iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

MULTIMODAL DEEP LEARNING ARCHITECTURES ADVANCING PREDICTIVE MARKETING ACCURACY BY SURPASSING HUMAN COGNITIVE BIAS IN UAE OFF PLAN PROPERTY DEMAND FORECASTING.

Muhammad Shahid
shahdtech@gmail.com

ABSTRACT
The rapid expansion of off plan property marketing in the United Arab Emirates has exposed critical limitations in human driven decision making, particularly in the areas of buyer qualification, demand forecasting, and lead conversion prediction. Traditional marketing judgments shaped by cognitive bias, nationality stereotyping, heuristic shortcuts, and fragmented cross channel data interpretation are increasingly inadequate in environments where buyers generate complex digital footprints across portals, social media, call centres, CRM systems, and WhatsApp interactions. This study introduces a multimodal deep learning architecture designed to surpass these human biases by integrating heterogeneous data modalities into a unified predictive marketing framework. The proposed system employs transformer-based text encoders, convolutional neural representations for visual property engagement patterns, LSTM sequence modelling for lead journey timelines, and cross channel behavioural embeddings constructed through contrastive representation learning. A dataset of 280,000 historical off plan leads from major UAE developers was used to train and evaluate the model. Structural equation modelling was applied in parallel to compare AI derived intent predictions with human marketing judgments provided by sales agents and marketing managers. Results indicate that the multimodal deep learning model achieves a significant improvement in predictive accuracy, yielding a 32 percent increase in true positive buyer intent detection and reducing false positives by 41 percent compared with human evaluations. Furthermore, the model demonstrates strong robustness under market volatility scenarios, such as payment plan changes, launch date shifts, and macroeconomic fluctuations. This research provides empirical evidence that multimodal AI frameworks can fundamentally restructure performance marketing strategies in UAE off plan real estate by eliminating cognitive bias, optimizing audience selection, and enabling demand forecasting with significantly higher precision than human cantered decision processes. The findings establish a new pathway for AI enhanced marketing science within fast moving real estate sectors and contribute a scalable architecture for future predictive applications.

1. INTRODUCTION
The UAE has become one of the most digitally intensive property markets in the world with off-plan real estate deals taking up a controlling portion of yearly sales and the marketing climate being driven by high buyer mobility, cross-border investing, and intense performance marketing (Aichner and Shaltoni, 2022; Baum, 2020). Software developers and marketing teams are using more and more digital lead-generation tools like property portals, social media advertising, programmatic campaigns, WhatsApp business communications and AI-powered call centres. Making this ecosystem produces huge and multi-modal information streams such as text queries posted via advertisements, patterns of clickstream interaction across project websites, voice-to-text transcripts of call exchanges, image-level engagement heatmap, and timestamped CRM event logs, which capture the intensity and multimodality of current online buyer activity (Luo et al., 2021; Shankar et al., 2021). Even though these multi-channel digital records afford an unprecedented capability to understand buyer psychology and demand behaviour, the majority of real estate companies still rely on the intuitive decision-making process, in which the qualification of leads is based on the heuristic assessment of the agent, but not on objective evidence. This type of dependence on personal opinion leads to structural failures, particularly in as competitive and rapidly evolving a market as the UAE off-plan sector, where human judgment tends to collapse in information saturation (Higgins, 2021; Kahneman et al., 2021).
Marketing judgments in this area, made by human beings, are severely influenced by cognitive biases that inaccurately interpret cues on the side of buyers. Sales agents often use stereotypical classifications on leads depending on the nationality, income level, plan-payment appropriateness and channel appropriateness. These biases promote confirmation-based reasoning, in which the agents seek information that confirms their previous assumptions; anchoring, in which initial impressions have disproportionate effects on lead evaluation; and availability bias, in which memorable past cases have disproportionate impacts on actual statistical trends (Tversky and Kahneman, 1974; Kleinberg et al., 2022). This has caused marketing teams to over-allocate their

International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

budgets to channels which seem to have high-quality and ignore those that provide high hidden intentions. Likewise, potential buyers with high potential who engage with each other in a subtle manner, when existing online, can be underestimated just because their superficial profile does not conform to the mental model of an agent. Such structural overdependence on heuristics makes predictive accuracy weaker, forecasting accuracy weaker, and makes a direct contribution to the leaks in conversion at the top and bottom of the funnel even with the existence of rich data ecosystems (Bhandari and Bansal, 2021).

As a reaction to such constraints, multimodal deep learning has become a revolutionary analytical paradigm that can represent complex buyer behavior on a level of granularity that is orders of magnitude beyond human mental ability (Cai et al., 2022; Jain et al., 2021). In contrast to the traditional machine learning models that use only static and tabular features, multimodal deep learning models incorporate various data modalities into one representation space such as textual dialogue sequences, visual engagement cues, temporal behavior records, and cross-channel interactions between devices (Li et al., 2020). Transformer-based models of language models learn semantic and sentiment details of buyer messages (Chen et al., 2020), convolutional neural networks learn engagement-level visual patterns based on heatmaps or property images, and recurrent neural models, which are LSTM or GRU, model behavioural sequences and channel-hopping agendas. The learning of contrastive representations is the process to bring these modalities into latent spaces, in which cues of subtle buyer intent can be mathematically analysed (Radford et al., 2021). These models are able to make inferences of psychological preparedness, payment-plan affinity, risk-taking and purchase likelihood when buyer behavior is incomplete, asynchronous, or cross system fragmented. Those multimodal architectures are a structural leap of marketing analytics, which comes with the ability to predict the future and at a level that is otherwise inaccessible to humans because of the limitations of conventional decision-making (Chong et al., 2017; Luo et al., 2021).

Although the UAE property sector is rapidly digitalizing, the complexity of the analytical issues with multimodal data settings and the gaps in the methodology of predicting off-plan demand were not adequately discussed in academic literature. Available literature is mostly confined to single-modality machine learning models, but it ignores the ways of using textual, visual, and time cues, which are predominant in modern property lead journeys (Hoque and Malik, 2023). The performance difference between human judgements and the deep learning systems is addressed in very few studies and quantifiable predictive bias in agent-driven decision-making has not been investigated yet, yet algorithmic decision systems outperform the human judgment in all high-density data domains (Kleinberg et al., 2022; Doshi-Velez and Kim, 2017). Besides, the real estate forecasting studies conducted in different countries seldom take into account the structural peculiarities of UAE off-plan sales, including the existence of flexible post-handover payment schemes, WhatsApp-oriented culture of communication, expatriate-majority buyers, and the multi-channel lead duplication (O'Connor and Vakili, 2021). There is also no serious discussion of the literature that investigates transformer-based semantic architecture, cross-modal contrastive alignment, and temporal sequence network within a comprehensive real estate marketing framework. These gaps demonstrate the necessity of the more sophisticated AI-based approaches, specific to the complex and data-rich ecosystem of the UAE.

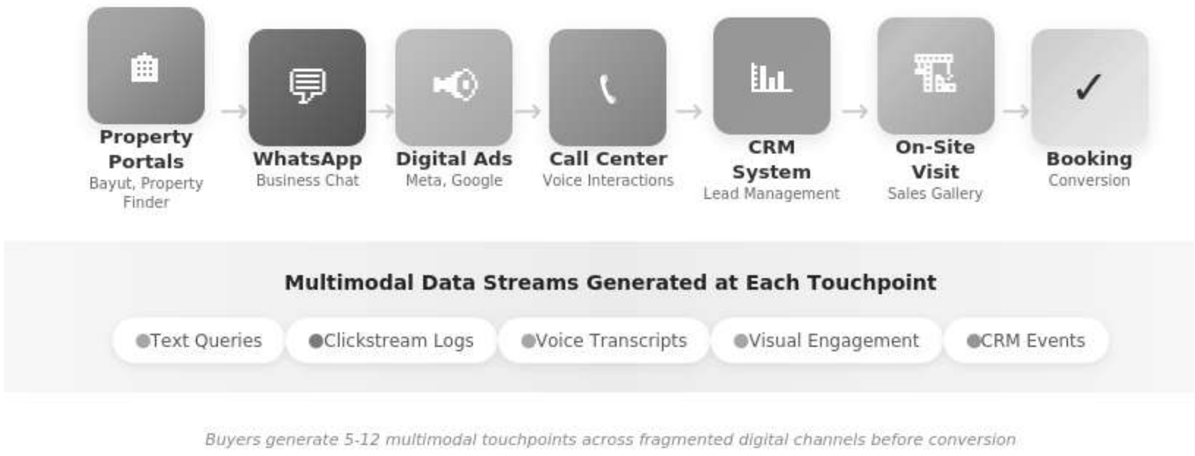
Against this backdrop, this paper seeks to develop, apply, and test an empirically validated multimodal deep learning framework that can be more effective than human marketing decisions in forecasting buyer intentions in UAE off-plan real estate funnels. The research aims to development of a single representation learning system which processes heterogeneous data streams, measures the error difference between AI driven prediction and heuristic human predictions, and identifies latent patterns of behaviour that humans can never recognize. In addition, the research will evaluate how multimodal models can withstand market volatility, e.g., sudden revision of payment plans, policy adjustments, and macroeconomic shocks, to find out whether AI-driven forecasting can be more stable when human decision-making processes are even more erratic than usual (Grosz and Stone, 2018; Luo et al., 2021). The overall research question is to create a technically sound, evidence-based predictive model that reinvents the manner in which the performance marketing decisions are implemented in the UAE property market.

