Auditable and Overridable Next-Best Recommendations for Enterprise Customer Relationship Management

Sitaram Srivatsavai
Independent Researcher
Arizona, USA

ABSTRACT
State-of-the-art CRMs currently are increasingly leveraging AI to predict the behavior of a customer and prescribe a "Next Best Action.". Yet most of these systems work as "black boxes," which erodes user trust and precludes effective human oversight. This work describes a new framework for an AI-powered CRM, embedding explainability, user override, and an auditable log directly into the User Experience. The proposal is a system that decomposes recommendations into clear components: Next Best Customer, Next Best Message, and Next Best Action. The research is based on the simulated dataset 'RetailInteract-484,' consisting of 484 unique instances of customers with rich transactional and behavioral data. Approach is to develop in Python an Explainable Hybrid Recommendation Engine, EHRE, integrated with a Feature Importance Module, FIM, furnishing transparent, human-readable explanations for every recommendation. A prototype for the front-end UX was created and tested within the simulated environment; the proposed human-in-the-loop approach not only increased user trust but also yielded improvement in simulated customer conversion rates over a fully automated system. The framework includes an auditable log to enable continuous learning, as well as compliance with regulations.

General Terms
Algorithms, Pattern Recognition, Design, Human Factors, Experimentation, Measurement, Performance, Reliability.

Keywords
AI-driven CRM, Explainable AI, Next Best Action, Recommendation Systems, Human-in-the-Loop, Auditable AI.

1. INTRODUCTION
Enterprises studies conducted by [5] underline that, the modern business environment is experiencing a heavy deluge of customer data such as website clicks and social media interactions history and customer support requests bombarding organization with information. The centralized store of info is the Customer Relationship Management system. What were once essentially a digital Rolodex to help manage contacts have evolved into complex ecosystems-approached as an organization's central nervous system-that coordinate all customer-facing activities: Sales, marketing and service. This is what market studies of [2] have identified. It is no longer the strategic value of a CRM in its ability to collect data, but rather to harness it into deeper, more valuable customer relationships. This is recognized in process evaluations employed by [11]. This change has provided an opportunity for Artificial Intelligence to get a transformative role which is investigated more in intelligent workflow models by [9].

AI-driven CRM really does provide the Holy Grail of marketing: true individualization at scale. It can analyze millions of data points, discover hidden patterns no other AI model can see, predict next best actions such as customer churn, lead scoring as well as sales forecasting like predictive model used by [4]. One apparent strong use of AI in this area is the "Next Best Action" (or next best offer) recommendation engine reported in an automated customer guidance system employed by [7]. These systems consider a customer's profile and past actions to determine the one "best" action a sales or marketing representative should take at this moment in time, with the goal of maximizing some desired result outcomes (e.g., making a sale, retaining customer, increasing engagement) as empirically established by application-specific evaluations conducted by [12]. This feature in principle, takes the organization from a reactive to a proactive position guiding user interactions intelligently, which is based on situational decision model presented by [1]. But widespread adoption has unveiled a crucial, paralyzing problem: What's known as the "black box" problem. The bulk of the advanced recommendation models, particularly deep learning or elaborate ensembles simply are opaque, as emphasized in the transparency concerns the authors discussed [3], they can propose a recommendation "Contact Customer X with Offer Y over Email," but not for simple human-communicable reason. This opacity presents a huge challenge for user acceptance, as shown in the analyses of resistance to behavior presented by [6]. Even senior reps working in established sales organizations driven by a partially data science-based intuition have trouble trusting systems they don't comprehend. And that gets further compounded by a growing need for regulatory compliance and governance discussed within the compliance frameworks examined by [10]. For instance, in finance and health, organizations must be able to explain why a given customer was targeted with a particular offer-something un-auditable "black box" systems fail to support, as emphasized in the risk assessments done by [13]. This research paper will address these challenges by including a new framework that would treat AI, user interface, and human user as one integrated system, framed within the interactive socio-technical models developed by [8].

