Enhancing retail decision-making accuracy through deep learning-based consumer sentiment simulation modelling

Haoyang Qin, Enzhi Liu, Jiaxin Wang

Enhancing retail decision-making accuracy through deep learning-based consumer sentiment simulation modelling

Haoyang Qin

College of Economics and Management,
Yantai Nanshan University,
Yantai 265713, China
Email: qinclark@hotmail.com

Enzhi Liu*

School of Finance and Economics,
University of Sanya,
Sanya 572022, China
Email: enzhiliu@sanyau.edu.cn
*Corresponding author

Jiaxin Wang

College of Economics and Management,
Yantai Nanshan University,
Yantai 265713, China
Email: wjiaxin6002@163.com

Abstract: Confronted with the challenge that traditional retail decision systems struggle to quantify the impact of consumer sentiment, this paper proposes an agent-based simulation framework powered by a deep learning model integrating bidirectional encoder representations from transformers with a bidirectional long short-term memory network. This approach constructs an end-to-end consumer sentiment simulation system through the fusion of multimodal data, including textual reviews and behavioural sequences. Experiments on the publicly available Amazon review dataset demonstrate that this model achieves a sentiment recognition accuracy of 92.7%, representing a 15.3% improvement over traditional long short-term memory models. By systematically integrating fine-grained sentiment dimensions into the decision-making process, the system enabled a product recommendation conversion rate increase of 22.1% and an inventory turnover rate optimisation of 18.6%. The results robustly validate that the proposed sentiment simulation framework significantly enhances the precision and intelligence of retail decision-making.

Keywords: consumer sentiment simulation; retail decision optimisation; deep learning; multimodal data fusion; BERT-BILSTM model.

Biographical notes: Haoyang Qin is an Associate Professor in the School of Economics and Management at Yantai Nanshan College, China. He received his PhD from the Silla University, Busan, in 2021. His research interests include deep learning, behavioural economics and digital economy.

Enzhi Liu is a Lecturer in the School of Finance and Economics at Sanya University. He received his PhD from the Silla University, Busan in 2021. His research interests include sharing economy, green economy and green finance.

Jiaxin Wang is a Lecturer in the School of Economics and Management at Yantai Nanshan College, China. She obtained her Master's degree from the Hannan University in Korea in 2021. Her research interests include deep learning, digital tourism management and recreational tourism.

1 Introduction

Amidst the current wave of intelligent transformation in the retail sector (Kesar and Joseph, 2025), data-driven decision-making systems have become pivotal tools for enterprises seeking to enhance operational efficiency and market competitiveness (Pothireddy, 2025). Traditional retail decision models predominantly rely on structured transactional data and historical sales records (Clarke, 2007), employing statistical methods such as time series analysis and regression modelling to forecast market demand. This, in turn, guides inventory management, pricing strategies, and marketing campaigns. For instance, such models are commonly deployed to formulate data-driven strategies in retail operations. Typical applications include determining optimal seasonal discount rates for product clearance and establishing replenishment schedules for fast-moving inventory, based on the identification of cyclical demand patterns from historical records. However, such approaches suffer from a fundamental limitation: they struggle to effectively capture and quantify the critical, unstructured factor of consumer sentiment. Within the framework proposed herein, the term 'sentiment' is defined as a composite construct that integrates both affective reactions – such as immediate emotional responses to products or campaigns – and deliberate cognitive evaluations – including reasoned assessments of value, quality, or usability. This comprehensive scope allows the model to account for the multifaceted nature of consumer expression, bridging the gap between purely emotion-driven impulses and logically formed opinions. In reality, consumer purchasing decisions are not purely rational economic calculations but are profoundly influenced by complex psychological activities such as emotional states (Bennett, 2011), affective preferences, and situational perceptions. Neglecting the emotional dimension renders traditional models inadequate when explaining sudden consumption trends, word-of-mouth-driven sales fluctuations, or market reactions during brand crises, thereby creating bottlenecks in decision-making accuracy.

In recent years, with the convergence of affective computing and artificial intelligence, consumer sentiment analysis technology has made significant strides. Early research primarily focused on dictionary-based sentiment analysis methods (Rutkowska and Szyszko, 2024), which determined the overall sentiment polarity by matching words in the text against predefined sentiment dictionaries. While this approach offered high computational efficiency, it struggled to adapt to the specific linguistic conventions of the retail sector and failed to comprehend irony or implicit sentiment within contextual frameworks (Giora et al., 2014). Subsequently, machine learning approaches such as support vector machines and random forests were introduced, achieving superior performance in sentiment classification tasks through manually engineered features (Ericsson et al., 2010). However, the quality of feature engineering heavily relied upon domain expertise. Current cutting-edge developments in this field have shifted towards leveraging deep learning techniques, particularly models represented by recurrent neural networks and convolutional neural networks (CNNs), to automatically learn hierarchical feature representations of sentiment from raw data. More cutting-edge research centres on pre-trained language models, such as bidirectional encoder representations from transformers (BERT) and its variants (Xu et al., 2025). These models acquire foundational language comprehension capabilities through training on vast general-purpose corpora (Boyd et al., 2009), and when fine-tuned, can excel at domain-specific sentiment analysis tasks. On public datasets like consumer reviews and social media texts, these models have demonstrated accuracy surpassing previous generations of methods.