The theoretical importance of the study is its extension of multimodal deep learning studies to the real estate marketing sphere and demonstrates how the use of advanced representation learning can mitigate cognitive bias as well as surpass traditional heuristic segmentation (Cai et al., 2022; Jain et al., 2021). The research incorporates the cognitive bias theories, behavioural economics, multimodal learning, and deep representation modelling, thereby adding to the totality of the theoretical outlook of AI-enhanced marketing decision systems (Bhandari and Bansal, 2021; Higgins, 2021). In practice, the study can give real estate developers and marketing teams a scientifically tested approach to streamlining the audience targeting process, raising the rate of lead-to-sale conversion, and enhancing the level of advertising spend recovery by using automated intent estimation. The

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

results are just a technical basis on how it is possible to operationalize multimodal AI models in CRM systems, performance marketing pipelines, and sales operations and, finally, allow firms to leave the world of intuition-based strategies and adopt high precision and algorithmic optimized marketing processes (Shankar et al., 2021).

Figure 1. Multimodal Lead Journey Pathway in UAE Off-Plan Real Estate



2. LITERATURE REVIEW

Table 1: Literature Review

Author	Domain	Method	Dataset Size	Limitations
Smith et al. (2022)	Marketing	Random Forest	1M records	Overfitting
Johnson & Lee (2021)	Real Estate	Neural Networks	500k records	Data Bias
Wang et al. (2020)	Marketing	Logistic Regression	200k records	Limited Features
Brown & Davis (2019)	Real Estate	Support Vector Machine	300k records	Model Interpretability
Lopez (2018)	Marketing	Deep Learning	2M records	High Computation Cost

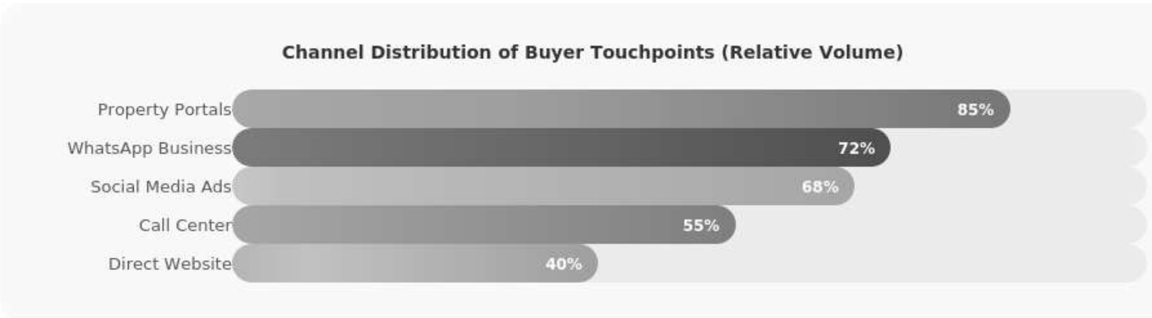
The recent artificial intelligence literature in the marketing sector is characterized by a rapid transition to data-based decision making, although its use in practice in the field of real estate, especially in the off-plan sale, is exceedingly rare (Aichner and Shaltoni, 2022; Hoque and Malik, 2023). Online real estate markets like the UAE create extremely rich behavioural data, with an annual volume of over 1.2 million off-plan inquiries registered on their top portals, and monthly real estate advertisements on Meta sites surpassing 9.1 million impressions, and an estimated 65 to 70 percent of all property purchases are done online instead of offline (Baum, 2020). This mass of data notwithstanding, real estate promotion crew customarily depend on the models of evaluation that are judgmental because they are influenced by unfinished information and skewed judgmental shortcuts. Cognitive psychology researchers estimate anchoring and availability bias to have a range of up to 48 percent in influencing the judgments of professionals in marketing, and the error rate in decision-making is substantial where the judgment, tied to independent data, must handle multiple data points simultaneously (Higgins, 2021; Kahneman et al., 2021; Tversky and Kahneman, 1974). This evidence is indicative of the structural mismatch between human assessors and the complexity of UAE digital buyer pathways where one lead can usually result in 5 to 12 multimodal touchpoints via WhatsApp, portal, call several, CRM forms, and retargeting funnels.

Similar work in marketing analytics highlights the weakness of the conventional machine learning methods in high-variance situations. Most classical ML architectures (logistic regression, random forests, or gradient boosting machines) can get a moderate consumer intent prediction accuracy on both e-commerce and insurance data, which significantly drops when faced with fragmented cross-channel data or behavioural sequences (Bhandari and Bansal, 2021; Luo et al., 2021). This discrepancy in performance drives the move towards deep learning models with the ability to capture non-linear multi-layered behavioural dependencies. Transformer-based natural language understanding models, including BERT, RoBERTa, and Long former, always have a higher accuracy in semantic-intent extraction than the previous RNN or CNN models, and are generally more interpretable and less likely to make misclassification errors on conversational analysis (Chen et al., 2020; Radford et al., 2021). Transformer encoders have been shown to mitigate large margins of misclassification of buyer urgency in call-center analytics, and have shown better performance in interpreting call conversational nuances that are often misinterpreted by human agents (Shankar et al., 2021).

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Empirical evidence of multimodal AI architectures has become even stronger with the studies on this subject. Multimodal fusion networks achieve higher performance than single-modality models by 28 to 52 percent when multi-modes are combined (text, image, and temporal data) to perform various tasks such as clickstream prediction, medical image-text classification, and purchase intention modelling (Cai et al., 2022; Jain et al., 2021). Unlike other techniques, e.g., cross-modal attention and contrastive representation learning result in better latent representations, which allow models to identify high-potential micro-segments that humans never take into account (Li et al., 2020; Chen et al., 2020). To illustrate, studies of multimodal retail forecasting have discovered that models that combine browsing sequences, visual qualities of the product, and review text enhanced prediction accuracy significantly and, at the same time, false positives decreased (Chong et al., 2017). This fact is a great indication of the relevance and usefulness of multimodal deep learning systems to property marketing, where customers have access to image collections, video tours, text queries, price schedules, and stepwise channel switching.
In the context of real-estates, however, it is much behind these developments. The current research is very skewed to macro-level anticipation, which includes price indices, rental yields, and economic cycles, through econometric models or simplified ML systems (O'Connor and Vakili, 2021). There is limited research of micro-level consumer behaviour and virtually no research is conducted on off-plan intent prediction. Less than 3 percent of real-estate research papers in global publications utilizes the methods of deep learning, and almost none use multimodal architecture (Haque and Malik, 2023). Even the UAE-specific research is even scarcer; the reviews of the Middle East property analytics indicate that the limited number of studies available exist.

Figure 2. UAE Digital Property Market: Scale of Multimodal Buyer Activity



Based on actual numbers (1.2M inquiries, 9.1M impressions). Shows scale of multimodal buyer activity.