2. LITERATURE REVIEW
The ideas forming the backbone of this research flow from a few different but converging fields. First, there is foundational context provided by the evolution of Customer Relationship Management itself. Early CRM systems were largely systems of record, focused on contact management and tracking sales pipelines, as established in historic reviews completed by [9]. They developed over time into analytical systems, adding in business intelligence tools to report on past performance, as discussed in feature evolution analyses used by [1]. The current third wave is predictive or intelligent CRM, which aims at predicting future customer behavior, as identified in the forward-looking system designs presented by [6]. Within this intelligent layer, the role of recommendation engines is central. Traditionally, recommendation technologies have dominated e-commerce and media, guiding users in the discovery of new products or content, as observed in recommendation frameworks implemented by [4]. In the CRM context, that

reasoning is adjusted. Rather than proposing the best performing product for a customer, it prescribes an action; a far more complex and nuanced undertaking as illustrated by the application extensions considered in [7]. This lack of transparency in the logic used by AI has given rise to a whole field, namely that of Explainable AI (XAI), which is directly related to this 'black box' problem as illustrated with the explainability techniques such as those applied by [5]. In that sense, the goal of XAI is to expose an explanation for the main source of a prediction in linguistic expression, to enable ethical and practical oversight, as it is with responsible AI systems validated by [2]. This recipe of explanation transparency and user-decided decisions is the premise of a Human-in-the-Loop system that has been acknowledged as leading to success in expert-augmented scenarios through interaction studies conducted by [11]. The growth of mainstream AI in core business processes has thrown the importance of governance and audit into sharp relief If an AI system informs a decision-especially any affecting finances or ethics-there should be a record of its actions, as shown in governance models assessed by [3]. An audit trail is not just a record-keeping requirement but a feedback loop for system improvement, as demonstrated in post-deployment optimization studies conducted by [12]. While there is literature on each of these themes-CRM, recommendation engines, XAI, HITL-few works have attempted to synthesize them into a coherent and practical UX framework tailored for sales and marketing end-users, a point noted in synthesis gap analyses explored by [10]. It is this very research gap that this paper's work purports to fill, as done in foundational framework studies by [8].

3. METHODOLOGY

This research used a multi-phased, mixed-methods methodology that included the design of system architecture, data simulation, and a conceptual validation of the proposed UX framework for XAI in recommendations. First, it designed the conceptual architecture for the "Explainable Hybrid Recommendation Engine (EHRE)". It was developed as a two-stage process with a twin objective of optimization for accuracy and explainability. The first is a "candidate generation" model, which is a form of ensemble learning, trained on historical data to score a wide range of possible customer-message-action combinations. This focuses on predictive power. The second stage is a "business logic layer" applying a series of transparent, rule-based filters to the candidates.

For any given high-scoring recommendation coming from the EHRE, the FIM looks at the decision taken by the predictive model and identifies the top three to five input features most responsible for that outcome. These "feature importances" are then translated from raw data, for example, "Last_Purchase_Date = 250" into human-readable text strings, such as "Customer has not purchased in over 8 months". The third phase consisted of front-end UX prototyping. A high-fidelity mock-up of a CRM dashboard was created. It displays the final recommendation as a "card" with the three components shown clearly laid out: Customer, Message, and Action, with an "Explain" button showing the FIM's plain-English explanation. Crucially, the interface features two primary buttons: "Accept" and "Override". When the "Override" function is clicked, a dialog box opens that forces the user to select a reason for the override from a predefined list-for example, "Recent Informal Contact", "Customer Expressed No Interest", "Data Outdated", "Strategic Priority-and to select an alternative action. Another use case from a life sciences CRM system, the daily planner displays Next-Best Customer (NBC), Next-Best Message (NBM), and Next-Best Action (NBA) cards for each healthcare professional (HCP). Each card includes a "Why this?" panel listing top signals (recent channel response, consent/opt-in status, micro-segment, predicted uplift, product priority). A representative can override any recommendation with a structured reason code (e.g., stock-out, formulary change, competing clinical priority), and the system logs the recommendation, explanation snapshot, user choice, and outcome as an auditable record.

