the received recommendation and measurable manifestations of customer behavior can be facilitated more easily. This matters because it enhances the transparency and verifiability of AI use in sales decisions. The framework illustrates shifts in consumer buying behavior by examining purchase frequency, recent purchase volume, and revenue. It also helps communicate with users without forcing them to wade through dozens of dashboards and reports to get the results they want. The proposed research study offers an architecture and evaluation framework to inform future research implementations. Taking it further with enterprise CRM data, sales-pipeline details, and customer-interaction information would clearly confirm that it can be applied more broadly than transactional retail analyses.

2. Literature Review

2.1 Evolution of Sales Intelligence and Enterprise Business Intelligence

Today, sales intelligence has evolved from basic spreadsheets to CRM, business intelligence, predictive business analytics, and AI-driven decision-making. Without an enterprise-wide platform to share information on quality and performance, transactions, and customer data, enterprise data became fragmented, making sales reporting a manual process [1; 2]. Customer Relationship Management (CRM) solves these issues by grouping customer information, communications, opportunities, and sales processes into a single system. Later, BI enhanced sales analysis through dashboards, reports, and visual notifications, making it easier for sales managers to monitor revenue, customer activity, and performance without creating individual reports by hand [3; 4]. While traditional BI is still descriptive and relies on the consumers to deduce patterns between the indicators and make sense of changes in their behavior when interacting with a company or a product, artificial intelligence-based data architectures can aggregate data from various sources to detect patterns, anomalies, and trends to aid in more actionable insights [5]. Predictive analytics and its associated features, such as forecasting, segmentation, and purchasing-pattern analysis, followed. Another area that has developed is the ability to use natural language to access analytical information from AI and Generative AI systems. Users can ask questions about changes in customer behavior, customer revenue, and purchasing trends, and receive automatically generated interpretations. But natural-language access does not make the correctness of the underlying data disappear; the reliability of the calculations and appropriate contextualization remain essential. Even with these steps, errors at any stage can lead to incorrect sales insights, even if the generated explanation sounds convincing. Metrics like revenue, repeat purchases, average spend, and recency can be inferred directly from transactions but don't predict customer activity on their own. The relationship between the indicators, as well as the customer's purchase history, should be considered in their interpretation. As such, this study regards sales intelligence as multi-layered: sales transaction information is converted into customer-level analytical evidence, retrieved, and interpreted to inform sales decisions. The proposed approach is therefore split between calculating customer indicators and interpreting the implications of those indicators. The determinate ANALYTICAL PROCEDURES present the numerical evidence, and the AIY COMPONENT is used to give these meaning or provide explanations based on the evidence retrieved. This matters because the language model doesn't want to create its own customer metrics.

2.2 Artificial Intelligence in Sales and Customer-Level Decision Support

In sales, AI can support sales forecasting, customer segmentation, lead prioritization, opportunity assessment, and recommendations and support for next-best actions [6]. Much of this research is concerned with structured prediction – that is, predicting outcomes from past data, or categorizing customers into those that have been predefined. Generative AI can extend these applications by generating summaries and recommendations from structured, analyzed output in natural language. These systems need accurate evidence they can act on or rely on, and they need the opportunity to do so. Language-based customer recommendations can be complemented with quantitative recommendations such as transaction frequency, revenue, recency, or purchasing changes. These indicators, used regularly and together, provide a baseline for evaluating customer value and buying habits and identifying potential changes in account activity. This study does not aim to replicate all features of a CRM/B2B sales environment. Instead, it examines the process of making the necessary changes between the layer of individual transaction facts and the layer of customer-level intelligence, and making that information accessible through a user-friendly, AI-enabled interface [7]. This phase focuses on analysis and interpretation, not on managing sales pipelines, customer contact, or contracts. AI-generated content improves managerial decision-making by delivering analytics and information that guide the sales journey. But transaction data doesn't necessarily reflect all interactions and services between a customer and the company, such as ongoing sales engagements, service interactions, or commercial negotiations. Accordingly, you must interpret recommendations based on transaction records in the business context before taking any commercial action.

Figure 1 illustrates a predictive customer analytics framework in which transaction data is processed to identify customer patterns, generate predictive insights and support evidence-based sales decisions. The framework demonstrates how analytical outputs can inform recommendations while retaining human judgement in commercial decision-making.



Figure 1: Predictive Customer Analytics

2.3 Explainable AI in Enterprise Decision-Making

The use of AI in decision-making is gaining momentum, and transparency, explainability, and interpretability are increasingly important in AI use [8]. NIST believes it's essential to differentiate between what the AI system says and how it says it, and between interpretability and explainability. These qualities enable the assessment, recording, and control of AI systems. Explainability can be evidence-based at the customer level. The recommendation should include relevant analytic indicators, a description of the observed pattern of interest, and evidence. For example, an inactive customer may be matched with true inactivity in purchases, and a

high-value customer may mean a high-value customer with purchase behavior and/or finances. The NIST AI Risk Management Framework also includes explainability and interpretability, specifically through the characteristics of validity, reliability, security, transparency, privacy, and fairness [9]. This shift adopts a socio-technical perspective, highlighting that an AI system's outcomes result from the interplay among data quality, analytical techniques, the AI system's behavior, and humans. The study focuses on explainability primarily in the context of 'traces' of insights produced and supported by a customer's body of established evidence, gathered by the analyst. Customer metrics are not calculated directly from the language model; instead, an interpretive layer sits in between. This segmentation lets users use the supporting data in an explanation as the basis for analysis and choose whether to display the output for calculation analysis or human analysis using AI models.