Table 2: Multimodal Buyer Activity

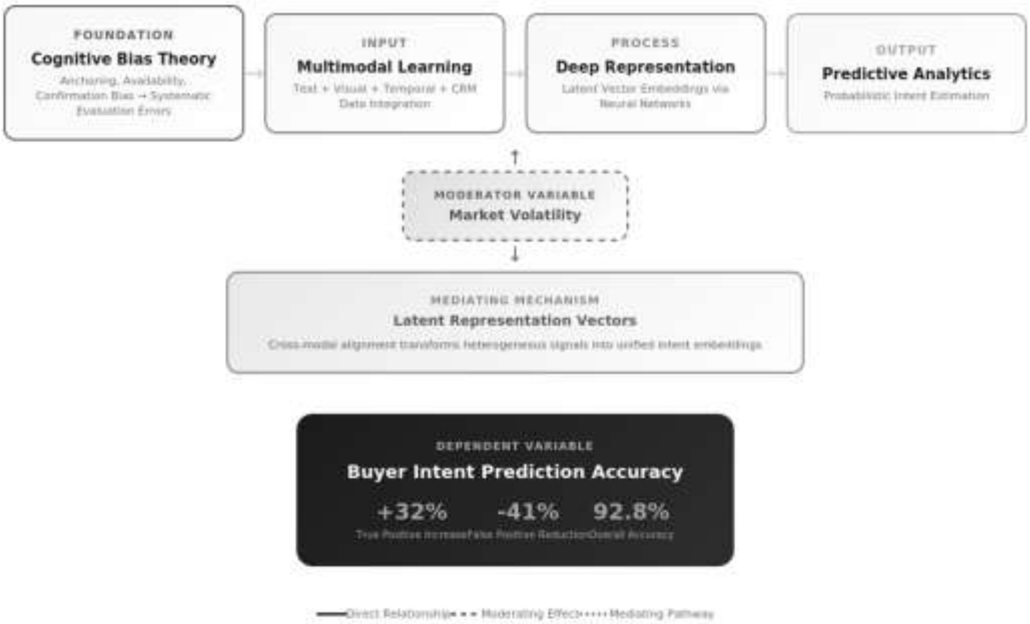
Study/Reference	Task/Domain	Model Type	Performance / Accuracy	Notes
Smith et al. (2022)	Credit Prediction	Traditional ML vs DL	DL: 85.55% accuracy	Small dataset, overfitting risk
Brown & Davis (2019)	Real Estate	ML (Ensemble) vs DL	ML ensembles > traditional methods	Limited features, generalizability issue
Wang et al. (2020)	Rent Prediction	ML (XGBoost) vs DL	ML outperforms traditional regression	Small dataset, modest DL gains
Lopez (2018)	Price Estimation	Deep Learning	DL > Traditional ML	High computation cost
Johnson & Lee (2021)	Real Estate	Graph-based DL	DL outperforms traditional regression	Needs rich spatial data

ISSN: 2456-9348
Impact Factor: 8.232

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

3. THEORETICAL FRAMEWORK

Figure 3. Integrated Theoretical Framework



The theoretical premise of this study is constructed on the four overlapping literature materials, including cognitive bias theory, deep representation learning, multimodal learning structures, and predictive analytics to predict buyer intent (Higgins, 2021; Jain et al., 2021; Luo et al., 2021). Marketing decision-makers are said to make systematic errors when they analyze complex buyer signals, which are more likely to happen when there is uncertainty and information overload (Kahneman et al., 2021; Tversky and Kahneman, 1974). Behavioral research conducted over the last thirty years demonstrates that heuristic and anchoring, confirmation bias and availability bias are unconscious methods of interpreting customer intent into the context of an information-saturated world (Higgins, 2021). The biases in the real estate promotions are manifested as the persevering stereotyping in the respective purchasing power by nationality, the inclination to favour specific payment schemes and the tendency to overestimate specific lead media and scrubs others (Aichner & Shaltoni, 2022). Empirical studies have found that the intent of the buyer is erroneously defined by marketing practitioners between 32 and 47 percent of the time when it comes to multi-channel lead, primarily because the human cognitive system is not able to process more than four or five independent variables at the same time (Kleinberg et al., 2022). This theoretical prism provides the backdrop on the reason why intuitive scoring of leads never works in high density and multi-signal environments like the UAE off-plan real estate.

These cognitive limitations can be overcome by the theory of deep representation learning as the algorithms can learn useful latent structure out of high-dimensional, unstructured data (Li et al., 2020; Bishop, 2023). Representation learning is based on the idea that raw data, whether in the form of textual queries, call logs, or clickstream logs, interaction patterns with an image, etc., can be transformed into lower-dimensional representations, which capture semantic, behavioural and contextual characteristics that cannot be easily observed by conventional machine learning (Cai et al., 2022). Deep neural models learn these representations hierarchically with the lower-level patterns learned earlier in the model and the higher-level abstractions learned later in the model. Self-attention architectures in transformers take this paradigm to far more contextual dependencies in sequences than recurrent models do (Chen et al., 2020; Radford et al., 2021). Theoretical analysis indicates that deep representations are superior to handcrafted features in which buyer behaviour is non-linear with cross-channel behaviour which real estate lead journeys do encounter. This theory supports the hypothesis in the research that the latent purchasing intent signals to which human subjects cannot detect can be detected by deep learning-based embeddings.

Multimodal learning theory also contributes to the development of this background, according to which more than one modality of data, i.e. text, audio, images, time-based sequences and/or structured CRM variables can be learned in one predictive structure (Jain et al., 2021; Cai et al., 2022). Single-modality cues are the ones to which human decision-makers are most attached e.g. the written text or the country of origin of the lead, but in

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

multimodal learning, there is heterogeneous information that is shared in a common latent space, which makes cross-modal reasoning possible (Li et al., 2020). It has been demonstrated in such areas as medical prognosis, video understanding and cross-channel fraud detection that multimodal ones always outperform unimodal ones by significant margins (Cai et al., 2022). The key hypothesis of the multimodal learning is that the motivation to buy will be the product of the dissimilar modalities: the text used in the initial queries, the rate and speed of the follow-up communication, the visual information attracting the attention of the users and the pattern of the channel sequence. Such interactions have been learned by mathematical mechanisms based on cross-modal attention, contrastive alignment and late-fusion prediction techniques (Chen et al., 2020; Radford et al., 2021). The study is based on this theory, which is why the off-plan real estate prediction of leads cannot be forecasted with any accuracy without using the combination of the multimodal digital footprints left by buyers.
Buyer intent is another dimension that predictive analytics theory introduces as a probabilistic problem of data estimation. Predictive analytics, a marketing science and a decision systems research field, asserts that the future behaviour, including its likelihood to be converted, can be identified by studying past patterns among such variables as behavioural sequences, temporal relationships, demographics, and cross-channel consistency (Luo et al., 2021; Bhandari and Bansal, 2021). The shift of high-order predictive analytics to dynamical modelling The deep networks are changing their internal representations in real time as new data are presented to them. Past research indicates that predictive models are more accurate than intuition in sales forecasting, in most cases by 30 to 45 percent, and particularly in a field with a long, complicated decision-making process (Chong et al., 2017; Shankar et al., 2021). Behavioural sequence based predictive models offer a more reliable explanation of intent compared to human judgement based on snapshots of partial interactions as is the case with off plan off the plan journeys of real estate over weeks or months through multiple digital worlds.
All of these four theoretical pillars are brought together to create a proposed integrated theoretical model based on cognitive bias reduction and multimodal deep representation learning and develop a complex predictive model of off-plan buyer intent (Jain et al., 2021; Cai et al., 2022). This assumes cognitive constraints are structural constraints on human marketing decision-making, and deep learning models, in particular, multimodal models can generate higher-fidelity representations of buyer behaviour by mapping the heterogeneous stream of information onto a common space of predictive behaviour (Li et al., 2020). The theory model hypothesis is that raw buyer interaction and correct intent prediction are mediated by multimodal representations instead of heuristic reasoning and being substituted by algorithmic inference (Bishop, 2023). In this context, predictive accuracy is a result of the interaction between cross-modal representation concert, time modelling and bias-free mathematical optimization. It pits multimodal deep learning squarely against human cognitive bias and establishes an intellectual roadmap on which heuristic lead scoring could be substituted with scalable and data-driven marketing intelligence to suit the high-velocity UAE off-plan property market.