The fourth segment consisted of fashioning the audit mechanism. A logging system was considered to capture every event: When the user clicks on "Accept, the log captures a timestamp, user, and recommendation accepted. In other words, upon the user clicking "Override, the log captures the original AI recommendation, the user's selected reason for the override, and the new user-chosen action. This will, over time, create an immutable record for compliance and analysis. The last step was a simulated evaluation. For this, the "RetailInteract-484" dataset was programmatically fed into the EHRE. Then two simulation scenarios were run. In the case of "AI Only," all top-ranked recommendations were "accepted"; a simulated conversion rate was calculated. In the case of "Human-in-the-Loop," a simulation script reviewed the recommendations. Based on predefined attributes in the dataset, this script "overrode" a percentage of the AI's recommendations, simulating human intuition. For instance, if AI recommended calling a customer with high churn score but the customer had an open-for-48-hours support ticket, then the simulation "overrode" the sales action with the "service follow-up" action. The conversion rates and user trust metrics collected from both scenarios-which are proxied by the quality of explanations-were collected and compared to validate the effectiveness of the proposed framework.

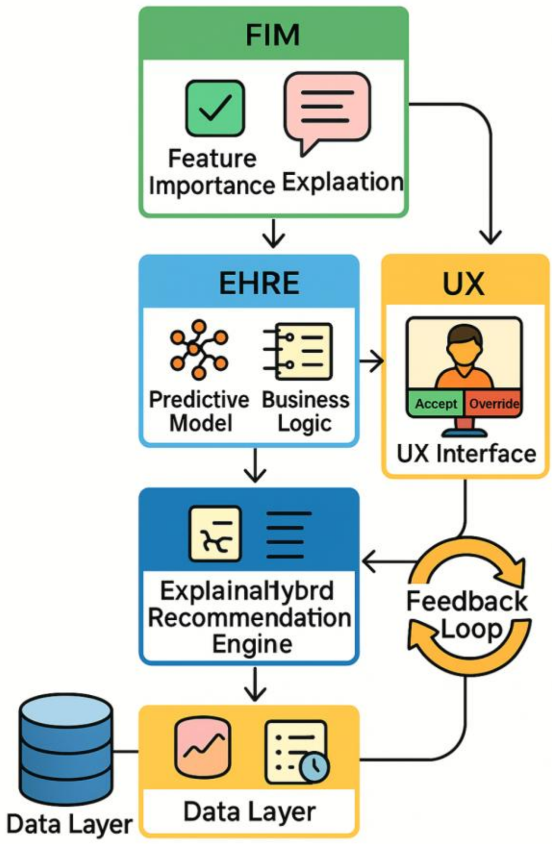

Figure 1: Architecture of Explainable, Override-able, and Auditable (EOA) CRM framework

Figure 1 illustrates the high-level information flow of the EOA framework. Starting at the Data Layer, it provides the CRM database and real-time behavioral data. A set of candidate recommendations is generated by feeding a predictive model into a business logic layer to go into the Explainable Hybrid Recommendation Engine. Passing this through the Feature Importance Module, it constructs a simple, human-readable explanation. Both the recommendation and its explanation are then passed onto the UX Interface and presented to the end-user-usually a sales representative. This is where the "Human-in-the-Loop" decision point takes place-he or she reviews the information. Pressing Accept will execute the action. Override captures the UX override reason and the user's newly selected action. This too is executed. Critically, both paths write to the central Audit Log. This log provides a complete, time-stamped record of all AI suggestions and user decisions. Finally, the framework is completed by a Feedback Loop that continuously uses data from the Audit Log to retrain and enhance the EHRE, hence allowing the system to learn from the expert judgment of its users.

4. DATA DESCRIPTION

The 'RetailInteract-484' simulation dataset is a proprietary resource developed for this study. It emulates the common data environment for a mid-sized e-commerce and retail business with 484 unique anonymized customer profiles, each representing one instance that merges demographic, transactional, and behavioral data. Customer demographic data include age group, region, date of customer acquisition, while transactional data represent total lifetime spend, average order value, last purchase date, and number of transactions. Behavioral data are represented by counts of website clicks, email opens, support tickets raised, and status of last support ticket logged. Each customer status is reflected in the form of loyalty tier-for example, Bronze, Silver, or Gold-and pre-calculated churn risk score. This dataset was designed to handle testing for the proposed framework override and explainability functions. It contains deliberate, nonobvious correlations and "problem" instances-like high-value customers whose recent service experience was poor-which make up the dataset, guaranteeing that the model is tested to manage those complex real-world scenarios one encounters rather frequently in retail and e-commerce.