2.4 Conversational AI and Natural-Language Sales Intelligence

For structured sales information, conversational AI can be an alternative to dashboards. Users don't need to follow predefined statements or conditions; they can ask questions about customer inactivity, product activity, purchasing, performance changes, revenue, etc. This can reduce technical barriers to accessing analytical information. Language generation isn't sufficient for reliable conversational analytics, however. A system must recognize what information it needs, what evidence it needs to make a judgment on, the calculation(s) it needs to undertake or, if appropriate, access, and the answer that it forms based on the evidence retrieved. This is especially crucial when the robot must use numerical data, since language generation can produce nonsensical values. Grounding is thus a key design criterion. The evidence layer relies on transaction data, with deterministic procedures calculating customer indicators and the AI section interpreting the verified results. These can be split into functions for numerical accuracy, evidence retrieval, and evidence-based explanation to assess the quality of the evidence-based explanation. It also serves as a tool for spotting incorrect values, values that are not relevant and aren't backed by evidence, and evidence that is not supported by values for recommendations. One relevant approach is retrieval-augmented generation, which uses information retrieval (IR) from an external source to augment language generation [10]. The proposed sales-intelligence framework uses accepted features of the retrieved information, meaning, it provides the AI component with relevant evidence for the customer, rather than information coded into the language model.

2.5 Product-Oriented AI Deployment in Sales Intelligence

For any AI sales-intelligence product to be truly effective, it must seamlessly connect data preparation, data analysis, evidence retrieval, model inference, logic explanation, and user communication. These functions are better thought of as a thread that runs through the product. It should integrate data preparation, data analysis, evidence retrieval, reasoning and modeling, and explanation and communication with the user. Model performance and faithful implementation are essential to consistent data and an effective human 'watchdog'. The UCI Online Retail Data includes transaction-level data, which can be used for customer-level analysis. It doesn't, however, include complete enterprise or B2B CRM or Sales features. It doesn't include sentiment, Contracts, sales interactions, opportunity stages, or account ownership. That makes the solution proposed more of a "customer intelligence solution that is transactional, but not an enterprise CRM. The product-oriented approach includes the linked analytical modules: transaction data preparation, deterministic customer analytics, evidence retrieval and interpretation, AI interpretation, and presentation of user-friendly messages. The structure enables numerical results to be drawn from the prepared data set and allows explanations to be traced back to the evidence on which the analysis is based. However, the recommended architecture still needs human intervention to interpret and apply the AI-generated recommendations. This may be especially true in sales, where transaction dynamics may indicate purchasing patterns; however, the intent or reason for the purchase may not be the same.

2.6 Research Gaps and Limitations in the Literature

The literature review reveals several gaps. Traditional business intelligence systems mainly present information and let users infer relationships among customer indicators and their meanings [11]. Second, many AI sales applications focus on specific targets (forecasting, classification, and segmentation) rather than a multi-signal approach to customer intelligence. Third, explainability is often discussed in terms of model behavior, but recommendations must also be clearly

linked to supporting business evidence. Fourth, conversational AI improves access to analytical information but also increases the risk of unsupported generation, inaccurate numbers, and inappropriate interpretation. These risks create a need for evidence-based architectures. Last, the metrics for assessing whole customer levels of sales-intelligence products must be considered from several angles – numerical accuracy, retrieval performance, recommendation relevance, explanation agreement, error rate, and response latency. The missing pieces call for a solution involving transaction-level data prep, deterministic customer insights, evidence retrieval, and model-driven AI interpretation and explainable conversational interactions [12]. This study, rather than a traditional CRM or B2B pipeline dataset, uses the UCI Online Retail dataset to explore how transaction data is turned into customer-level data for analysis. The dataset supports conclusions based on transactions but cannot capture all enterprise sales activities. In particular, it does not account for interactions within the sales team, opportunity stages, contractual specifics, or customer emotions. The model in this proposed framework is therefore product-based and needs validation for wider applicability using handset enterprise CRM, sales pipeline, and customer interaction data. The study process distinguishes the main gaps between empirical Transaction Analysis and the functionality offered by the proposed AI agent. It defines customer-level analytical measures and sets up an evidence-based architecture to assess future implementation in an operational system. The numerical criteria include accuracy, retrieval precision, recall, F1-score, recommendation relevance, explanation agreement rate, error rate, and response latency.

3. Methods and Techniques

3.1 Data Collection Methods

This study used quantitative research and secondary transaction data from the UCI Machine Learning Repository. The data, known as Online Retail, includes 541,909 transactions from an online retailer based in the United Kingdom from 1st December 2010 to 9th December 2011. It offers granular data that can be used to study customer transactions, customer influence, sales results, etc. The dataset has 8 main attributes: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. The value of InvoiceNo identifies an individual transaction; the StockCode describes the products purchased; the Description column contains the products' descriptions; the Quantity column contains the number of units purchased; the InvoiceDate column contains the date of a list of individual transactions; the UnitPrice column contains the price per unit; the CustomerID identifies an individual customer, and the Country column contains the country of a particular customer. A primary reason for choosing CustomerID as the new identifier for customer-level analysis was its ability to link transactions to individual customers.

For the analytical dataset, records without identifiable customer identifiers were too difficult to use to calculate customer-specific metrics and were omitted. No replacement customer identifiers were provided. The dataset was chosen because it contains a large amount of historical retail transactions to explore customer purchasing behavior and how sales intelligence could be gained from operational transaction data. However, it lacks customer profile details and does not include CRM opportunity records, sales communication histories, customer satisfaction details, or transactions with sales reps. As a result, the focus is mostly on observable purchasing behavior rather than more general relationship-management practices. The study relied solely on secondary transaction data, without interviewing, surveying, or actively recruiting human subjects. As a result, the transaction analysis is empirically grounded, including in data quality. In contrast, the proposed approach of providing analytical information extracted from transactional data is conceptually based on the framework of currently proposed AI technology.

Table 1 illustrates the principal variables in the UCI Online Retail dataset and their analytical roles. The variables capture transaction, product, customer, temporal and geographic attributes used to derive customer-level measures for sales analysis and explainable sales intelligence.