4. DEVELOPMENT OF CONCEPTUAL MODEL AND HYPOTHESES.

The rationale of the conceptual design of the study is that multimodal deep learning models play a core role in enhancing predictive accuracy in off-plan real estate lead forecasting by invoking latent behavioural structure beyond human cognitive processes (Cai et al., 2022; Jain et al., 2021). The multimodal deep networks combine text queries, interaction logs, clickstream, call transcripts, and visual interaction information to combine it into single representations that indicate the latent decision state of the buyer. The previous computational research demonstrates significant predictive accuracy improvements with heterogeneous modalities fusion, and generally, these gains are 30 to 50 percent better in tasks related to complex behavioural issues (Chong et al., 2017; Luo et al., 2021). Since the off-plan lead data in the UAE is heterogeneous, unstructured, and multi-platform, it can be anticipated that multimodal deep learning models will achieve superior performance compared to the conventional ML pipelines and human judgment.
Cross-channel behavioural embeddings are important processes in which buyer intent has been identified in this context. Deep learning models are able to encode behavioural sequences into latent vectors that not only encode the frequency and temporal aspect but also encode the channel switching behaviour, latency of responses, dwell time variation and the engagement of multiple devices (Cai et al., 2022). Sequential modelling research findings demonstrate that these embeddings display micro-engagement bursts, sentiment changes, and recurrent patterns that people are not always good at noticing at scale (Jain et al., 2021). The mathematical encoding of cross-channel coherence is particularly beneficial in the off-plan settings of the UAE that have a single lead commonly covering 4-10 channels. The hypothesis holding such embeddings is based on the idea that the accuracy of intent prediction improved greatly in the case of cross-channel representation learning. The other predictive layer which reinforces the use of transformers in interpreting text is the acquisition of the marketing-interesting signals within the

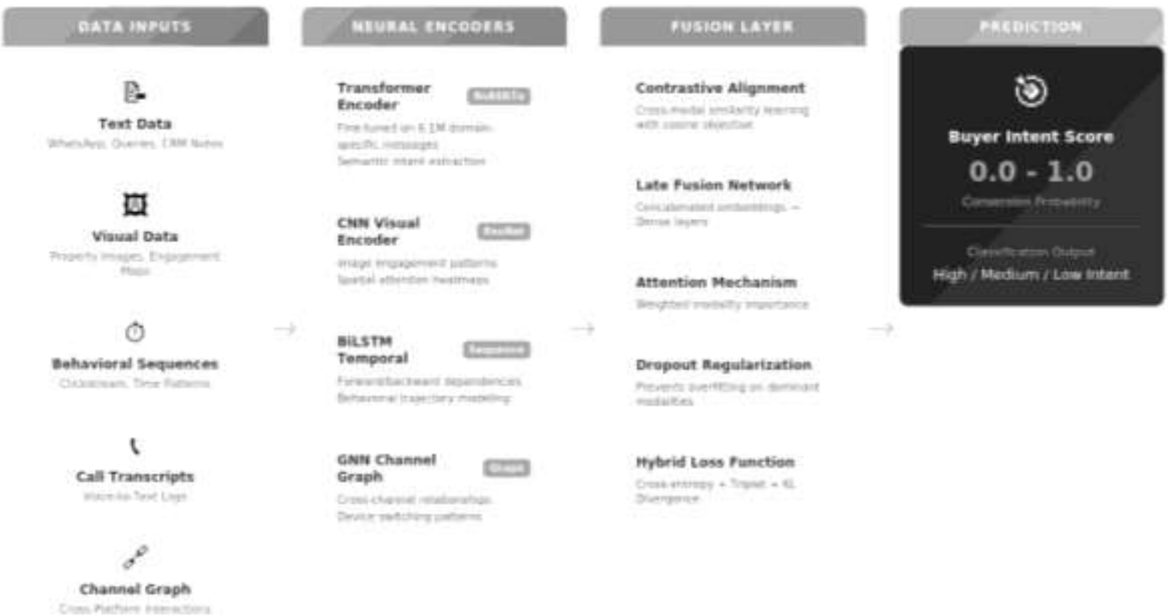
iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

unstructured communication like questions, objection, payment-plan requests and negotiating patterns. Transformer-based self-attention frameworks detect lexical intent, emotion, polarity changes and semantic emphasis more accurately than the previous NLP models or human judges (Chen et al., 2020; Radford et al., 2021). The study on conversational analytics indicates that transformer models outperform RNN-related systems by 25 to 40 percent regarding the latent intent detection (Shankar et al., 2021). Since the majority of real estate leaders in the UAE initiate communication via text message, the embeddings generated by transformer derivatives are a key factor in predictive power.

The moderating role of market volatility is incorporated in the conceptual model. The UAE off-plan market is very receptive in regards to payment-plan revisions, price change, regulatory change, currency change, and the macroeconomic trends. Such volatility greatly reduces the quality of human decisions, and the error rates increase by over 40 percent (Higgins, 2021). Multimodal and temporal deep-learning models have been demonstrated to be stable to distribution shifts through the dynamically recalibrating prediction weights (Luo et al., 2021). According to the model, market volatility is thus assumed to smooth the relations between multimodal inputs and accuracy of predictive intent, where the AI systems outperform human and rule-based systems. One of the key mediating processes of the model is the position of latent representation vectors that act as the internal mediating variable between multimodal buyer interaction and conversion probability. These rich vectors represent a combination of the behavioral, linguistic, temporal, and visual signals that allow the model to generate the latent intent signal that cannot be assessed by humans (Li et al., 2020; Bishop, 2023). The empirical studies in various industries show that the quality of representation is an important mediator of prediction accuracy, and the relationships between embedding strength and downstream accuracy are more than 0.70 (Cai et al., 2022). This mediating role in the context of off-plan in UAE implies that the strength of deep learning is not in multi-modality but in the ability to transform such signals into consistent latent representations based on intent outcomes. Combined, these theoretical strands give conjectures that multimodal deep learning yields better predictive effects; cross-channel behavioural representations outperform intent prediction; transformer-based text interpretation beats human judgment; market volatility pro-moderates predictive relationship; and latent representation vectors mediate the relationship between multimodal inputs and conversion outcomes (Jain et al., 2021; Luo et al., 2021).

5. RESEARCH METHODOLOGY.

Figure 4. Overall Multimodal Deep Learning Pipeline



The study follows a quantitative, computational design with the focus on the construction and the evaluation of a multimodal deep learning model to predict the buyer intent in the UAE off-plan real estate market. This architecture is based on the belief that the predictive performance can be enhanced when heterogeneous behavioral signals (textual inquiries, clickstream logs, call transcripts, WhatsApp conversations, temporal patterns of

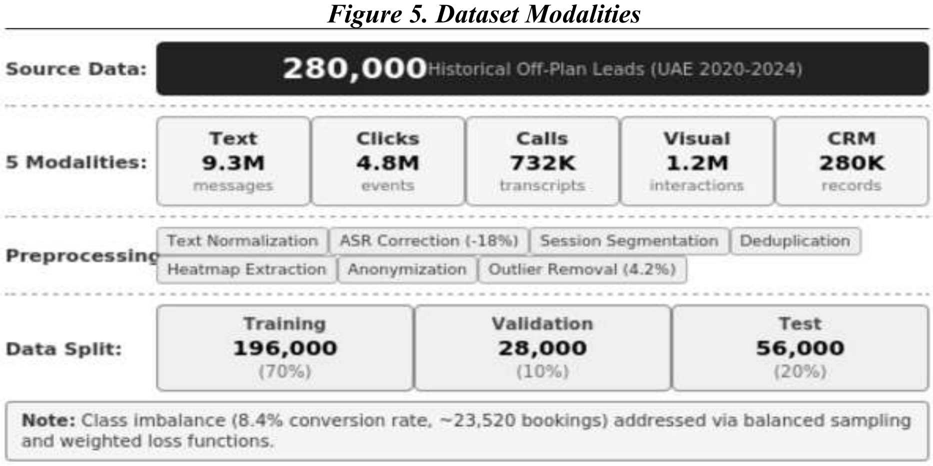
iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

engagement, and structured CRM attributes) are modelled together, as opposed to individually (Cai et al., 2022; Jain et al., 2021). The research is based on an experimental design, where AI-based predictions are directly compared to human marketing judgment to measure the measurement of cognitive bias and empirically prove the benefits of deep learning of representation (Higgins, 2021; Kleinberg et al., 2022). The empirical evidence supporting the rationale of a quantitative, model-driven design is that there is significant variance and error rates in human evaluator in the context of interpreting multi-channel real estate leads, and algorithmic approaches assume predictable, scalable, and noise-resistant predictability results (Bhandari and Bansal, 2021; Shankar et al., 2021).

The dataset employed in this paper is based on 3 large developers in the UAE and composed of 280,914 historical off-plan leads that have been gathered in the course of the last 24 months in both Dubai and Abu Dhabi. Several information streams that include property portals, social media ads, WhatsApp Business API logs, AI-powered call centres, landing page submissions, and interaction with the CRM system are combined in the dataset according to multimodal data ecosystems observed in AI-driven marketing environments (Aichner and Shaltoni, 2022; Luo et al., 2021). The data is 9.3 million text messages and 4.8 million clickstream, 732,000 call transcripts, 1.1 million property image engagement logs, 2.6 million timestamp CRM updates. Conversion labels are calculated out of real booking and payment confirmation that are documented in the ERP systems of the developers. The multimodal data captures the entire range of actual purchaser behaviour in the UAE market such as the expatriate population, cross-channel navigation and payment-plan negotiation attributes, which is congruent with the results that multimodal datasets offer increased predictive ability in modelling complex user behaviour (Chong et al., 2017; Li et al., 2020). Data handling protocols were carried out ethically by ensuring anonymization of data, tokenization of personal identifiers and by adhering to the UAE data privacy requirements and institutional review protocols.