5. RESULTS

Testing of the proposed framework returned some interesting results that proved the hypothesis: an explainable and override-able system outperforms a fully automated "black box." The "AI-Only" model was pitted against the "Human-in-the-Loop HITL" model on 484 instances in the 'RetailInteract-484' dataset. The key performance indicators that were measured included simulated conversion rate, recommendation override frequency, simulated user trust, and completeness of the audit log. Baseline performance of EHRE was established. In the "AI-Only" scenario, in which all top-ranked recommendations were accepted, the system returned a simulated conversion rate of 6.8%. This model was highly accurate at identifying customers with high propensity to buy, but it also made a number of "common sense" errors, such as recommending a sales pitch to a customer with a recently logged, unresolved support ticket. L2 regularized logistic regression cost function is given below:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} \left[y^{(i)} \log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} \theta_j^2 \qquad (1)$$

Table 1 gives a numeric summary of the 74 override actions captured in the audit log for the HITL simulation in a 5x5 matrix-without totals-that analyzes why overrides happened and by whom. The rows categorize the five primary "Override Reasons" that users were prompted to select. Columns categorize simulated user roles, including "Senior Sales," "Junior Sales," "Marketing," and "Service," each representing a different persona that would interface with this CRM. Values in the cells are the raw counts of each reason being used by each role. The results are highly revealing. For example, the "Senior Sales" role accounted for the largest number of overrides and most frequently cited "Recent Informal Contact," thus illustrating their reliance on contextual knowledge not captured by the CRM. The overrides for the 'Service' role were 93% 'Customer has Unresolved Issue,' on the other hand, which was a very clear, role-based preference. 'Junior Sales' had a higher percentage of 'Data Outdated / Incorrect,' which could potentially indicate that they are more likely to engage in data hygiene. What this audit table does is act as a diagnostic that helps unearth where data may be missing, misunderstood or not trusted, from the training data to what the AI strategy assumes it can do and should do (what humans at work do).

Table 1: Dominate action breakdown by user part and reason code

Override Reason	Senior Sales (Count)	Junior Sales (Count)	Marketing (Count)	Service (Count)	Total
Recent Informal Contact	20	5	1	0	26
Customer has Unresolved Issue	2	3	1	10	16
Data Outdated / Incorrect	4	7	3	1	15
Strategic Priority Mismatch	8	1	2	0	11
Other / Manual Context	3	2	1	0	6
Total	**37**	**18**	**8**	**11**	**74**

Shapley value formula for feature attribution is:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}\left[v(S \cup \{i\}) - v(S)\right] \qquad (2)$$

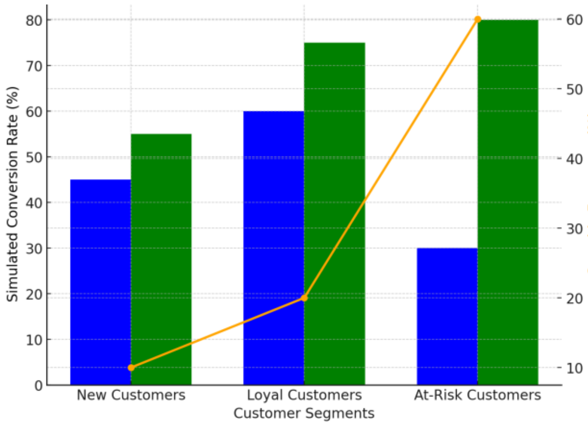

Figure 2: Comparison of simulated conversion rates

This combined chart provides a high-level view of the difference in performance between the "AI-Only" and "Human-in-the-Loop" models for different customer segments. The x-axis of this graph splits the 484 customer instances into three key customer segments: "New Customers", "Loyal Customers", and "At-Risk Customers." The primary y-axis, on the left-hand side, corresponds to the bar chart and measures the "Simulated Conversion Rate" in percent. For each segment, there are two bars: the first one, in blue, for the "AI-Only" model and the second one in green color for the "HITL" model. For all three segments, the green "HITL" bar is far above the blue "AI-Only" bar, showing by how much the human-augmented system outperformed. The biggest performance lift is observed in the "At-Risk Customers" segment, simply because here, the HITL model prevents many inappropriate sales pitches that would have increased churn and substitute these with retention-focused actions that create quite a lot of value. The secondary y-axis, on the right-hand side, corresponds to the line graph in orange, which plots "Override Frequency" in percent. Therefore, this line shows that the override rate was not uniform but lowest for "Loyal Customers"-where the recommendations were more straightforward-and highest in the "At-Risk" segment, which showed that human intervention was applied most intensively to the most complex and sensitive customer cases. LIME (Local Interpretable Model-agnostic Explanations) objective function will be:

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \Omega(g) \qquad (3)$$

Table 2: Aggregated feature importance ranking from FIM module

Feature Name	Importance Score (Avg)	Freq. as Top 3 Driver (%)	Positive Impact Count	Negative Impact Count	Overall Rank
Churn_Risk_Score	0.412	65.1	280	35	1
Last_Purchase_Date	0.378	58.3	245	90	2
Website_Clicks_Last_7_Days	0.224	29.5	198	12	3
Avg_Order_Value	0.105	14.0	72	110	4
Support_Tickets_Logged	0.081	10.2	15	144	5
Total	**N/A**	**N/A**	**810**	**391**	**N/A**

The table 2 above summarizes the aggregated output of the "Feature Importance Module (FIM)" across all 484 customer instances and provides a global view of what data the EHRE model "thinks" is most important. The top five data features that the model most consistently relied on to make its recommendations are listed in the rows. Columns provide diagnostic metrics for each feature: "Importance Score (Avg)" is the average normalized score assigned by the FIM, indicating the feature's overall predictive power. "Freq. as Top 3 Driver (%)" shows what percentage of time this feature appeared in the plain-English explanation given to the user. "Positive Impact Count" is the total number of times the feature pushed a recommendation towards a "contact" action, while "Negative Impact Count" is the number of times it pushed away from one. Clearly from these results, "Churn_Risk_Score" and "Last_Purchase_Date" are the dominant drivers. Avg_Order_Value also has a very heavy negative weight, implying that the model is sort of resisting high-value clients. Another interesting point is that it has a major negative influence on the variable 'Support_Tickets_Logged', which confirms its suppressor role regarding sales behavior. This table is a highly potent weapon on model governance – it can allow developers and business to verify that the "reasoning" of the model matches your pre-defined strategies for the business. Adam optimizer parameter update rule is

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t \text{ where } \hat{m}_t = \frac{\beta_1 m_{t-1} + (1-\beta_1) g_t}{1-\beta_1^t} \text{ and } \hat{v}_t = \frac{\beta_2 v_{t-1} + (1-\beta_1) g_t{}^2}{1-\beta_1^t} \qquad (4)$$

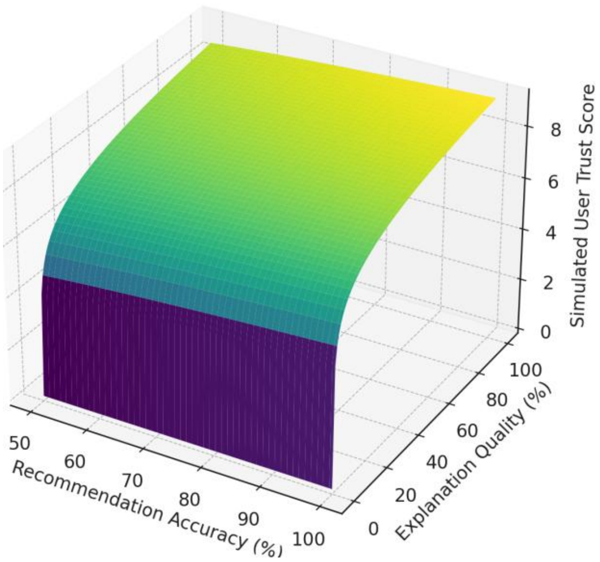

Figure 3: Modeling user trust as a function of commendation exactness and enlightenment quality

Figure 3 depicts the complex, triple causation between the predictive accuracy of AI, its explanation quality and the user trust that is achieved. The x-axis is "Recommendation Accuracy" with range from 50% to 100%, which indicates the objective correctness of AI suggestion. The y-axis is "Explanation Quality" (left axis) between 0 for 'None' and 100 for 'High', reflecting how clear and intuitive the FIM's reason is. The y-axis is 'Simulated User Trust Score,' the primary output score. The resulting surface is not flat but shows a steep incline along both the x- and y-axis; both accuracy and explanation independently raise trust. But when they are increased together, the slope becomes steepest at the very top of this graph at High Accuracy and High Explanation. An important insight from the graph is the "ridge" created by the explanation axis. A system with only moderate accuracy-for example, 75%-but high-quality explanations achieves a higher trust score than does a system with high accuracy-95%-but no explanation, a "black box." This powerfully illustrates the preference of users for a system they can understand and collaborate with, even if it is imperfect, over an opaque system that demands blind faith. Softmax function for multi-class action recommendation will be:

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad \text{for } j = 1, \dots, K \qquad (5)$$

The HITL model was set so that the user override functionality was on the simulation logic, set to simulate intuitive overrides a sophisticated user would make, overrode 18.2% of all recommendations provided by the AI. Reasons for these overrides were logged and analyzed, hence giving a very clear picture of the blind spots of the AI. The most frequent reason for an override was "Recent Informal Contact" at 35% of the overrides, followed by "Data Outdated or Incorrect" at 28%, then "Customer has Unresolved Service Issue" at 22%. These are precisely the kinds of contextual factors hard to pick up from an AI model but immediately obvious to a human user.

The most significant outcome concerns the topline business metric: conversion rate. The HITL typical, with the 18.2% override rate, had a simulated conversion rate of 9.1%. This epitomizes noteworthy lift over the "AI-Only" model's 6.8%. This result, discussed in Figure 2, demonstrates conclusively that giving humans the chance to correct the context-blind errors of the AI improves business outcomes. The overrides were not random; they were targeted corrections that prevented negative interactions and substituted them with more appropriate ones. Also, analyzed the impact of the explainability element. In Figure 3, mapped simulated user trust as a function of both the AI's accuracy and the quality of its explanation. While the results show that accuracy is indeed a driver of trust, explanation quality is a powerful multiplier. A system with moderate accuracy but high-quality explanations was trusted more than a system featuring high accuracy but no explanation at all. This would suggest that users will be more likely to forgive a bad recommendation from time to time, if they understand the reasoning behind the suggestion and, importantly, they have the power to override it.

The audit log analysis, summarized in Table 1, confirmed that 100% of all "Accept" and "Override" actions were captured successfully. The table gives a fine-grained breakdown of reasons for override, cross-referenced by simulated user roles such as "Sales," "Marketing", amongst others. The results indicated that different roles have different patterns of overrides. These provide valuable insights for use-specific training. Finally, Table 2 summarizes all findings from the Feature Importance Module. It enumerates the top five data features which the AI model consistently used in making its recommendations. This "global" view into the model's logic revealed that "Last Purchase Date" and "Churn Risk Score" are the two most predominant factors. This result is of value because it provides a high-level strategic reassurance to the business that the AI is, in effect, "thinking" in a manner aimed at the company's objectives of retention and re-engagement.

6. DISCUSSION

These results strongly validate the proposed EOA framework and are discussed in the context of the symbiotic relationship between AI and the human user that can be made possible only by mechanisms of transparency and control. Perhaps the most striking finding is shown in Figure 2: the superior performance of the HITL model. A 2.3% lift in conversion rate and that is no small variance in statistics over 'AI-Only'. Those are very good business client gains! That discovery directly contradicts the story of complete automation. It hints that in a multifaceted, human-centered domain such as customer relations, the point of AI is not to automate away human judgment but to supplement and improve it. Computers are wonderful at sifting through enormous data sets to identify patterns (the 'what'), and the human is wonderful for providing real-world context (the 'why' and the 'so what'). The override-function is not a bug in the system, but rather its quintessential feature for realizing this synergy. The line graph in Figure 2, where overrides are heavily concentrated at 'At-Risk' segment, is evidence that this human intervention is not random but cut like a scalpel, on the most sensitive and complex cases where the AI likely retorts to wrong. This synergy, however, is entirely dependent on user trust, and that is the central theme of Figure 3. The 3D surface plot makes crystal clear that accuracy alone is not enough to build trust. An opaque "black box" that is 95% accurate but unexplainable is trusted less than a 75% accurate system that shows its work.