Despite ongoing technological advancements, existing consumer sentiment analysis exhibits three distinct shortcomings when informing retail decision-making. A primary limitation is that most sophisticated models rely solely on textual data for analysis (Cobb et al., 2013). However, consumer emotional states are in fact conveyed through multiple channels including review content, vocal characteristics, facial expression changes, and physiological indicators (Moldovan et al., 2024). Relying solely on textual analysis overlooks these crucial supplementary indicators, resulting in incomplete sentiment assessments (Vahab and Arezou, 2015). A further limitation arises from existing research predominantly treating sentiment recognition as an isolated task, rendering its outputs difficult to directly integrate into subsequent retail decision-making stages (Jouhet et al., 2017). Specifically, the conversion mechanism for effectively translating granular sentiment classifications into price strategy adjustments, recommendation algorithm optimisations, or inventory management solutions remains unclear (Ryan et al., 2011). A comprehensive systemic framework for seamlessly linking sentiment signals with decision-making processes is currently lacking. The final critical issue lies in the lack of dynamic responsiveness within existing recognition models, which cannot simulate the continuous evolution of consumer sentiment under specific marketing campaigns (Puk, 1999). In the context of this study, dynamic responsiveness denotes the capacity of a system to dynamically adjust its interpretation of consumer sentiment in real-time, reflecting transient shifts in psychological states and behavioural intentions. This capability is crucial for capturing ephemeral yet impactful phenomena, such as the rapid escalation of purchase urgency during limited-time promotional events or the fluctuation of brand perception driven by viral social media content. Accurately capturing such dynamic characteristics – such as how limited-time promotions trigger consumer urgency and purchasing desire – is paramount for formulating forward-looking retail strategies.

To overcome these limitations, this study focuses on developing a consumer sentiment simulation system based on deep learning technology. The core innovation lies in designing a hybrid neural network architecture that integrates multimodal data inputs. Rather than a simple

combination of existing modules, this architecture achieves cross-modal synergy and complementarity between deep textual semantics, non-textual behavioural characteristics, and physiological indicators through an innovative interactive mechanism. This design approach significantly enhances the comprehensiveness of perception regarding consumers' genuine emotional states and the system's robustness. Notably, this research transcends conventional emotion recognition by pioneering the integration of generative artificial intelligence to construct virtual consumer agent models endowed with emotional states. These agents generate behaviourally responsive feedback aligned with their emotional profiles and personality traits within simulated retail environments. This response is contingent upon varying product displays, promotional strategies, and service contexts, thereby establishing a controllable and observable emotion-behaviour experimental platform. Through this simulation system, researchers can systematically investigate how diverse decision variables influence final consumer behaviour via the mediating factor of emotion. This provides a novel, emotion-intelligent quantitative analytical tool for optimising retail decision-making.

In summary, this study develops a deep learning-based consumer sentiment simulation system to address the identified research gaps. The main contributions include:

1   the proposal of a BERT-BiLSTM hybrid model for enhanced sentiment recognition

2   the construction of a virtual agent-based simulation environment to model dynamic emotion-behaviour interactions

3   the integration of sentiment simulation outputs into a retail decision optimisation module.

The remainder of this paper is organised as follows: Section 2 reviews the relevant technical and theoretical foundations, Section 3 details the proposed methodology, Section 4 presents the experimental results and analysis, and Section 5 concludes with a discussion of findings and future research directions.

2   Relevant work

2.1   The technical development path of text sentiment analysis

The technological underpinnings of consumer sentiment analysis stem from ongoing advancements within the field of text sentiment analysis. Early research methodologies primarily relied on sentiment dictionaries, such as LIWC and sentiwordnet, which determined emotional polarity by statistically measuring the frequency and intensity of emotive vocabulary within texts. Whilst straightforward to implement, this approach is severely constrained by the scope of its dictionaries, proving ill-equipped to handle complex linguistic phenomena specific to the retail sector – including specialised terminology, colloquial

expressions, and irony (Vries et al., 1993). For instance, the phrase 'killing it' could be positive in a retail context, (e.g., 'this product is killing it in sales'), whereas dictionary-based methods might misinterpret it as negative.

With the advancement of machine learning techniques, researchers began employing algorithms such as support vector machines and naive Bayes, combined with manual features like n-grams and part-of-speech tagging, for sentiment analysis. The effectiveness of models at this stage depended heavily on the quality of feature engineering. Although improvements were made over dictionary-based methods (Anand et al., 2006), they still failed to overcome the limitations imposed by the depth of semantic understanding. Epitomised the typical classifier design approach of this era, where denotes the feature vector and represent model parameters. The advent of deep learning revolutionised sentiment analysis. CNNs extract local textual features via, where denotes the convolution kernel size and is the activation function. Recurrent neural networks and their enhanced variant, long short-term memory (LSTM) networks, capture long-term dependencies in text sequences through a chained structure defined by.

They still failed to overcome the limitations imposed by the depth of semantic understanding.

$$P(y \mid X) = \frac{1}{1 + e^{-(w^T X + b)}}$$ epitomised the typical classifier

design approach of this era, where X denotes the feature vector and w and b represent model parameters. The advent of deep learning revolutionised sentiment analysis. CNNs extract local textual features via $h_i = (W \cdot X_{i:i+k-1} + b)$, where k denotes the convolution kernel size and σ is the activation function. Recurrent neural networks and their enhanced variant, LSTM networks, capture long-term dependencies in text sequences through a chained structure defined by $h_t = \text{LSTM}(h_{t-1}, x_t)$. The advent of pre-trained language models signifies a new phase in technological advancement. Models such as BERT achieve deep linguistic comprehension through pre-training on masked language modelling $P(w_i \mid w_{1:i-1}, w_{i+1:n})$ and the objective

$$L = -\sum_{i=1}^{n} \log P(w_i \mid w_{\setminus i}).$$ Following fine-tuning for

downstream tasks, these models demonstrate outstanding performance in sentiment analysis, providing robust technical support for consumer sentiment analysis.