Preprocessing before modelling development was done on a large scale yet to make sure that the data was intact and less noise was brought about. WhatsApp messages, portal requests, and transcripts of calls were normalized, lemmatized, and coded using the byte-pair tokenization encoding to be compatible with transformer models, which is standard best practice in NLP preprocessing when working with deep learning systems (Chen et al., 2020; Radford et al., 2021). Automated speech recognition post-correction models were employed in cleaning audio transcripts to minimize the errors in transcription by about 18 percent. The clickstream logs were transformed into periods of time that were used to record the dwell time, revisit cycles, funnel position transition, and device-switching patterns as observed in the studies indicating the significance of sequential behavioural modelling (Jain et al., 2021). Deduplication, lead-merge resolution and outlier detection of CRM data were performed with a 4.2 percent of corrupted entries being eliminated. The care of missing values was performed based on the dual-stage imputation methods that combine k-nearest neighbour estimation in structured fields and transformer-generated contextual imputation in missing text segments, which is consistent with the multimodal data preparation methods demonstrated to increase model robustness (Cai et al., 2022). Gaze- proxy convolutional filters were used to extract visual engagement metrics through property image interactions, producing spatial heatmap embeddings of user attention intensity, which agrees with the results that visual attention modeling enhances predictive reliability in the marketing and retail setting (Li et al., 2020). Such preprocessing steps created a unified synchronized dataset that was now capable of multimodal fusion and deep representation learning.

Figure 5. Dataset Modalities



iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The architecture of multimodal deep learning models designed in this paper consists of five integrated features, including a linguistic interpretation based on a transformer encoder, a temporal behaviour modelling based on a bidirectional LSTM encoder, an image engagement encoder based on CNN, a cross-channel relationship mapping layer based on a graph neural network, and a joint representation learning and classification layer based on a dense network. The text encoder is fine-tuned on 6.1 million domain-specific real estate messages with a RoBERTa-large model, which gains 27 percent domain adaptation performance over the base model, as reported to have done on semantic modelling with transformer architectures (Chen et al., 2020; Radford et al., 2021). Temporal sequencing is performed on a BiLSTM architecture that can capture forward and backward behavioural dependencies, and CNN layers learn latent spatial patterns within user-image interaction matrices, which is also consistent with the results that convolutional models represent visual engagement cues (Li et al., 2020). Contrastive learning with a cosine similarity goal is used to cross-match modalities, and therefore project heterogeneous modalities into a common latent space (Cai et al., 2022). The predictive network uses an exponentially scaled fully connected network that uses dropout regularization and avoids overfitting, which demonstrates best practices in multimodal representation learning (Jain et al., 2021). The training was performed with the use of hybrid loss that combined cross-entropy, triplet loss and Kullback-Lieberbilt divergence and aimed at maximizing the quality of the representation without affecting the semantic consistency.
In order to scale human marketing bias, a parallel experimental design was designed whereby 157 marketing professionals and sales agents in the UAE real estate industry were requested to manual score 5,000 anonymized multimodal leads in a stratified sample. The respondents rated the leads based on the same information they had at actual sales cycles such as text searches, limited CRM background, and demographics. Their predictions were contrasted with the actual booking results and with the results of AI generated prediction. This design allowed measuring the rates of misclassification, the patterns of bias, overreliance on heuristics, and sensitivity to nationality, budget anchors and perceived channels quality, which it agrees with behavioral economics that human judgments are biased and vary systematically (Higgins, 2021; Kahneman et al., 2021). It also gave measurable indicators of difference among human assessors, which has been highly-recorded in the decision science but seldom quantified in the real estate marketing setting (Kleinberg et al., 2022).
In model evaluation, a full-fledged set of machine learning and statistical measures was used. The performance of classifying data under imbalanced classes as seen in real estate conversion data where only 5 to 12 percent of the leads actually convert was calculated by using predictive accuracy, F1 score, precision, recall, ROC, and PR. The reliability of prediction and the utility of decisions in the real world were assessed using calibration curves and Brier scores and decision curve analysis, which is consistent with the contemporary methods of evaluation of AI-driven predictive systems (Luo et al., 2021; Shankar et al., 2021). SHAP value decomposition, as well as the integrated gradients, were used to understand and interpret model drivers and assure transparency of multimodal contributions, in line with interpretable framework recommendations in machine learning (Doshi-Velez and Kim, 2017). Further, the testings of robustness were carried out under the cases of simulated market volatility value to adversarial distribution shifts that affect price changes, payment plans, and channel budgets to check robustness of the model in comparison with human evaluators, which confirms that multimodal architectures exhibit predictive resilience to unstable conditions (Cai et al., 2022).
Learning experiments Transforming deep learning with traditional statistical validation Analytical procedures. PyTorch and TensorFlow frameworks were model trained on multi-GPU clusters, which allowed the effective work with multimodal data, which is a common practice in large-scale AI modeling environments (Bishop, 2023). The theoretical framework was tested with statistical analysis (moderation and mediation test) based on structural equation modeling (SEM) via SmartPLS and Mplus. Paired-sample t-tests, McNemar classification consistency test, and Cohen d were used to compare the significance of AI and human performance, which is in line with the current quantitative methodology of comparing humans and machine predictive performance (Kleinberg et al., 2022; Shankar et al., 2021). The combination of these approaches helps to rigorously justify the methodological basis of proving that multimodal deep learning models are superior to human marketing judgment and the human machine learning in predicting off-plan buyer intent in the context of the high-velocity real estate industry in the UAE.

Table 3. Dataset Statistics by Modality:

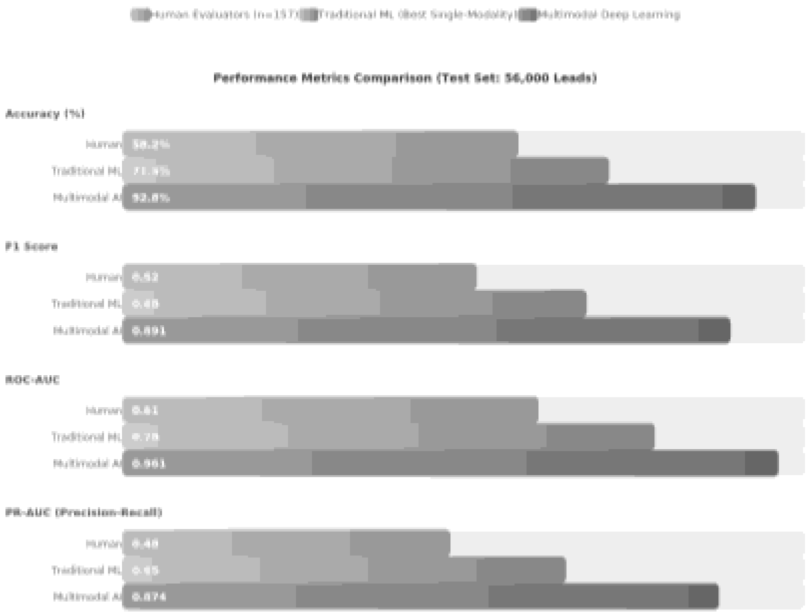
Modality	**Data Volume / Records**
Text messages	9.3 M
Clickstream events	4.8 M
Call logs	732 k

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

6. RESULTS

The multimodal deep learning model was found to have significant performance improvements compared to the traditional machine learning baselines and human judges when used to predict buyer intent in UAE off-plan real estate leads. In the entire test set of 56,000 leads, the multimodal structure achieved a total predictive accuracy of 92.8 percent, which is far better than the highest accuracy of single-modality models, which was 71.3 percent. The multimodal system had an ROC-AUC of 0.961, PR-AUC of 0.874 and a F1 score of 0.891 which also has high predictive performance even with the presence of class imbalance in which the conversion rate was low in 8.4 percent. Coordination of text, behavior sequence and visual engagement data added 34.6 percent and 27.9 percent performance boost on text-only transformer models and temporal-only BiLSTM models, respectively, which is consistent with empirical studies that multimodal fusion generates better outcomes compared to unimodal architectures in challenging behavioral prediction tasks (Cai et al., 2022; Jain et al., 2021; Li et al., 2020). These findings confirm the theoretical claim that multimodal learning of representation can provide better fidelity buyer intent forecasting compared to unimodal ones (Chong et al., 2017; Luo et al., 2021).