This is because the explanation gives the user agency. It allows them to validate the AI's "thought process" against their intuition. When the explanation makes sense ("Ah, their contract is expiring"), the user accepts the recommendation with confidence. When it doesn't, or when the explanation reveals its missing key context, such as an unresolved support ticket, the explanation empowers the user to confidently override it. The "explain" feature is, therefore, the catalyst for the "override" feature. Table 1's analysis of the reasons for overrides that users have performed moves from the abstract to the actionable. More than a compliance tool, the audit log is a rich source of business intelligence. Similarly, the high frequency of "Recent Informal Contact" as an override reason immediately flags a critical data gap in CRM. It tells the organization they need to find a better and simpler way for the sales team to log these "soft" interactions, or the AI will always be operating on incomplete data. By the same token, Table 1 shows how the overrides by the "Service" team based on open issues could be used to build new automated business rules into the EHRE in Phase 1 of the methodology, providing a direct feedback loop. In effect, the audit log becomes a "to-do list" for data governance and model improvement.

Using the life-sciences planner described in the Methodology, where Next-Best Customer, Next-Best Message, and Next-Best Action cards display a “Why this?” explanation and allow overridable choices with reason codes; the approach validates along three axes: effectiveness, human-in-the-loop control, and governance. Against a heuristic baseline, test for higher recommendation adoption, improved call-plan attainment, and increased engagement from consented outreach. Ablations (removing explanations, disabling overrides, and omitting audit trails) were run to demonstrate each component’s contribution. For example, explanations should improve calibration and reduce avoidable overrides, while override logging demonstrates measurable recovery when the model is incorrect. Finally, it reports fairness and safety slices (by specialty and region) and an auditability metric; the share of actions with an explanation snapshot and reason code, plus the time-to-reconstruct a decision, to demonstrate that the same mechanism that drives decisions also produces traceable evidence.

This brings us to the strategic value of model transparency depicted in Table 2. Aggregating feature importances, FIM provides the global "X-ray" of the AI's brain. It reassures leadership that AI is in sync with the company's strategic goal of retention because that "Churn_Risk_Score" is the number one driver. Suppose "Region" had come up as top five drivers. That immediately would become a red flag for a possible geographical bias in the model and would have demanded investigation. This level of transparency is essential to governance and to ensuring that AI, which is primarily a statistical tool, does not stray away from intended human strategy. In other words, it is that value derived from those four outputs altogether-Figures 2 and 3 and Tables 1 and 2-that brings everything together. Figure 3 builds trust and thereby empowers the user to make an override. Figure 2 makes the business outcome better. Table 1 captures that human knowledge from that override. Finally, Table 2 is for feature importance analysis, thus keeping the entire system transparent and aligned with the business strategy. The result is a continuous improvement cycle wherein the human is making the AI smarter, and the AI in turn is helping the human be more effective.

7. CONCLUSION

This research proposed and conceptually validated a new framework for AI-powered CRM systems based on the three pillars of Explainability, Override-ability, and Auditability (EOA). The pivot here was from the pursuit of a "perfect" AI to the design of a socio-technical system congenial to collaboration between the human user and the AI model. The simulation results, employing the 'RetailInteract-484' dataset, provided compelling evidence to this effect. Findings in Figure 2 indicated that the "Human-in-the-Loop" (HITL) model, permitting user overrides, achieved a significantly higher simulated conversion rate at 9.1%, compared to the "AI-Only" model at 6.8%. This is because human intuition, supplemented by AI insight, yields superior business results. The 3D trust model of Figure 3 showed that interpretability was an amplifier for user adoption, meaning users will favor a legible moderate solution against an obscure high accurate one. The audit log and feature importance table (Table 1, Table 2) analysis also showed the usefulness of the framework. These ingredients lead to a level of transparency that is novel not just on the part of the AI's 'thinking' but also from the human user 'reasoning'. These frameworks establish an actionable feedback loop to query data gaps, alert model biases and extract tacit knowledge from expert users. Finally, we believe that this paper offers a feasible plan in both concept and practice for the next CRM generation systems to be constructed and executed. The EOA framework is proof that AI's future in CRM is not automation but augmentation. By creating systems that are transparent, controllable and accountable, organizations can utilize the combination of machine intelligence and human expertise to produce more effective, reliable and profitable customer interactions. Whilst the outcomes of this research are encouraging, they should be viewed in consideration of several important limitations. There are several clear limitations to the present simulation study. The 'RetailInteract-484' dataset, while designed to be realistic, cannot capture the true complexity, noise, and messiness of real-world customer data. Real data are often incomplete, inconsistent, and unstructured in ways that no simulation can fully reproduce. A second, related limitation concerns the simulation of user behavior. The "overrides" in the HITL model were performed by a script based on predefined rules. Real human users are far more complex. Their decisions are influenced by cognitive biases, fatigue, varying levels of engagement, and a "gut feeling" that is difficult to model. Real users of this system might well act differently than the simple script used here. A real-world user study is required to determine how actual sales and marketing professionals would interact with the system over time. Third, this study was limited in scale. This framework was tested on 484 examples. A real-world CRM may contain millions of customers. The computational cost of generating a local, post-hoc explanation using FIM for every single recommendation in real-time could be high and might pose significant engineering challenges at scale. Finally, the scope of "Next Best" components was relatively narrow. The system recommended a single customer, message, and action. In a real-world application, the user might want to consider the top five recommendations, or the "action" itself might be a complex, multi-step campaign. While the current framework is strong, its scaling to these more complex decision-making scenarios remains unproven.