Table 1: ***UCI Online Retail Dataset Variables***

Variable	Analytical Role
InvoiceNo	Transaction identification and cancellation filtering
StockCode	Product identification
Description	Product information
Quantity	Purchase volume
InvoiceDate	Recency and purchase behaviour
UnitPrice	Revenue calculation
CustomerID	Customer-level analysis
Country	Customer geographic profile
Revenue	Derived as Quantity × UnitPrice

3.2 Data Preparation and Cleaning

Data preparation enhanced the consistency and appropriateness of transaction records for further analysis [13]. The cleaning process targeted data that could skew revenue figures or prevent accurate revenue attribution to customers. The cancellation items referred to the following cleanup process: Cancellation records were removed, transactions lacked CustomerID data, and records lacked or contained invalid Quantity and UnitPrice data. Invoicing numbers starting with the cancellation identifier "C" were used to identify the records for cancellations. One reason is that some transactions were missing customer information. Missing information was not filled in or replaced because transactions without customer identifiers are hard to analyze at the customer level with the expected reliability. The InvoiceDate field was standardized to a date-time format for analyses based on the data's chronological dimension and for comparing observations over time. A revenue variable was obtained by multiplying the quantity recorded by the price of the unit: [Revenue = Quantity × UnitPrice], where Revenue is the revenue of the transaction line. No “synthetic” customer IDs or artificial customer records were created. After all cleaning-specified operations are performed, the final analytical record count is made. Determine the final retained count and the number of records excluded at each cleaning stage in the same way (following the documented procedure). With the dataset ready, customer revenue can be determined, customer purchase frequency can be analyzed, average order value can be calculated, customer recency can be assessed, and customer segmentation can be done based on the data [14].

3.3 Data Analysis

The analytical layer translates transaction records into customer-level measures. Total revenue is the sum of transaction revenue per customer, and average order value is the sum of customer revenue (out of all transactions) divided by the number of distinct invoices (out of all orders). Frequency is the count of different invoices that are associated with a customer. Recency is the number of days since a customer's most recent purchase relative to the reference date (determined by the observation period). These measures are the foundation of customer behavior analysis. Revenue contribution is calculated to identify customers' large orders, and customers are categorized into different price groups using frequency and recency to measure existing price trends among active, declining, and inactive customers. Time comparisons also identify changes in purchases. Customers who have sold in the past but haven't

purchased for a long time, for example, can then be identified as potential reactivation customers [15]. Customer segmentation combines recency, frequency, and monetary value with purchasing behavior. The purpose isn't just to create statistical groups; it's to make sense of customer categories that can support sales decisions. For instance, customers can be segmented by high monetary contributors, frequent buyers, recent activity, recent inactivity, or long-term inactivity. The concept of product-level purchasing behavior is also explored, with the analysis focused on StockCode and Description, to identify differences in purchasing behavior across customer groups. The business rules information on the account-health indicators are based on measurable transaction behavior. Any decrease in activity is recognized by comparing how much the purchaser bought recently with a suitable past time frame; inactivity is recognized from a set of criteria that define when a purchaser doesn't want the product anymore. Potential reactivation candidates combine evidence of previous purchase value/frequency with extended periods of no identifiable transactions in the record. This helps to determine that the recommendation comes from definite analytical data and not from free interpretation of AI.

3.4 Explainable AI Agent Development

The AI sales-intelligence agent consists of a multilevel structure of: a cleaned transaction dataset, analytical processing layer, retrieval mechanism, AI model and conversational interface/explanation layer [16]. The analytical layer generates accurate customer data outputs prior to applying the AI model. These outputs include insights into revenue, frequency, average order value, product activity, customer segments and account-health indicators. The retrieval mechanism is linking the natural-language questions with the corresponding customer information data and analytical results. For instance, the language model is not used to perform a numerical comparison to the specified conditions when a request is made for dropping customers. Evidence picked up is provided to the AI system as supporting context. The model is used to "read" and "simplify" the evidence, but does not lead to numerical truth.

The conversational interface can be used to ask questions about customer groups, customer inactivity, product behavior, revenue, questions about individual customers, and questions about potential customer rewinners [17]. These are the types of questions that can be asked using the conversational interface: customer groups, customer inactivity, and questions about product behavior, questions about revenue, questions about individual customers, and questions about potential customer rewinners. If the information requested is not provided with the information available provided from the transactions, the system is designed to communicate the limitation rather than unsupported information. Explainability is achieved by linking recommendations to the indicators responsible for the interpretation, creating the sequence: data evidence → analytical calculation → evidence retrieval → AI interpretation → explanation.

Figure 2 illustrates the enterprise AI agent architecture, showing how cleaned transaction data moves through analytical processing, retrieval, AI interpretation and conversational interaction. The architecture demonstrates how verified evidence supports explainable recommendations while maintaining human oversight of sales decisions.