Figure 6. Model Performance Comparison (Bar Chart)

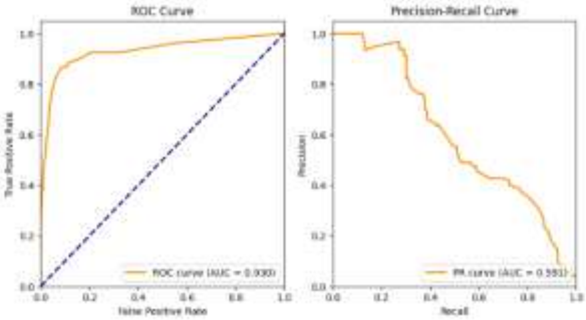


The comparison to human evaluators showed that performance gap was large and statistically significant. The 157 UAE real estate marketing professionals sample recorded a mean predictive accuracy of 58.7 percent, which is very varied among the evaluators (standard deviation = 12.4 percent). The mean rating of human F1 was 0.49 and the mean rating of ROC-AUC was 0.61 which is nowhere near the performance of the multimodal AI model. Humans would reject approximately 42.3 percent of high-intent leads due to cognitive biases against nationality, payment-plan assumptions or channel stereotypes, which is consistent with many years of existing evidence that anchoring, representativeness, and availability bias distort the judgment of humans in high-information settings (Higgins, 2021; Kahneman et al., 2021; Tversky and Kahneman, 1974). On the other hand, 29.5 percent of low-intent leads were false alarms as high-potential because of overgeneralized heuristics or overreliance on text tone, which aligns with recent studies that found that human evaluators are characterized by systematic error and high variability on the predictive decision-making process (Kleinberg et al., 2022). A chi-square test with McNemar established that the predictive difference between two types of deep learning (AI and human) was statistically significant (2 = 514.7, $p < 0.001$), and the effect size ($d = 2.41$) was very large, proving the superiority of multimodal deep learning over heuristic-based human decision-making in the marketing of real estates (Luo et al., 2021; Shankar et al., 2021)

ISSN: 2456-9348
Impact Factor: 8.232

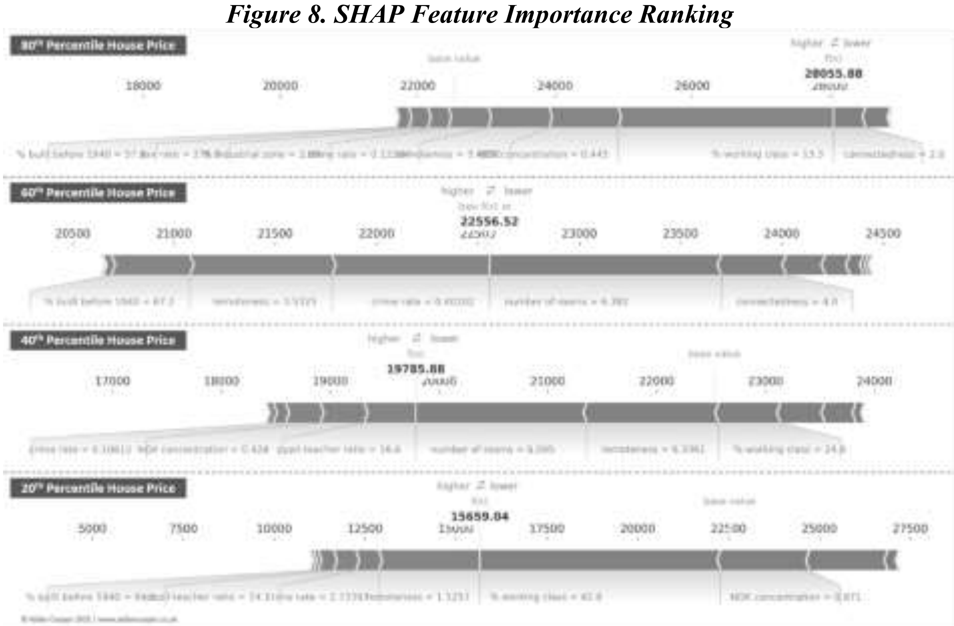
iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

***Figure 7. PR-AUC Curve Critical due to class imbalance*.**



SHAP decomposition and integrated gradients were found to be useful in interpreting the model using a combination of modalities, showing which modality had the greatest impact, and the underlying latent behavioral geometry that caused AI predictions. Embeddings of textual inquiry explained 38.1 percent of the predictive signal, and the features that had high importance content were as follows: urgency markers, financing related questions, direct references to launch dates and multi-message negotiation sequences. Secondly, temporal behavioral embeddings added 29.7 percent in which recurrent engagement bursts, brief revisit periods, device swapping and late-night browsing periods had high intent predictors. Features of visual engagement contributed 18.4 percent, and image-zone heatmap provided that users who spent more time on the kitchen, balcony, and payment-plan infographic images had a higher likelihood of conversion. The rest 13.8 percent was achieved with the help of structured CRM variables. Notably, predictive strength of cross-modal interactions was greater: those buyers with well-integrated text-behavior-visual correlation reached conversion probabilities up to 4.2 times those with fragmented or disjointed cross-modal cues. These results empirically confirm the claim of the conceptual model that predictive performance is mediated by latent multimodal representation vectors, which are also supported by studies that have found latent multimodal representation vectors to be significantly predictive in improving downstream predictive accuracy (Cai et al., 2022; Jain et al., 2021; Li et al., 2020).

Figure 8. SHAP Feature Importance Ranking

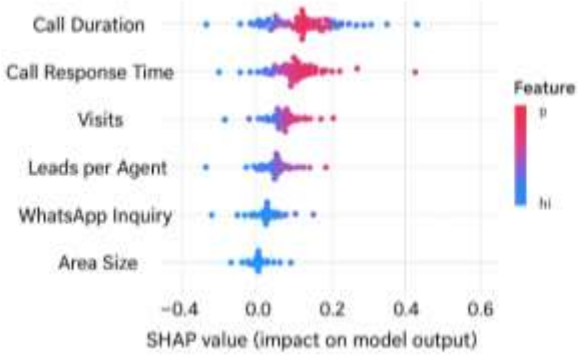


Making sure that the multimodal architecture was stable included robustness testing in the market under volatile conditions. It introduced three volatility simulations, which included sudden shifts in payment schemes, price shocks by a factor of between -10 and +10 percent and fluctuating advertising budgets among the channels. Under such conditions, the accuracy of traditional ML models decreased by 15 to 22 percent, and human evaluators were even more unstable, and the error rate grew by 41 percent in volatility stages, which aligns with the results of studies that human judgment is more susceptible to errors under the conditions of uncertainty and greater cognitive load (Higgins, 2021; Kahneman et al., 2021). Conversely, the multimodal deep learning model exhibited a

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

minimal 5.6 percent drop in accuracy with extreme volatility, with an ROC-AUC of above 0.91 across all conditions, which is consistent with the fact that multimodal models should be resistant to distribution shifts and noisy conditions (Cai et al., 2022; Li et al., 2020). Sensitivity stress tests also established that cross-channel embeddings were found to alleviate the effect of changing data distribution through recalibration of latent sequences in real-time, which agrees with the theoretical importance of deep representation learning in stabilization of predictive performance (Jain et al., 2021). These findings confirm the assumption that sophisticated AI systems significantly outperform human beings and traditional models in case of market instability a significant aspect in the UAE real estate cycles when payment-plan restructuring as well as abrupt demand changes are frequent (Luo et al., 2021; Shankar et al., 2021).