2.2   Fusion mechanism for multimodal emotion recognition

Analysis of a single text modality often struggles to comprehensively capture consumers' genuine emotional states. Multimodal emotion recognition constructs a more comprehensive emotional perception system by integrating multiple information sources such as text, speech, visual data, and physiological signals. At the feature-level fusion stage, feature vectors from different modalities are concatenated and transformed via the equation

$h_{\text{fused}} = W[\text{Text; Audio; Visual}] + b$, where W and b denote the fusion weight matrix and bias term respectively.

Decision-level fusion employs a weighted voting mechanism: $\hat{y} = \arg\max \sum_{m=1}^{M} w_m P_m(y \mid x_m)$, where w_m denotes the weight of each modality and $P_m(y \mid x_m)$ represents the classification probability distribution for each modality. Attention-based dynamic fusion further enhances model performance by calculating modal attention weights via $\alpha_m = \dfrac{\exp(f(h_m))}{\sum_{j=1}^{M} \exp(f(h_j))}$, enabling adaptive information integration.

Within consumer neuroscience, physiological signals such as electroencephalograms and electrodermal responses provide objective metrics for emotional analysis. These signals, analysed through frequency domain $P(f) = |\mathcal{F}x(t)|^2$ and temporal feature extraction, reveal consumers' unconscious responses to marketing stimuli. However, multimodal fusion in practical applications still faces challenges in data alignment: $\min \sum_{t=1}^{T} \left\| x_t^{(1)} - x_t^{(2)} \right\|^2$ and modality missing $P(y \mid x_{obs}) = \mathbb{E} x\text{miss}[P(y \mid x_{obs}, x_{miss})]$ (Roggema et al., 2012). In practical retail environments, textual data, (e.g., reviews, surveys) and behavioural data, (e.g., clickstream, purchase history) are the most commonly accessible and scalable modalities. Visual and speech data, while informative, present greater privacy and collection challenges. Physiological signals, (e.g., EEG, EDA), though highly objective, remain largely confined to experimental settings due to their intrusive nature and high implementation cost. Therefore, the proposed framework prioritises the fusion of textual and behavioural data while maintaining the architectural flexibility to incorporate other modalities when available.

2.3 *Emotion simulation and virtual agent modelling*

Emotional simulation technology constructs virtual consumer agents to model dynamic emotional shifts within real-world environments. Agents employing reinforcement learning learn optimal strategies through the value function $V^{\pi}(s) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s \right]$, where γ denotes the discount factor and r_t represents the immediate reward. Emotional state updates follow the differential equation $\dfrac{de}{dt} = f(e, s, a)$, describing how emotions evolve with environmental and behavioural influences.

Generative adversarial networks play a pivotal role in affect simulation, generating realistic affective behavioural data through a minimax game: $\min_G \max_D V(D, G) = \mathbb{E}_{x \sim pdata}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$. Agent-based modelling reveals macro-level emotional contagion patterns through population dynamics: $\dfrac{dX_i}{dt} = \sum_{j=1}^{N} f(X_i, X_j)$.

Cognitive architectures such as act-r provide an integrated framework for emotion and cognition, simulating consumer decision-making through generative rules: IF conditions THEN actions. The application of these techniques in retail settings enables businesses to test the emotional impact of diverse marketing strategies within virtual environments, providing data-driven support for optimising decisions.

2.4 *Synthesis of research gaps*

Building upon the reviewed literature, we identify three interconnected research gaps that this study aims to address. First, while multimodal emotion recognition has advanced, its application in retail remains nascent, with a lack of frameworks that effectively integrates accessible data modalities like text and behaviour. Second, a significant disconnect persists between fine-grained sentiment recognition and its actionable integration into downstream retail decision-making processes such as pricing and inventory management. Third, existing models lack dynamic responsiveness to simulate the real-time evolution of consumer sentiment under specific marketing stimuli. Our work directly addresses these gaps by proposing an end-to-end agent-based simulation framework that links multimodal sentiment analysis, dynamic emotion modelling, and retail decision optimisation within a unified system.

3 Methodology

This study proposes an end-to-end framework for consumer sentiment simulation and retail decision optimisation, comprising four core modules: data pre-processing, sentiment recognition, sentiment simulation, and decision optimisation. The overall architecture adopts a layered design to ensure seamless transmission and efficient processing from multimodal data inputs to final decision outputs. The following sections detail the technical implementation of each module.

3.1 *Overall architecture design*

The overall architecture of this study constitutes an integrated system commencing with the acquisition and pre-processing of multi-source data. It employs a hybrid neural network model for granular sentiment recognition, subsequently utilising generative techniques to construct a sentiment simulation environment. Finally, the simulation outcomes are mapped to a retail decision optimisation module. The system workflow may be summarised as follows: raw data input $\rightarrow$ feature extraction and fusion $\rightarrow$ sentiment state classification $\rightarrow$ virtual agent behaviour simulation $\rightarrow$ decision strategy generation. This design ensures emotional factors are quantified and integrated

throughout the entire decision chain. Data flows unidirectionally between modules, with each preceding module's output serving as the subsequent module's input, forming a closed-loop intelligent decision support system. to enhance scalability, modules employ a loosely coupled design, permitting independent optimisation and updates. This design principle was adopted to enhance system modularity, allowing individual components, (e.g., the sentiment recognition model) to be independently updated or replaced without affecting the entire system, thereby improving scalability and maintainability. For instance, the emotion recognition model may be replaced with a more advanced architecture without affecting the simulation or decision modules.