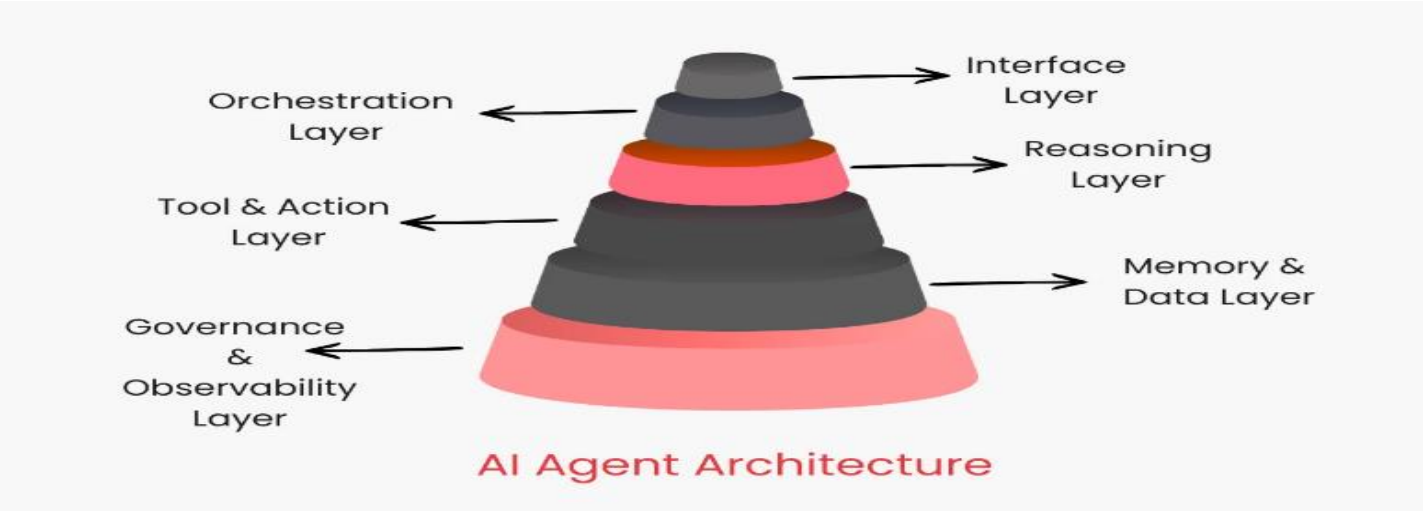


Figure 2: Enterprise AI Agent Architecture

3.5 System Evaluation Techniques

Because the proposed AI agent was not implemented in its own right and was not further test-run as part of the study evidence, this section provides an evaluation framework for future testing and validation of the agent's performance against the study evidence that will be available. The proposed evaluation dimensions focus on numerical reliability, evidence retrieval, recommendation quality, consistency of explanation, and operational responsiveness. Concerns about the accuracy of numerical measures should focus on the system's ability to report a customer's revenue level, number of purchases, average order value, and purchase recency, using independent measures of system performance to compare with reference values [18]. Test queries should be compared with outputs generated by the validated deterministic analytical layer. Retrieval performance should test whether the system can identify information within customer records and related analytical evidence. Precision indicates what % of retrieved evidence is relevant, and recall indicates what % of relevant evidence is retrieved. F1-Score is a fusion of precision and recall. The suggested selling actions need to be analyzed against customer purchase behaviors and analytical indicators to assess the relevance of the recommendation to those indicators. Assessment can include fixed standards and/or quality-driven independent assessments by a sales or business-intelligence practitioner. The quality of the explanation agreement should include an evaluation of the precision of the system's natural-language descriptions of the underlying analytical evidence. Evaluators should check that the explanations match the reported indicators and that the recommendations link to the supporting information at the customer level. The error rate is the percentage of errors that are out of place, either because the student didn't say the correct number, because what they said is not valid, or because it is an inappropriate interpretation. Response latency should represent the time it takes to pull evidence from the system and return a response after presenting a query to the user. Future evaluations should record the language-model version, the analytical environment, the query set, a test procedure, reference outputs, and the scoring criteria. This information would help scientists establish the accuracy of deterministic analyses relative to the conversational AI system's performance. The method provides an analytical process to extract customer-level sales intelligence from existing transaction logs. It offers an explainable sales intelligence architecture that makes the extracted intelligence available through conversational interaction at the customer level. This distinction allows the study to add proposed AI-agent development and evaluation and ensures consistency between the suggested AIs' construction and the supporting evidence [19].

4. Explainable Sales Intelligence Model

4.1 Product-Oriented System Architecture

The goal-oriented explainable sales-intelligence model is not intended to be developed as a language model (LM) per se, but rather as an integrated product for analytical use in the sales domain [20]. It organizes and cleanses transaction information to create customer-level insight, delivers the right evidence, and provides AI interpretation, enabling analytical results to be delivered via a conversational interface. The proposed architecture is based on the empirical data of the UCI Online Retail dataset. Includes data about transactions, products bought, quantities, transaction date, unit price, customer information, and customers' countries. The first layer of the architecture is not a CRM profile, an opportunity pipeline, or a record of sales interactions, etc., but rather a record of transactions. The data-preparation layer filters out "canceled" transactions, transactions where CustomerID is missing, and transactions with no quantity or an invalid price or unit. It also normalizes the date of a transaction and computes the amount of revenue based on the "quantity" multiplied by the "unit price. It is also useful for normalizing the date for a transaction and also computing the value of revenue based on "quantity" times "unit price. No artificial identifying labels are added. The CustomerID validation allowed transactions to be grouped for analysis to derive customer revenue, purchase frequency, recency, avg. order value, and product activity. These customers form the Customer-intelligence layer, which is used to study purchasing behavior and customer contribution. The evidence-retrieval layer supports finding analytical records related to given customer-level evidence, based on the natural-language query. The recovered proof is then passed to the AI interpretation layer, which provides clarifications and possible sales recommendations based on its analytics knowledge. The retrieved evidence is then passed to the AI interpretation layer, which can provide explanations and potential sales suggestions based on its analytics knowledge. Incoming answers are handled in a dialogue

format. This architecture separates numerical computation from language generation. The analytical layer still generates customer metrics, but now, when it comes to communicating and interpreting the verified results instead of performing a calculation to arrive at the sales figure, it is the language model that is needed. The architecture presented here combines five key components: preparing transaction data, processing analytical data, retrieving transaction evidence, analyzing content using AI feeds, and conversational interaction. This does not mean this study provided or deployed a fully working AI agent, as each component represents the workflow intended for the desired AI agent.

4.2 Customer Intelligence and Recommendation Framework

Based on transaction records, the customer-intelligence framework is a structured representation of customer behavior. It includes five analytical indicators: customer revenue, Purchase frequency, Recency, Average order value, and Product activity. Together, these indicators offer a more detailed understanding of customer contribution and buying behavior than revenue alone. The number of distinct customer purchases indicates purchase frequency, and customer revenue represents the total amount of money the customer contributes. Recency is the time since the customer's last purchase, and Average Order Value is the average dollar amount of their orders. Product activity provides additional data on purchased products and product usage levels in customers' buying practices. These indicators help classify customers using an operational definition developed in Chapter Three. Customers are classified into a high-value group by customer cumulative revenue (top 25%). Customer activity is evaluated using a look-back period covering the last 90 days of the Sprint support period (13 June to 10 September 2011) and the last 90 days of the Sprint support period (11 September to 9 December 2011). The classification system groups customers into high-value, inactive, growing, declining, and reactivation-ready segments. The actual growth/decline is analyzed mostly in terms of increases or decreases in customer revenue, and purchase frequency provides supporting data on engagement.