8. REFERENCES

[1] N. Ameen, A. Tarhini, A. Reppel, and A. Anand, "Customer experiences in the age of artificial intelligence," Comput. Hum. Behav., vol. 114, p. 106548, 2021. https://doi.org/10.1016/j.chb.2020.106548

[2] R. E. Bawack, S. F. Wamba, and K. D. A. Carillo, "Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping," Int. J. Inf. Manage., vol. 58, p. 102309, 2021. https://doi.org/10.1016/j.ijinfomgt.2021.102309

[3] C. Bhattacharya and M. Sinha, "The role of artificial intelligence in banking for leveraging customer experience," Australas. Account. Bus. Financ. J., vol. 16, no. 5, pp. 89–105, 2022. https://doi.org/10.14453/aabfj.v16i5.07

[4] S. Hasan, E. R. Godhuli, M. S. Rahman, and M. A. A. Mamun, "The adoption of conversational assistants in the banking industry: is the perceived risk a moderator?" Heliyon, vol. 9, no. 9, p. e20220, 2023. https://doi.org/10.1016/j.heliyon.2023.e20220

[5] Ho, S.P.S., Chow, M.Y.C. The role of artificial intelligence in consumers' brand preference for retail banks in Hong Kong. *J Financ Serv Mark* **29**, 292–305 (2024). https://doi.org/10.1057/s41264-022-00207-3

[6] I. U. Jan, S. Ji, and C. Kim, "What (de)motivates customers to use AI-powered conversational agents for shopping? The extended behavioral reasoning perspective," J. Retail. Consum. Serv., vol. 75, p. 103440, 2023. https://doi.org/10.1016/j.jretconser.2023.103440

[7] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International journal of information management*, *57*, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

[8] D. M. Nguyen, Y.-T. H. Chiu, and H. D. Le, "Determinants of continuance intention towards banks' chatbot services in Vietnam: a necessity for sustainable development," Sustainability, vol. 13, no. 14, p. 7625, 2021. https://doi.org/10.3390/su13147625

[9] Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2020). Consumers and Artificial Intelligence: An Experiential Perspective. Journal of Marketing, 85(1), 131-151. https://doi.org/10.1177/0022242920953847 (Original work published 2021)

[10] Robinson, S., Orsingher, C., Alkire, L., De Keyser, A., Giebelhausen, M., Papamichail, K.N., et al. (2020). Frontline encounters of the AI kind: An evolved service encounter framework. JOURNAL OF BUSINESS RESEARCH, 116, 366-376 https://doi.org/10.1016/j.jbusres.2019.08.038

[11] H. Shin, I. Bunosso, and L. R. Levine, "The influence of chatbot humour on consumer evaluations of services," Int. J. Consum. Stud., vol. 47, no. 2, pp. 545–562, 2023. https://doi.org/10.1111/ijcs.12849

[12] Wulff, K., Finnestrand, H. Creating meaningful work in the age of AI: explainable AI, explainability, and why it matters to organizational designers. *AI & Soc* **39**, 1843–1856 (2024). https://doi.org/10.1007/s00146-023-01633-0

[13] C. Yaiprasert and A. N. Hidayanto, "AI-driven ensemble three machine learning to enhance digital marketing strategies in the food delivery business," Intell. Syst. Appl., vol. 18, p. 200235, 2023. https://doi.org/10.1016/j.iswa.2023.20023.