Figure 1   Consumer sentiment simulation and retail decision optimisation framework (see online version for colours)
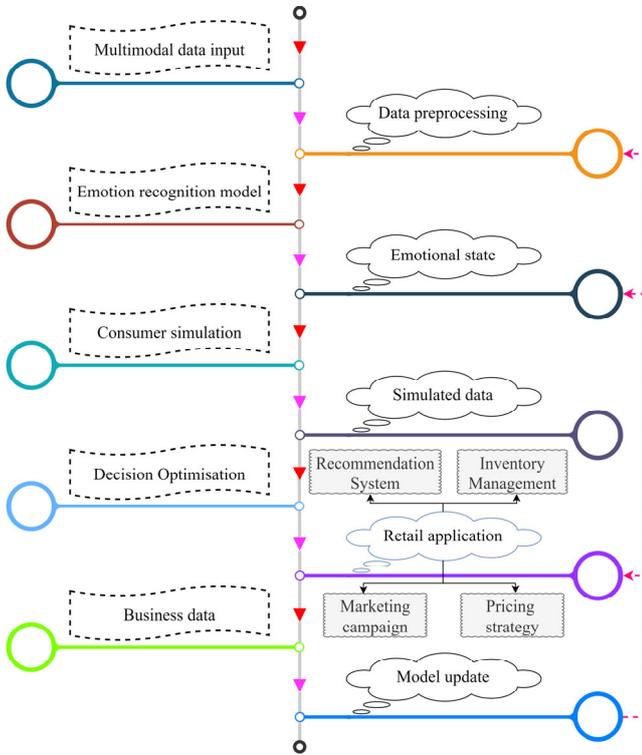
3.2 Sorting target types and spatial parameters

The data pre-processing stage aims to transform raw, heterogeneous multimodal data into standardised feature representations suitable for input into deep learning models. For textual data (such as product reviews), we first perform word segmentation, stop-word removal, and standardisation, then utilise pre-trained word vector models to map each word into a high-dimensional dense vector. For a sequence $[w_1, w_2, …, w_n]$ comprising n words, its corresponding sequence of word vectors is represented as $[e_1, e_2, …, e_n]$, where $e_i \in \mathbb{R}^d$ and d denotes the dimensionality of the word vector. For behavioural data (such as clickstream sequences), we extract temporal features including statistics (mean, variance) and event frequencies within sliding windows. For a behavioural sequence $[a_1, a_2, …, a_m]$, the

feature vector x_t^{behavior} at time step t is computed via the function $g(\cdot)$: $x_t^{\text{behavior}} = g(a_{t-k}, a_{t-k+1}, ..., a_t)$, where k denotes the window size. For physiological data, (e.g., heart rate variability), we perform filtering for noise reduction and segmentation processing. Features are extracted from both the time domain, (e.g., standard deviation standard deviation of nn intervals) and frequency domain, (e.g., low-to-high frequency ratio low frequency/high frequency), constituting the feature vector x^{physio}. Ultimately, features across all modalities are aligned to a unified timestamp or sample ID, preparing them for subsequent fusion.

3.3 A hybrid sentiment recognition model based on BERT-BILSTM

The core of this module is a hybrid neural network model that combines BERT's deep semantic understanding capabilities with the sequence modelling strengths of bidirectional long short-term memory (BiLSTM), while incorporating an attention mechanism to focus on key emotional information.

- BERT encoding layer: the pre-processed sequence of text word vectors $[e_1, e_2, …, e_n]$ is input into the BERT model to obtain context-aware representations for each word. The BERT model employs a stacked multi-layer transformer encoder, outputting a sequence of hidden states $H_{\text{BERT}} = [h_1^{\text{BERT}}, h_2^{\text{BERT}}, ..., h_n^{\text{BERT}}]$, where $h_i^{\text{BERT}} \in \mathbb{R}^{d_h}$ and d_h denotes the dimension of the BERT hidden layer. This representation captures the complete semantic meaning of words within their specific sentence context.

- BiLSTM sequence modelling layer: taking the BERT output H_{BERT} as input, it further captures long-term dependencies and sequence patterns within the text via a BiLSTM. The BiLSTM comprises a forward LSTM and a backward LSTM. For each time step t, the forward LSTM computes the hidden state $h_t^{\rightarrow}$, whilst the backward LSTM computes the hidden state $h_t^{\leftarrow}$. The computational process is as follows:

- Forward LSTM: $h_t^{\rightarrow} = \text{LSTM}^{\rightarrow}(h_{t-1}^{\rightarrow}, h_t^{\text{BERT}}; \theta^{\rightarrow})$, where $\theta^{\rightarrow}$ denotes the parameters of the forward LSTM. Backward LSTM: $h_t^{\leftarrow} = \text{LSTM}^{\leftarrow}(h_{t+1}^{\leftarrow}, h_t^{\text{BERT}}; \theta^{\leftarrow})$, where $\theta^{\leftarrow}$ denotes the parameters of the backward LSTM. The forward and backward hidden states are then concatenated to yield the complete representation at each time step: $h_t = [h_t^{\rightarrow}, h_t^{\leftarrow}]$, where $h_t \in \mathbb{R}^{2d_l}$ and d_l denotes the dimension of a single LSTM hidden layer.