Customers who have not made any purchases in the last 90 days are categorized as inactive, and previous buyers with no purchases in the last 90 days are considered potential reactivation customers. In addition, it accounts for the fact that past customer values may not reflect healthy accounts. An established customer with a large turnover but who is now cutting back is in the category of "attention please. Likewise, a customer who previously made lots of purchases but has a long break between purchases could be a potential reactivation opportunity. The classifications are not necessarily exclusive. A high-value customer could also be inactive or declining, providing a mechanism to determine the customer's commercial value and the need for further attention. Apply the updated thresholds only to the cleaned dataset, and derive the numerical classifications and customer counts from this data, not from previously reported segmentation numbers. Instead of relying entirely on language-model interpretation, pre-defined analytical conditions generate recommendations. For each recommendation, users should include the corresponding customer indicators and supporting evidence so they can view the justification for the recommended action [21]. Because the framework is built on a transactional basis, it can only provide business-device suggestions based on observed purchasing behavior and is not intended to be a worldwide appraisal of the customer relationship.

4.3 Explainability and Governance Framework

Explainability is achieved through traceability of the relationship between records of the work actions, indicators generated, evidence retrieved, and interpretation (products) from the AI algorithm. If a customer activity decline results in a revenue decline, for instance, the system should provide insights indicating the relevant revenue changes to support that observation. This allows differentiation between language generated by the AI component of the system and the analytical information that can be calculated. Key governance aspects of the proposed framework include data validation, evidence grounding, output constraints, and human checking. In the analytical layer, numerical claims are generated, and the evidence-retrieval component in the AI interpretation layer provides the information needed to support the response. When the available transaction record does not contain enough information to answer a query, the system should not try to explain what caused it; it should simply report that the claim is not supported by the transaction records. These recommendations are at staff discretion and should not influence commercial decisions without review. Because the sale dataset may contain more information about

customers and other factors affecting the sale, information not tracked may also be relevant to a sale. The analysis of the public dataset does not show future deployment requirements, but rather requirements for security, access control, and auditability. The model then places explainability and governance within the proposed architecture to keep the customer intelligence informed, traceable, and under human control.

Figure 3 illustrates the proposed traceable governance architecture, showing how data validation, evidence grounding, output constraints and human oversight work together to support accountable AI decision-making. The architecture emphasizes traceability, controlled outputs and human review throughout the analytical process.

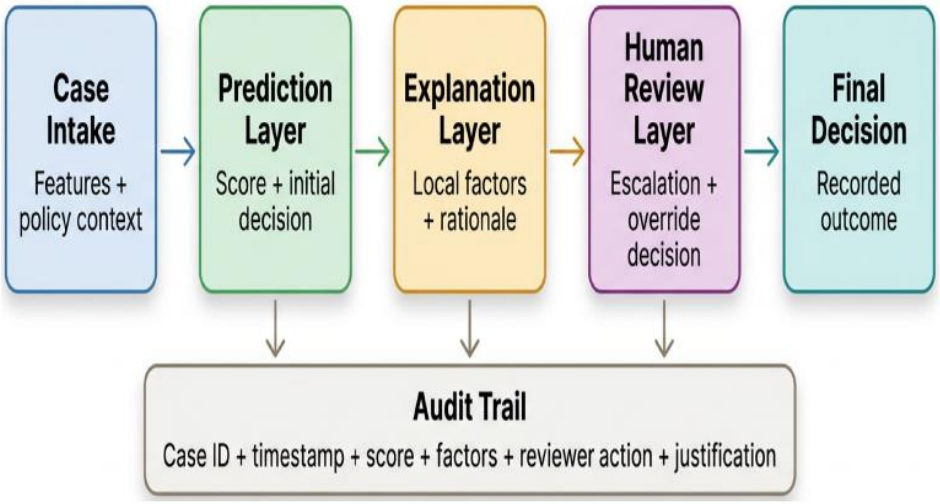


Figure 3: Proposed traceable governance architecture for accountable AI-enabled institutional decision support.

5. Results and Discussion

5.1 Dataset and Sales-Portfolio Results

The empirical data for analyzing customer purchasing behavior and developing the proposed sales-intelligence framework were obtained from the UCI Online Retail dataset. The raw data includes 541,909 transaction records from a UK online shop between 1 December 2010 and 9 December 2011. The records include details such as unit price, invoice, product, quantity, transaction date, customer, and country information.

The analytical dataset excluded cancellations and transactions missing CustomerID, quantity, or unit price data. The number of transactions multiplied by the unit price equals the amount of money the company received for each unit sold. Where a customer identifier was missing, it was not replaced with artificial values because customer-level transaction analysis cannot be accurately carried out without an identifier. The final analytical record count and descriptive statistics are not reported as confirmed results because the original analysis describes its treatment of missing values inconsistently. These figures revise the previously reported amounts and are not independently verified through the new cleaning process: 397,884 transaction lines; 18,532 invoices; 4,338 customers; 3,665 products; 37 countries; and £8,911,407.90 kept revenue.

These data can be used to learn more about every purchase, including transactions and customers. Complementary KPIs include customer contribution and purchase activity: RVE (revenue), Frequency (purchases), Average Order Value (AOV), Recency (purchases), and Product Activity. Customer-level indicators are very important, as total revenue does not necessarily or always correlate with a successful individual customer – it might show some high spenders, but it may also show a high number of people buying "little bits" – if that is the case. Therefore, this approach will emphasize differences in customer buying behavior rather than absolute sales numbers.