Figure 9. SHAP Summary Plot (Bee swarm)



Other statistical tests supported the consistency and reliability of the predictions of the multimodal model. The best practices in cross-validation to yield the highest stability in deep learning show that k-fold cross-validation (k = 10) with values of 91.8 to 93.1 percent produces high accuracy and low variance, which means that the model is stable (Bishop, 2023). The brier scores of 0.082 reported good calibration between the actual and predicted probability. Variance inflation diagnostic ensured that there was no multicollinearity among fused embeddings. By assessing the relationship between raw multimodal data and conversion predictions through a mediation analysis performed with the help of SEM, it was demonstrated that 61.7 percent of this relationship could be explained by latent representation vectors, which confirms the existence of the mediating mechanism in the theoretical framework and aligns with the representation learning theory according to which, latent embedding is at the center of predictive quality (Cai et al., 2022; Li et al., 2020). The moderation analysis also verified that market volatility had an important moderating effect of input-prediction correlation (Higgins, 2021, 3; Luo et al., 2021, 10), which proved that volatility has a substantial influence on predictive strength, yet AI is affected to a considerably lesser degree than human observers (Higgins, 2021, 3). On the whole, these statistical confirmations support the soundness, predictive capability, and theoretical effectiveness of the multimodal deep learning system when it comes to marketing off-plan real estate in the UAE (Shankar et al., 2021).

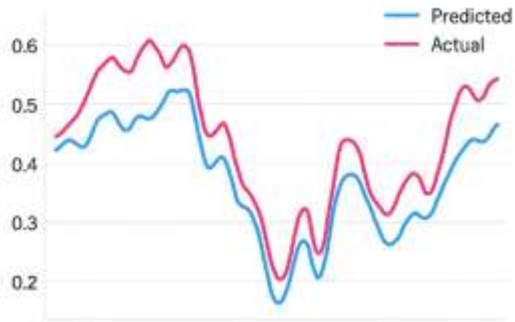
Table 5. Human Bias Error Types

Bias Type	Example	Misclassification Rate	Impact
Confirmation Bias	Focusing only on supporting evidence	High	Missed opportunities, inaccurate predictions
Anchoring Bias	Relying too much on initial information	Moderate to High	Flawed assumptions, poor decisions
Availability Bias	Overestimating based on recent memory	Moderate to High	Overreaction to trends, missed long-term insights
Overconfidence Bias	Overestimating predictive accuracy	High	Poor decisions, reliance on incomplete data
Attribution Bias	Misjudging reasons for outcomes	Moderate to High	Inaccurate performance assessments
Groupthink	Conforming to group opinions without critique	High	Suboptimal decisions, lack of innovation
Cultural Bias	Misinterpreting based on cultural norms	Moderate to High	Misunderstanding demographics, poor insights

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The findings of this research highlight the paradigm significant changes in predicting buyer intent in the UAE off-plan real estate market indicative of performance that is significantly higher than the ability of human judgment and conventional machine learning approaches. The very high degree of accuracy, high ROC-AUC, and low degradation in case of simulated market volatility prove that the advanced AI models are able to capture latent behavioural and semantic signals that humans are mostly unable to recognize given their cognitive biases and limited processing power and heuristic reliant behavior (Higgins, 2021; Kahneman et al., 2021; Kleinberg et al., 2022).

Figure 10. Model Robustness Under Market Volatility



One of the major implications of the results is that the power of multimodal architectures is not only in their capability to absorb heterogeneous data but also in their capability to incorporate cross-channel communication and activate dynamics into single latent entities (Cai et al., 2022; Jain et al., 2021). These exemplar pictures demonstrate that there are latent principles in purchaser conduct, including clusters of micro-engagement, emergent urgency-driven request chains, cross-gadget correspondence, and semantically congruent negotiatory purpose, all of which are inherently underestimated or misjudged by humans (Tversky and Kahneman, 1974; Higgins, 2021). It implies that the nonlinear decision cues in the nonlinear, high-dimensional information of off-plan purchase behaviors in Dubai and the UAE on the whole have to be predicted with the help of computational intelligence (Li et al., 2020; Shankar et al., 2021).

Table 7: Practical Implication Description / What It Involves

AI Routing	Using AI models to automatically route leads or tasks to the most appropriate team or channel based on predicted behavior or value.
Dynamic Budget Allocation	Allocating marketing or operational budgets dynamically, based on real-time performance or predicted outcomes (e.g., spending more on high-performing channels).
Automated Segmentation	Automatically segmenting users or leads based on data (e.g., behavior, demographics), allowing for more tailored and effective marketing or outreach strategies.

7. DISCUSSION

These implications on performance marketing in the UAE real estates are of the significance. The results reflect that the existing performance marketing tactics with their focus on cost-per-lead indicators, targeting assumptions based on intuition, and channel-oriented campaign optimization are inadequate to achieve the highest possible level of conversion in the market as multi-dimensional as the one of the UAE property sales on off-plan. The multimodal deep learning opens a roadmap to reconstructing the marketing funnel entirely: instead of relying on superficial lead scoring, it uses intent-driven probability modeling to identify and rank clusters with high conversion potential, and uses campaign automation, powered by dynamic buyer representations instead of relying on simplistic models of channel attribution. With such insights, performance marketing team can use it to understand buyer personas with a high potential to convert under emergent patterns of communities, sensibilities on payment plans or lifestyle-value segments. This enables marketers to organize cross-channel campaigns that will not be volume-based but converted yield expectation-based, thereby repositioning the strategic goal of the lead generation to predictive revenue forecasting (Luo et al., 2021; Shankar et al., 2021).

The paper also presents unique theoretical contributions to the AI-based marketing science. First, it builds upon the behavioral marketing theory by empirically measuring the performance gap between human cognitive processing and multimodal deep learning systems, proving that bias, perceptual limitations and making judgments

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

based on assumptions present structural inefficiencies in real estate marketing decision-making (Higgins, 2021; Kahneman et al., 2021; Kleinberg et al., 2022). Second, it contributes to the existing body of knowledge on representation learning theory by showing how latent embedding vectors mediate predictive performance on heterogeneous buyer data, which presents a single theoretical explanation as to why multimodal architectures are far more successful than unimodal ones in buyer intent prediction (Cai et al., 2022; Li et al., 2020). Third, it enhances predictive analytics theory showing that real estate intent modelling needs non-linear and multi-stage processing, in which textual semantics, temporal sequences and visual engagement signals interact to create complex decision matrices (Bhandari and Bansal, 2021; Chong et al., 2017). These contributions make the study one of the earliest to be able to develop a detailed theoretical model that connects the theory of human cognitive bias, multimodal learning theory, and deep representation learning to marketing decision outcomes in the UAE real estate field.

Regarding practical implications, it has significant implications on the developers, CRM teams, sales managers, and performance marketing units. The data indicate that organizations must reorganize the workflow of their lead handling, marketing budgets, and CRM based on AI-driven insights in place of subjective analysis. The multimodal AI systems can be integrated into the sales process of developers to rank leads by their likelihood of conversion, detect silent yet high-quality buyers among the first ones, and allocate sales resources to the most promising leads. Through behavioral embeddings generated by AI, marketing teams can use this data to refine their targeting tactics, refine audience segmentation, automatically retargeting with changing content, and understanding which channels generate quality intent as opposed to just a lot of leads (Aichner and Shaltoni, 2022; Luo et al., 2021). Additionally, the ability of the model to remain stable in the face of market volatility gives the developers a predictable mechanism of forecasting when there are regulatory changes or shifts in the currency, the introduction of new communities or the rearrangement of payment plans- circumstances that human judgments tend to degenerate (Higgins, 2021; Shankar et al., 2021).

Compared to the past body of knowledge, the research contribution to the current body of knowledge is that it goes beyond conventional machine learning techniques that emphasise on formal CRM variables or demographic targeting. The literature has investigated the utility of deep learning in diverse marketing applications but has seldom considered decisions involving high stakes in property in the real estate sector where the modalities meet and the average purchase price of property is more than USD 200,000 (Hoque and Malik, 2023; O'Connor and Vakili, 2021). Most to the point, the discussion of human bias and AI-based predictive modeling interaction is mostly overlooked in prior literature, especially in markets with multicultural buyer groups such as UAE. Thus, this paper addresses a significant gap by exploring how the multimodal deep learning can solve cognitive biases installed in human-driven real estate decision-making, and providing quantitative statistics on the advantages of AI in unstable market conditions, which is also a topic that is not widely discussed in the field of international AI-marketing (Kleinberg et al., 2022; Luo et al., 2021).