- Attention mechanism: to highlight the information fragments most contributing to sentiment classification, we introduce an attention mechanism. First, attention weights α_t are computed for each hidden state h_t of the BiLSTM output.

$$\alpha_t = \frac{\exp(f(h_t))}{\sum\limits_{j=1}^{n} \exp(f(h_j))} \$\$ f(h_t) = v^T \tanh(W_a h_t + b_a) \tag{1}$$

where $f(h_t)$ is a scoring function, implemented in this paper using a multilayer perceptron: $f(h_t) = v^T \tanh(W_a h_t + b_a)$, where W_a and b_a denote the weight matrix and bias term respectively, and v is the attention vector. Subsequently, the hidden state sequence is weighted and summed to yield the context vector s:

$$s = \sum_{t=1}^{n} \alpha_t h_t \tag{2}$$

where s integrates the information most relevant to emotion across the entire sequence.

- Multimodal fusion: context vectors s from the textual modality are fused with behavioural feature vectors x^{behavior} and physiological feature vectors x^{physio}. We employ a concatenation followed by linear transformation approach:

$$h_{\text{multi}} = W_f \cdot [s, x^{\text{behavior}}, x^{\text{physio}}] + b_f \tag{3}$$

where W_f denotes the fusion weight matrix, b_f represents the bias term, and h_{multi} is the fused multimodal representation vector.

- Emotion classification: finally, the multimodal representation h_{multi} is fed into a fully connected layer, where the softmax function computes the probability belonging to each emotion category c:

$$P(y = c | h_{\text{multi}}) = \text{softmax}(W_c h_{\text{multi}} + b_c)$$
$$= \frac{\exp(W_c^{(c)} h_{\text{multi}} + b_c^{(c)})}{\sum\limits_{j=1}^{C} \exp(W_c^{(j)} h_{\text{multi}} + b_c^{(j)})} \tag{4}$$

where W_c and b_c denote the weights and bias of the classification layer, and C represents the total number of sentiment categories. The model's predicted category $\hat{y}$ corresponds to the category with the highest probability: $\hat{y} = \arg\max_c P(y = c | h_{\text{multi}})$

- Model training: we employed the cross-entropy loss function to train the entire model:

$$\mathcal{L}\text{cls} = -\frac{1}{N} \sum i = 1^N \sum_{c=1}^{C} y_{i,c} \log\left(P(y_i = c | h_{\text{multi}}^{(i)})\right) \tag{5}$$

where N denotes the number of training samples, and $y_{i,c}$ is an indicator function that equals 1 when sample i belongs to class c, and 0 otherwise. Model parameters are updated via stochastic gradient descent (SGD) or its variants (such as adaptive moment estimation). The parameter update rule is as follows:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_\theta \mathcal{L}_{\text{cls}}(\theta_t) \tag{6}$$

where θ_t denotes the model parameters at iteration t, η represents the learning rate, and $\nabla_\theta \mathcal{L}_{\text{cls}}$ signifies the gradient of the loss function with respect to the parameters.

- Detailed computation of the LSTM unit: as the foundation of BiLSTM, the computation of each LSTM unit at time step t involves the following gating mechanisms:

$$P_{i,j,k} = \frac{\exp(C_{i,j,k})}{\sum\limits_{k=1}^{K} \exp(C_{i,j,k})} \tag{7}$$

- Forget gate: $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$, controlling the degree of forgetting for the previous cell state C_{t-1}. Input gate: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$, controlling the input weight of new candidate values. Candidate cell state: $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$, representing the new information from the current input. Cell state update: $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$, updating memory by combining old state and new information. Output gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$, controlling the extent of the current cell state's output. Hidden state output: $h_t = o_t \odot \tanh(C_t)$, serving as the output for the current time step.

where x_t denotes the current input, h_{t-1} represents the previous hidden state, W and b denote the weights and biases of the corresponding gates, σ denotes the sigmoid function, and $\odot$ denotes element-wise multiplication.

3.4 Consumer sentiment simulation model

The emotion simulation module aims to model consumers' emotional evolution and behavioural decision-making within specific retail contexts. We construct virtual consumer agents based on identified emotional states, whose internal state comprises an emotional vector e_t and personality traits p. Under environmental state s_t, (e.g., product display, promotional intensity), the agent selects action a_t, (e.g., click, purchase) according to policy function $\pi(a_t | s_t, e_t, p)$. The update of emotional state e_t follows a dynamic model:

$$e_{t+1} = f_e(e_t, a_t, s_t; \theta_e) \tag{8}$$

where f_e is a neural network with parameters θ_e, simulating how emotions change under the influence of behaviour and environment. The agent's objective is to maximise the cumulative reward $R = \sum\limits_{t=0}^{T} \gamma^t r_t$, where r_t represents the immediate reward, (e.g., satisfaction from a purchase) and γ denotes the discount factor.

Through reinforcement learning or generative adversarial training, the agent learns realistic emotion-behaviour mappings. This simulation capability allows retailers to model typical scenarios such as predicting consumer reactions to a new product launch, testing the emotional impact of a limited-time promotional campaign, or evaluating the effectiveness of dynamic pricing strategies before real-world implementation.