5.2 Customer Segmentation and Account-Health Results

Customer segmentation is a suitable way to segment customers based on their buying behavior and commercial contribution. The framework combines

revenue, recency, purchase frequency, average order value, and product activity to build multidimensional insight into customer performance. The method for determining high-value customers is the top 25% of identifiable customers, as determined by cumulative customer revenue. This relative threshold provides an equitable basis for prioritizing accounts by their contribution to sales. There was also some interesting additional information in case purchases weren't frequent enough, including purchase frequency and average order value, which show a distinction between customers that bring a high number of purchases and those that bring a high revenue per purchase. Account health is assessed by comparing the second 90-day period of same- or related-product purchases with the first 90-day period. Data for the recent period are from 11 September to 9 December 2011, and for the previous period are from 13 June to 10 September 2011. Customers who made at least one purchase in the last period are considered active, while those who made no purchases are inactive. The best indicator of growth and decline is an account's revenue change across both periods. At the same time, purchase frequency works in tandem; it is not the cornerstone for determining growth and decline.

When patterns differ, consider both revenue and frequency, and do not interpret either in isolation. Reactivation candidates include prior-purchase customers who have not made qualifying purchases in the past 90 days. The category provides a basis for prioritizing accounts that could benefit from reengagement. The classifications are not mutually exclusive, as performance has two aspects: customer value and account activity. A historically high transaction value may also be inactive - a customer's account could be applicable for both reactivation and retention efforts. The tools complement and support better account prioritization decisions than individual sales measures. The framework doesn't use revenue as the only indicator of a customer's importance. Still, it does account for the relationship between the customer's commercial contribution, purchasing consistency, recent purchasing activity, and product preferences. This way, sales folks can identify accounts on the lists that need work and retention, accounts that support growth, and accounts that might need re-harvesting.

5.3 Proposed AI Agent Evaluation

Results from the actual implementation of the proposed AI agent should be used for performance results and not results from other models with similar data that are not related to the proposed model. The gold-standard outputs are calculated independently and compared with the agent responses. Two aspects of the numerical accuracy—agreements with verified revenue and agreements with invoice counts and customers indicators. The task of retrieval precision, recall and F1-score is to determine if the correct customer records are retrieved. One measure of relevancy is the degree of agreement to the pre-defined conditions which were analyzed and one measure of the correctness of explanations is whether the reasons given match the evidence for each recommendation.

The hallucination/ error rate is the percentage of claims, attributes, wrong numbers/given numbers, and explanations that could not be found in available transaction evidence [22]. The distinction is crucial because the dataset is comprised of transactions but without any calls made by the sales person, no complaints from customers or obvious causal explanations of the reasons for a variation in customers purchasing. Additional measurements of operational performance are successful-query rate and response latency.

Figure 4 illustrates a proposed ReAct-based single-agent architecture in which observations, reasoning and actions form an iterative decision loop. The architecture supports evaluation-based feedback by connecting agent responses with evidence, reasoning processes and subsequent actions during AI-assisted decision-making.

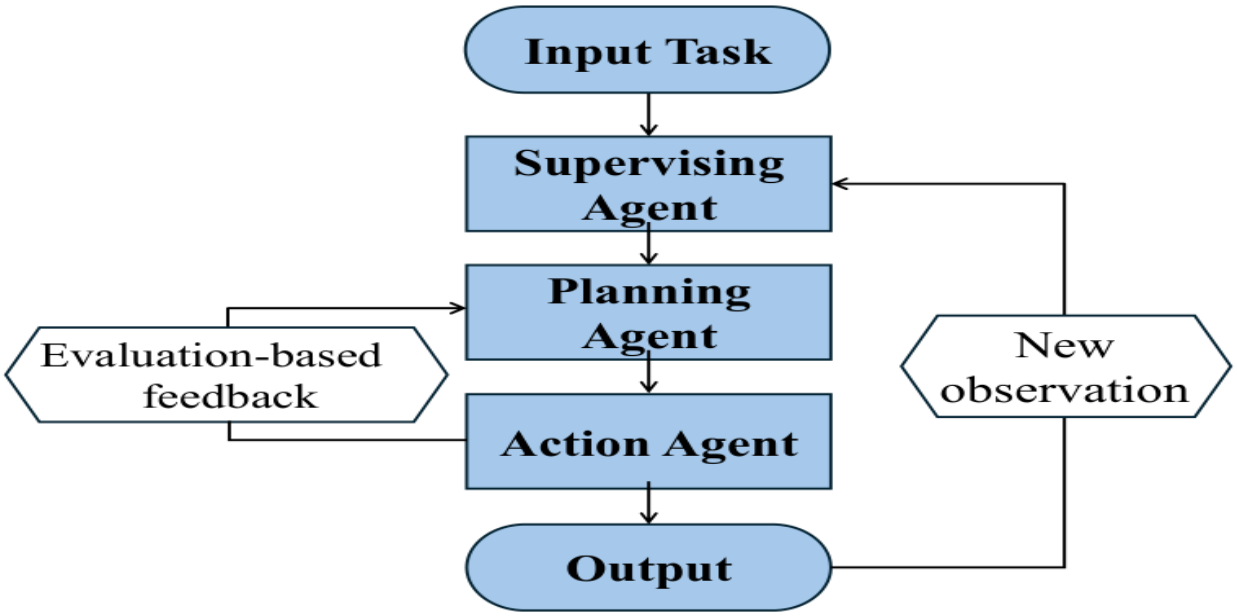


Figure 4: Proposed ReAct-based single-agent architecture with observation-reasoning-action loops and evaluation-based feedback.

5.4 AI Agent versus Conventional Dashboard Workflow

Traditional sales dashboards offer a graphical overview of business metrics, including sales figures, sales activity, product performance, and more. They enable monitoring and reporting by presenting information in charts, tables, filters, and predefined metrics. However, users sometimes have to add up a few indicators before they can figure out why a customer's performance dropped or what to do about it. The proposed AI-led method takes this analytical process a step further by offering a conversational interface through which the customer can ask for explanations and recommendations at the customer level. The sales representative wouldn't have to manually go to different parts of the dashboard to ask questions about the customer's buying behavior, income segments, or declining order frequency.