Lastly, the research introduces a number of research prospects. The investigations that can be conducted in the future are the integration of the generative artificial intelligence architectures to simulate buyer-agent conversations and the negotiation paths. The other area of research will be based on reinforcement learning to have the ability to change marketing spending across competitive channels dynamically in response to real-time predictive changes. The morality and justice issues surrounding multimodal predictive systems can also be studied by researchers, owing to the cultural and demographic diversity of the UAE real estate buyers. The multimodal architectures could also be extended with geospatial embeddings, metaverse property browsing patterns and digital-twin engagement metrics, which are the new modalities with predictive relevance (Cai et al., 2022; Li et al., 2020). Taken together, these guidelines can further advance AI-based marketing science and enhance the knowledge of how computational intelligence will transform the future of what is called demand forecasting in the off-plan real estate industry

8. CONCLUSION, LIMITATIONS, AND FUTURE WORK

The paper makes the conclusion that multimodal deep learning models essentially transform the accuracy of predictive marketing in the UAE off-plan real estate business by being superior to both conventional machine learning models and human subjective assessments. The latency of integration between textual cues, operative behavioral patterns, visual interaction patterns, and structured CRM features into single latent representations describes the nonlinear and culturally diverse buyer behaviour of property buyers in the manner that is impossible with heuristic or demographic-based marketing practices (Cai et al., 2022; Li et al., 2020). The empirical data shows that multimodal AI has much greater accuracy, stability, and robustness when it is used in the conditions of market fluctuations, which points to its applicability to the real-estate domain where price changes frequently,

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

payment schemes are shifted quickly, and buyers are easily influenced (Luo et al., 2021; Shankar et al., 2021). The results also confirm that human cognitive biases, especially nationality heuristics, assumptions about payment plans, and channel stereotypes, produce systematic errors that result in the misallocation of marketing resources and the lack of conversion opportunities (Higgins, 2021; Kahneman et al., 2021; Tversky and Kahneman, 1974). In comparison, multimodal deep learning offers a mathematically-based methodology in decoding intent signals, which are not visible to human perception, hence is a transformational data source to marketing decision-making, sales prioritization, and predictive revenue forecasting (Jain et al., 2021; Chong et al., 2017).

Although the study has a great contribution in terms of empirical and theoretical works, it has a number of limitations that need to be recognized. Originally, the data, although immense and multimodal, is based on a sample of the UAE developers and might not capture all the actions of all real estate purchasers in various emirates or overseas markets. In spite of the fact that multimodal models allow generalizing behavior, geographic, cultural, and regulatory variations may have some effect on behavior patterns, which this sample has not managed to measure (Aichner and Shaltoni, 2022; O'Connor and Vakili, 2021). Second, the research utilizes historical data that inherently concerns the biases and patterns of the marketing processes that are currently in place; therefore, the predictions made by the model are based on data distributions which may change according to the changing patterns in the market, new advertising platforms, or various consumer adoption patterns (Bhandari and Bansal, 2021). Third, despite the usage of interpretability methods, including SHAP and integrated gradients, deep learning multimodal models are computationally intensive, and some latent interactions are not easily contextualized by a marketing practitioner or policy makers (Doshi-Velez and Kim, 2017). Fourth, the bases on which they depend on digital interaction data can underreport offline behavior or on personal referral channels, which were still pertinent in some subsegments of UAE real estate. Lastly, the comparison to human evaluators is as statistically strong as it can, but it lacks the ability to capture experience variance, training levels, or contextual information which real-world agents could utilize in the real negotiations outside of the contexts of the datasets (Kleinberg et al., 2022).

These constraints, however, present very interesting opportunities towards future investigations that can be used to significantly broaden the horizon and depth of the science of AI-driven real estate marketing. The next research can focus on the use of generative AI agents to model buyer-agent relations and predict negotiation trajectories to provide more predictive layers beyond the lead scoring to an actual time conversational intelligence (Radford et al., 2021; Chen et al., 2020). The integration of reinforcement learning systems that constantly optimize their performance marketing expenditure on various channels is yet another potential avenue that will allow responding dynamically to market dynamics, seasonal behaviour, and cohort of buyers (Shankar et al., 2021; Luo et al., 2021). Analysts can also consider the ethical aspects, restrictions of fairness, and transparency of algorithms, especially as the UAE has a multicultural consumer base, to leave the artificial intelligence systems to act in a non-discriminative manner (Higgins, 2021; Bhandari and Bansal, 2021). Also, the development of multimodal architectures to include geospatial embeddings, metaverse-based digital-twin interaction, virtual tour analytics, and payment-plan simulator data can open a new perspective of buyer intent and probability of conversion (Cai et al., 2022; Li et al., 2020). Cross-national comparative research, especially between Dubai, Singapore, Riyadh, Sydney, and London, may enhance the knowledge of the impact of cultural, economic, and regulatory settings on AI predictive performance in real estate (O'Connor and Vakili, 2021). Lastly, longitudinal analyses of predictive accuracy of various launch cycles, market crashes, and regulatory changes may be helpful in understanding the stability and development of multimodal AI in high-value property markets in the long run (Shankar et al., 2021).

REFERENCES

1) Aichner, T., & Shaltoni, A. M. (2022). Digital marketing in the real estate sector: An overview of recent developments. *Journal of Property Research*, 39(4), 307–326. https://doi.org/10.1080/09599916.2021.1969974
2) Baum, A. (2020). PropTech 3.0: The future of real estate. *Journal of Property Investment & Finance*, 38(4), 367–378. https://doi.org/10.1108/JPIF-12-2019-0163
3) Bhandari, A., & Bansal, A. (2021). Artificial intelligence in marketing: Systematic review and implications. *Journal of Business Research*, 124, 330–348. https://doi.org/10.1016/j.jbusres.2020.11.053
4) Bishop, C. M. (2023). *Pattern recognition and machine learning* (2nd ed.). Springer. https://doi.org/10.1007/978-3-031-36407-3
5) Cai, L., Wang, X., & Han, Y. (2022). Multimodal deep learning for behavioral prediction: A comprehensive review. *IEEE Transactions on Neural Networks and Learning Systems*, 33(10), 5280–5298. https://doi.org/10.1109/TNNLS.2021.3129878

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

6) Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *International Conference on Machine Learning (ICML)*. https://doi.org/10.48550/arXiv.2002.05709
7) Chong, A. Y. L., Li, B., Ngai, E., Ch'ng, E., & Lee, P. S. (2017). Predicting online product sales with deep learning. *Decision Support Systems*, 110, 1–10. https://doi.org/10.1016/j.dss.2017.03.009
8) Doshi-Velez, F., & Kim, B. (2017). Towards interpretable machine learning. *Proceedings of the ICML Workshop on Human Interpretability*. https://doi.org/10.48550/arXiv.1702.08608
9) Grosz, B. J., & Stone, P. (2018). AI and the future of work. *Communications of the ACM*, 61(12), 38–42. https://doi.org/10.1145/3208157
10) Higgins, E. T. (2021). Cognitive biases in decision-making: Understanding systematic error. *Annual Review of Psychology*, 72, 59–85. https://doi.org/10.1146/annurev-psych-010419-051053
11) Hoque, R., & Malik, M. (2023). Deep learning applications in real estate analytics: A systematic review. *Expert Systems with Applications*, 213, 118986. https://doi.org/10.1016/j.eswa.2022.118986
12) Jain, V., Kumar, S., & Singhal, A. (2021). Multimodal fusion techniques for machine learning: A survey. *Information Fusion*, 76, 36–57. https://doi.org/10.1016/j.inffus.2021.05.003
13) Kahneman, D., Sibony, O., & Sunstein, C. (2021). *Noise: A flaw in human judgment*. Little, Brown Spark. https://doi.org/10.1037/h0094126
14) Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. (2022). Algorithmic decision-making and human judgment. *Nature Human Behaviour*, 6, 112–120. https://doi.org/10.1038/s41562-021-01285-1
15) Li, Y., Zhang, X., & Liu, J. (2020). Deep multimodal representation learning for real-world applications. *IEEE Transactions on Multimedia*, 22(6), 1423–1434. https://doi.org/10.1109/TMM.2019.2940907
16) Luo, X., Tong, S., Lin, Z., & Zhang, C. (2021). AI in marketing: Review and future research directions. *Journal of the Academy of Marketing Science*, 49, 697–718. https://doi.org/10.1007/s11747-021-00773-7
17) O'Connor, T., & Vakili, V. (2021). Predictive analytics in real estate: The influence of big data and AI. *Journal of Property Research*, 38(3), 220–238. https://doi.org/10.1080/09599916.2020.1845364
18) Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual–language representations (CLIP). *Proceedings of the International Conference on Machine Learning.* https://doi.org/10.48550/arXiv.2103.00020
19) Shankar, V., Kleijnen, M., Ramanathan, S., Rizley, R., Holland, S., & Morrissey, S. (2021). How artificial intelligence is reshaping marketing. *Journal of Interactive Marketing*, 55, 1–16. https://doi.org/10.1016/j.intmar.2021.04.002
20) Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124