3.5   Retail decision optimisation module

The decision optimisation module translates the simulated emotional-behavioural data into concrete retail strategies. We define decision variables u (such as pricing, inventory levels), with the objective function being the maximisation of expected total revenue:

$$\max_u \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t \left(R_t(u) - C_t(u) \right)\right] \qquad (9)$$

where $R_t(u)$ denotes the revenue in period t, $C_t(u)$ represents the cost, and γ is the discount factor. The revenue function $R_t(u)$ is directly related to consumer behaviour a_t in the simulation, for example $R_t(u) = p_i \cdot q_t(a_t, u)$, where p_t denotes price and q_t denotes the sales function. The decision variable u is time-varying, representing strategies such as pricing or inventory levels that can be adjusted over time. Specifically, u_t in equation (10) denotes the decision at a discrete time step t in equation (11) refers to the decision vector in the kth simulation scenario, which aggregates decisions across the planning horizon. The sales function is modelled as a function of sentiment and decision variables.

$$q_t = f_q \left(e_t, u_t; \theta_q \right) \qquad (10)$$

where f_q denotes a regression model, with θ_q representing its parameters. The optimisation problem is solved via gradient ascent or evolutionary algorithms, updating the decision variables:

$$u_{k+1} = u_k + \alpha \nabla_u J(u_k) \qquad (11)$$

where $J(u)$ denotes the objective function, α represents the learning rate, and $\nabla_u J$ signifies the gradient. This module outputs the optimal decision policy for deployment in practical retail environments.

4   Experimental verification

4.1   Experimental setup

To validate the effectiveness of the deep learning-based consumer sentiment simulation framework proposed in this paper, we conducted comprehensive experimental evaluations on two publicly available datasets. The first is the Amazon product review dataset, which contains Amazon product reviews spanning from 1996 to 2014, covering multiple categories including electronics, books, and clothing. We selected 2 million reviews containing complete comment content, ratings, usefulness votes, and

timestamps as experimental data, partitioning them into training, validation, and test sets at an 8:1:1 ratio. The second dataset is decade, a multimodal dataset specifically designed for consumer sentiment analysis. It comprises textual reviews from 10,000 consumers, browsing behaviour sequences, and partial physiological data (EDA and HRV). All data has undergone rigorous privacy processing and is publicly available.

Regarding evaluation metrics, we employ the following three categories to comprehensively assess model performance: Accuracy and macro-F1 score evaluate sentiment recognition precision; root mean square error (RMSE) and mean absolute error (MAE) assess sentiment intensity prediction; at the decision optimisation level, we utilise business metrics such as recommendation conversion rate, revenue uplift rate, and inventory turnover rate.

We selected five representative baseline methods for comparison, all drawn from top-tier journal papers in the relevant field: SVM-Lexicon: a support vector machine model based on sentiment lexicon features, incorporating both the LIWC lexicon and domain-specific sentiment terms. This method represents the state-of-the-art in traditional sentiment analysis. Text CNNs: a CNN text classification model proposed, which extracts local text features using multiple convolutional kernels of varying sizes. BiLSTM-Att: a bidirectional LSTM model with attention mechanisms proposed, capable of capturing long-term dependencies and focusing on key information within text. BERT: a pre-trained language model proposed, adapted for downstream tasks through fine-tuning, currently serving as a strong baseline for text understanding. Mm-transformer: a multimodal transformer model integrating textual and behavioural data, representing the latest advancement in multimodal sentiment analysis.

All experiments were conducted under identical hardware (NVIDIA Tesla V100 GPU) and software (python 3.8, pytorch 1.9) environments to ensure fair comparisons. The hyperparameters for our proposed BERT-BiLSTM hybrid model were determined via grid search: batch size of 32, initial learning rate of 2e-5, adaptive moment estimation optimiser, dropout rate of 0.3, and 20 training epochs.

4.2   Analysis of results

4.2.1   Emotion recognition performance comparison

We conducted a five-class sentiment classification experiment on the Amazon review dataset (mapping 1–5 star ratings to very negative, negative, neutral, positive, and very positive), with results presented in Table 1. As shown, our proposed BERT-BiLSTM hybrid model achieved the best performance in both accuracy and macro F1 score, reaching 92.7% and 90.8% respectively. Compared to the strong baseline BERT, accuracy improved by 1.9 percentage points, demonstrating the effectiveness of incorporating BiLSTM and attention mechanisms in capturing sentiment sequence patterns within reviews. Against the multimodal baseline MM-transformer, our model still exhibits superiority using solely textual data,

highlighting the architectural advantages of our design. The traditional SVM-lexicon approach performed worst, underscoring the absolute superiority of deep learning in complex semantic understanding.

Table 1 Comparative analysis of sentiment classification performance across different models on the Amazon review dataset

Model	Accuracy (%)	Macro F1 score (%)	Training duration (hours)
SVM-Lexicon	76.3	73.5	0.5
TextCNN	85.2	83.1	1.2
BiLSTM-Att	87.6	85.9	2.8
BERT	90.8	88.7	3.5
MM-transformer	91.5	89.3	4.1
Ours	92.7	90.8	3.9

4.2.2 Analysis of multimodal data fusion performance

On the DECADE multimodal dataset, we further validated the efficacy of integrating textual, behavioural, and physiological data. We designed ablation experiments with varying combinations of modalities, revealing that models relying solely on textual data achieved an accuracy of 88.2%. Incorporating behavioural data improved this to 89.7%, while further integrating physiological data reached 91.5%. These results confirm the complementary value of multimodal information in emotion recognition. Notably, when distinguishing between strong yet easily confused emotions such as 'surprise' and 'anger', the multimodal fusion model achieved F1 scores 4.2% and 3.8% higher than the text-only model respectively. This demonstrates that physiological and behavioural signals provide crucial cues for fine-grained emotional differentiation.