The device would carry the relevant analytical evidence and, based on it, build an explanation. The difference between the two use cases lies in how users interact with and interpret sales data, not in replacing hard-coded analytical tools. Dashboards help track overall performance, compare trends, and provide a visual summary. Conversational AI can add a new level of interaction and help ease the retrieval and interpretation of customer-level evidence. This proposed architecture thus defines the AI agent as providing additional decision-making support. Its reliability depends on the accuracy of the data provided for analysis, the evidence retrieved, and the consistency of the explanations it provides with the identified indicators. [23] This means recommendations are only information to support a sales action decision and should not be interpreted as independent, actionable recommendations.

5.5 Discussion and Practical Implications

It explains the importance of deterministic customer analytics and the need for customers to understand the attributes of conversational AI that make it explainable in enterprise sales intelligence. Converting transaction records into customer-based indexes yields a structured, customer-based understanding of commercial activity and purchasing behavior. Together with evidence retrieval and natural-language interpretation, these indicators can provide more user-friendly sales information and, in turn, support informed decision-making. One important aspect of the proposed framework is the split between AI-driven numerical computation and interpretation. The AI part accesses and clarifies the rationale behind the metrics for the customer, while the customer's analytical procedures define the metrics. This distinction will reduce unsupported revenue numbers, improve purchase-relationship recognition, and better track generated insights through a combination of revenue, purchase frequency, product activity, average order value, and recency details. They can help sales teams focus on top customers, track outdated accounts, and re-engage disconnected accounts. These categories, however, are analytical signals only and should not be used as absolute descriptions of customer

behavior; they should be considered alongside other business context. The other key principle, explainability, is also a focus, along with accountability in AI-powered sales decisions. Recommendations should be quantifiable in key customer metrics and transaction evidence, and users should be able to prove their impact before implementation. Therefore, the AI agent serves as a supplementary tool to support decision-making rather than replacing human judgment and traditional dashboards. The results are limited by the dataset and the proposed system's stage of development. The UCI Online Retail dataset excludes customer relationship history, sales interaction details, contract details, opportunity stages, and customer sentiment in addition to transaction data. For this reason, it cannot determine the root causes of shifts in buying behavior. Moreover, the planned conversational architecture has not yet been tested in operation. The retrieval accuracy, explanation quality, and relevance to recommendations need to be tested empirically with typical enterprise data from companies that sell products or services and validated against independent sources.

5.6 Future Research

Future research should first adopt the proposed architecture and compare performance using the proposed reproducible analytical procedures. Document data preparation, customer identification, revenue calculations, segmentation thresholds, and customer "account health" classification as scriptable logic so it is reproducible and verifiable. The system should also be tested to ensure it produces consistent sales answers, in terms of numbers, whether it's asked in one way or another. Further research is needed to explore the effectiveness of evidence-based conversational AI with representative enterprise sales data. Such datasets should include relationship information, sales history, product information, opportunity details, and other organization-relevant data. This would enable scientists to study the framework and determine whether it can support more detailed account evaluations than transaction reports provide. Conventional dashboards could also be compared with conversational sales intelligence systems to evaluate differences in how information is retrieved, tasks are accomplished, how accurately information is interpreted, and user experience. A human evaluation with sales representatives and managers would provide insights into whether the explanations and recommendations are practically useful for decision-making. Future investigations will also examine data privacy and access control, auditability, model reliability, and human oversight. For instance, in areas where sales information is critical and business decisions are being made, these factors are crucial to implementing AI-driven sales intelligence. Over time, this study would build continuity in the field by adapting the framework to other industries, assessing its success and effectiveness, and exploring its potential for enterprise sales management.

6. Conclusions

The study presented a product-centric sales intelligence framework that merges customer-level analytics, proof-retrieval resources, and conversational AI to support explainable enterprise sales decision-making. The UCI Online Retail dataset, a daily retail dataset from a Norwegian online retail store, was used to build an analytical approach that generated customer insights such as purchase frequency, recency, average order value, product activity, and daily revenue. These metrics can be used as a Cue for customers (For Customer Segmentation) and as a Health Check for customer accounts, retention, and reactivation opportunities. The proposed framework addresses a key limitation of current sales intelligence systems: connecting analysis results to clear, fact-based, verified natural-language explanations. It helps break the interpretative cycle of AI-driven content and provides a more structured approach to improving the traceability, consistency, and interpretability of sales insights. Updates to the conversational interface will complement dashboards by making information more customizable and understandable for customers, while also giving humans the time needed to make decisions. Sales Intelligence remains a conceptual and analytical point of view in the context of 'Explainable AI', but it still lacks real-world implementation and empirical testing to prove its effectiveness. Future work can then use representative enterprise sales data to assess the accuracy of the numbers, retrieval performance, explanation performance, and the relevance of the recommendations. In brief, the study offers hints for making enterprise-level sales management AI solutions more transparent, accessible, and fact-based.

References

1. Operations Management Efficiency and Employee Kivinen, S. (2020). Enhancing Experience through

Data. https://urn.fi/URN:NBN:fi:amk-2020063019452

2. Prakashkumar, P. K. R. (2025). Analytical Study of Vertex–Oracle Cloud Integration in Enterprise ERP Systems for Automated Tax Compliance and Efficient Financial Closings. International Journal of Accounting and Management Sciences, 4(4). https://scipubhouse.com/international-journal-of-accounting-and-management-sciences-ijams/content/ijams-vol-4-no-4-october-2025/analytical-study-of-vertex-oracle-cloud-integration-in-enterprise-erp-systems-for-automated-tax-compliance-and-efficient-financial-closings/
3. S. Rangu, K. S. Chadha and S. Samala, "AI-Powered Cloud Architecture, Agile Analytics, and Enterprise DevOps -Enabled Framework for Healthcare and Insurance Ecosystem Optimisation," 2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2025, pp. 1-7, doi: 10.1109/ICONSTEM65670.2025.11374414
4. Gondi, S., Sepuri, S., & Narayan, P. (2025). AI ETHICS FOR PUBLIC FINANCIAL SYSTEMS: A CROSS-SECTOR PERSPECTIVE. International Journal of Apllied Mathematics, 38, 2212-2233. DOI : https://doi.org/10.12732/ijam.v38i12s.1543
5. Sepuri, S. (2026). Comparative Analysis of Artificial Intelligence Test Generation Methods for Compliance Validation in Regulated Enterprise Workflows. Journal of Intelligent Decision Making and Information Science, 3(2s), 107-123. DOI: https://doi.org/10.59543/jidmis.v3.632
6. Bonthu, C., Malik, G., & Dhanagari, M. R. (2025, November). AI-Driven Secure and Scalable Data Architecture with Multi-Domain MDM and Cybersecurity for Real-Time Fintech Decision-Making. In 2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) (pp. 1-7). IEEE. https://ieeexplore.ieee.org/document/11374416
7. Rainsberger, L. (2023). Sales technology: An ocean of possibilities. In *Digital transformation in sales: How to turn a buzzword into real sales practice–A 21-step guide* (pp. 189-267). Wiesbaden: Springer Fachmedien Wiesbaden. https://link.springer.com/chapter/10.1007/978-3-658-38887-4_4
8. Balasubramaniam, N., Kauppinen, M., Hiekkanen, K., & Kujala, S. (2022, March). Transparency and explainability of AI systems: ethical guidelines in practice. In *International working conference on requirements engineering: foundation for software quality* (pp. 3-18). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-98464-9_1
9. AI, N. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile. *NIST Trustworthy and Responsible AI Gaithersburg, MD, USA*. https://pathologyinnovationcc.org/s/R8b_1722118549750.pdf
10. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, 9459-9474. https://proceedings.neurips.cc/paper_files/paper/2020/hash/6b493230-Abstract.html
11. Osakwe, J., Mutelo, S., & Obijiofor, N. (2023, December). Integrating Customer Relationship Management and Business Intelligence to Enhance Customer Satisfaction and Organisational Performance. A Literature Review. In *Proceedings of the International Conference on Information Systems and Emerging Technologies (ICISET)*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4664792
12. Gros, C. (2020). Life events in the design of public services: Creating communities of service in the Finnish national AuroraAI program. https://aaltodoc.aalto.fi/items/69826dbd-b12a-41c5-92a9-b181baaad806
13. H. Y. Paik, X. Xu, H. D. Bandara, S. U. Lee, and S. K. Lo, "Analysis of data management in blockchain-based systems: From architecture to governance," *IEEE Access*, vol. 7, pp. 186091–186107, 2019. https://ieeexplore.ieee.org/iel7/6287639/8600701/08938787.pdf
14. Aldhahery, M. A. (2026). An End-to-End Machine Learning Workflow for Customer Value Classification in Online Retail. *IEEE Access*. https://ieeexplore.ieee.org/abstract/document/11700919/
15. Holtrop, N., & Wieringa, J. E. (2023). Timing customer reactivation initiatives. *International Journal of Research in Marketing*, *40*(3), 570-589.

https://www.sciencedirect.com/science/article/pii/S0167811623000290

16. Baruti, R. (2023). *Analysis and Implementation of a Business Intelligence QlikView application for logistic and procurement management. Sews Cabind case for the shortage problem* (Doctoral dissertation, Politecnico di Torino). https://webthesis.biblio.polito.it/27551/
17. M. Shahabikargar, "Learning complex customer behaviors for intelligent recommendation and customer churn prediction," 2025. https://figshare.mq.edu.au/ndownloader/files/66631412
18. R. Heldt, C. S. Silveira, and F. B. Luce, "Predicting customer value per product: From RFM to RFM/P," *Journal of Business Research*, vol. 127, pp. 444–453, 2021. https://www.academia.edu/download/107373773/j.jbusres.2019.05.00120231114-1-mga1ds.pdf
19. S. Balaskas, "From recommendations to delegation: A systematic review mapping agentic AI in e-commerce and its consumer effects," *Information*, vol. 17, no. 3, Art. no. 222, 2026. https://www.researchgate.net/profile/Stefanos_Balaskas/publication/401253739_From_Recommendations_to_Delegation_A_Systematic_Review_Mapping_Agentic_AI_in_E-Commerce_and_Its_Consumer_Effects/links/69aa23a2a91b826e434775ae/From-Recommendations-to-Delegation-A-Systematic-Review-Mapping-Agentic-AI-in-E-Commerce-and-Its-Consumer-Effects.pdf
20. E. Sofryano, K. B. Subhash, and I. Rachmawati, "Artificial intelligence capability in business-to-business relationships: A systematic review of value creation mechanisms and relational outcomes," *IEEE Access*, 2026. https://ieeexplore.ieee.org/iel8/6287639/6514899/11753526.pdf
21. Konstantinovskiy, L., Price, O., Babakar, M., & Zubiaga, A. (2021). Toward automated factchecking: Developing an annotation schema and benchmark for consistent automated claim detection. *Digital threats: research and practice*, *2*(2), 1-16. https://dl.acm.org/doi/abs/10.1145/3412869
22. Vázquez-Ingelmo, A., Garcia-Penalvo, F. J., & Theron, R. (2019). Information dashboards and tailoring capabilities-a systematic literature review. *IEEE access*, *7*, 109673-109688. https://ieeexplore.ieee.org/abstract/document/8789402/
23. A. Mertzani, P. G. Giannopoulos, V. Malamas, P. T. Chountalas, and T. K. Dasaklis, "A neurosymbolic multi-agent system for explainable and traceable decision-making in organizational contexts," *IEEE Access*, 2026. https://ieeexplore.ieee.org/iel8/6287639/6514899/11527203.pdf,