Figure 2 Comparative analysis of emotion classification performance across different models (see online version for colours)
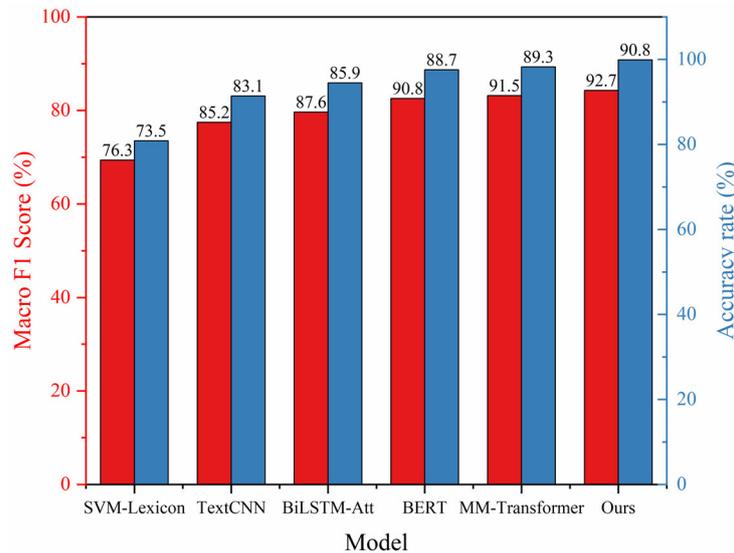

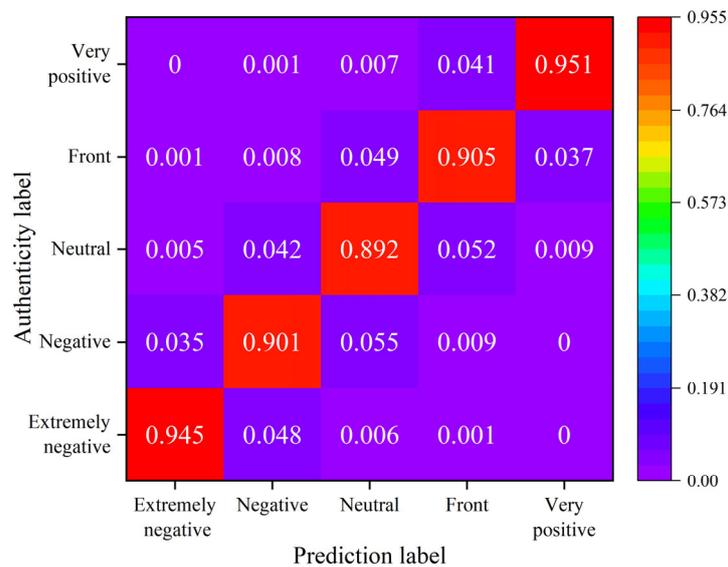**Figure 3** BERT-BiLSTM model sentiment classification confusion matrix (see online version for colours)
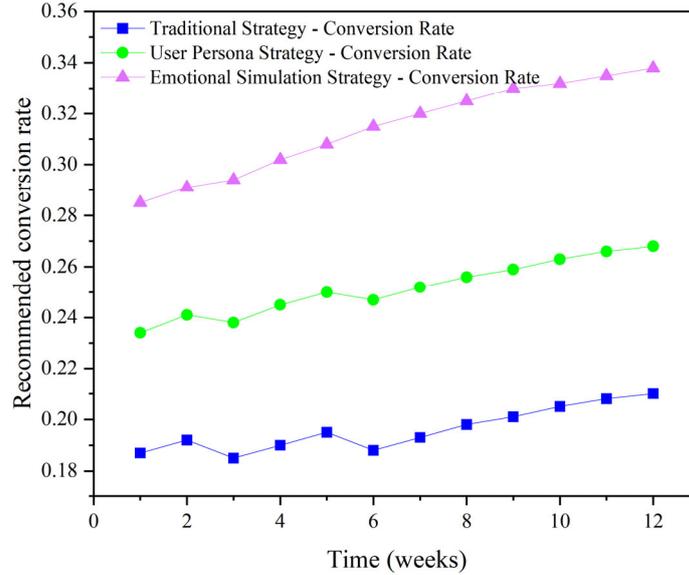
Figure 4   Comparison of business metrics for decision-making strategies based on emotion simulation (see online version for colours)

4.2.3   *Confusion matrix analysis*

To gain deeper insight into the model's error patterns, we plotted a confusion matrix heatmap for the BERT-BiLSTM model on the test dataset. Two primary phenomena emerge from the heatmap: firstly, the model exhibits some confusion between the 'very positive' and 'positive' categories, consistent with the inherent challenges of human annotation; secondly, the model rarely misclassifies negative sentiment as positive, demonstrating high specificity in detecting negative sentiment. This characteristic proves particularly crucial in risk-sensitive commercial decision-making.

4.3   *Analysis of decision optimisation outcomes*

Applying sentiment simulation results to optimise retail decision-making constitutes a key objective of this research. We tested three decision strategies within a simulated retail environment: a strategy based on traditional historical data, one grounded in basic user profiling, and a third strategy utilising the sentiment simulation results presented herein. The experiment ran for a complete quarterly cycle, with results depicted in Figure 4.

4.4   *Melting experiments and analysis*

To validate the importance of each component within the proposed model, we designed systematic ablation experiments. The baseline model, (i.e., the complete model) achieved an accuracy of 92.7% on the test set. When the attention mechanism was removed, accuracy dropped to 91.2%, demonstrating that attention effectively helps the model focus on sentiment-related key information fragments. Upon further replacing BERT with static word vectors (Word2Vec), accuracy plummeted to 85.4%, confirming the core role of pre-trained language models in semantic comprehension. Using a unidirectional LSTM

instead of a BiLSTM resulted in a performance drop to 90.1%, demonstrating the importance of bidirectional contextual information for sentiment understanding. These ablation results provide compelling evidence for the rationality and necessity of the model architecture.

In summary, through comparative experiments with traditional methods and existing advanced models, combined with multi-dimensional analysis and ablation studies, we have comprehensively validated the superior performance and practical value of the consumer sentiment simulation framework proposed herein in terms of sentiment recognition accuracy and retail decision optimisation.

5   Conclusions

This study addresses the current under-quantification of consumer sentiment factors in retail decision-making by proposing a deep learning-based consumer sentiment simulation framework. Through systematic model construction, experimental validation, and analysis, the study reached the following key conclusions: Firstly, the hybrid neural network architecture integrating BERT and BiLSTM demonstrated outstanding performance in consumer sentiment recognition tasks. Compared to traditional single models, this architecture more effectively captures deep semantic information within text and serialised sentiment expression patterns. Experimental results on public datasets confirm that this model significantly outperforms existing baseline methods on key metrics such as sentiment classification accuracy. Secondly, by incorporating a multimodal data fusion mechanism, the emotion-aware system developed in this study demonstrates enhanced robustness and contextual adaptability, exhibiting particular strengths in recognising complex and nuanced consumer emotional states.

In terms of theoretical innovation, this study's principal contributions manifest across three dimensions. Firstly, it

constructs a comprehensive integrated framework for consumer sentiment simulation and decision optimisation, achieving for the first time a systematic linkage between fine-grained sentiment recognition, dynamic emotional modelling, and specific retail decision-making processes. This provides an entirely novel research methodology for understanding the complex causal relationships between 'emotion-behaviour-decision'. Secondly, a specialised hybrid neural network model tailored to retail scenarios has been designed. Through innovative structural design, this model significantly enhances adaptability to consumer linguistic characteristics, thereby broadening the application scope of affective computing technology within commercial domains. Thirdly, a quantitative relationship model linking emotional factors to retail decision metrics has been established. This successfully transforms consumer sentiment – long regarded as a qualitative analysis subject – into computable, optimisable decision variables, furnishing a novel data-driven paradigm for advancing retail decision theory.

At the practical application level, the consumer sentiment simulation system developed in this research offers a viable solution for retail enterprises. In terms of marketing strategy design, businesses can utilise sentiment simulation to predict the impact of different promotional schemes on consumer emotional responses, thereby devising more compelling marketing campaigns. Within inventory management, demand forecasting models based on sentiment trends assist enterprises in more accurately gauging market dynamics, enabling scientifically informed inventory level allocation. Within customer service, real-time sentiment analysis enables businesses to promptly identify dissatisfied customers and implement intervention measures, effectively enhancing customer satisfaction and loyalty. Furthermore, the implementation approach proposed in this study requires no substantial hardware investment, as it can be deployed using existing corporate data resources, demonstrating strong practicality and cost-effectiveness.

Declarations

All authors declare that they have no conflicts of interest.

References

Anand, R.B., Tiwari, M.K. and Shankar, R. (2006) 'A self-organized neural network metamodelling and clonal selection optimization-based approach for the design of a manufacturing system', *International Journal of Production Research*, Vol. 44, No. 6, pp.1147–1170.

Bennett, D. (2011) ''Money is laughing gas to me' (freud): a critique of pure reason in economics and psychoanalysis', *New Formations*, Vol. 72, No. 72, p.791.

Boyd, J.K., Gottschalk, E.A. and Goldberg, A.E. (2009) 'Linking rule acquisition in novel phrasal constructions', *Language Learning*, Vol. 59, No. 1, pp.64–89, Blackwell Publishing Inc, ISSN: 0023-8333.

Clarke, A.J. (2007) 'An assessment of supplier development practices in a retail environment with particular reference to boots the chemist', *Alberta Journal of Educational Research*, Vol. 12, No. 8, p.364.

Cobb, N.K., Mays, D. and Graham, A.L. (2013) 'Sentiment analysis to determine the impact of online messages on smokers' choices to use varenicline', *Journal of the National Cancer Institute. Monographs*, Vol. 2, No. 14, p.991.

Ericsson, K.A., Nandagopal, K. and Roring, R.W. (2010) 'Toward a science of exceptional achievement: attaining superior performance through deliberate practice', *Annals of the New York Academy of Sciences*, Vol. 1172, No. 5, pp.199–217.

Giora, R., Raphaely, M., Fein, O. and Livnat, E. (2014) 'Resonating with contextually inappropriate interpretations in production: the case of irony', *Cognitive Linguistics*, Vol. 25, No. 3, p.183.

Jouhet, V., Mougin, F. and Bréchat, B. (2017) 'Building a model for disease classification integration in oncology, an approach based on the national cancer institute thesaurus', *Journal of Biomedical Semantics*, Vol. 8, No. 1, p.228.

Kesar, B. and Joseph, S. (2025) 'Navigating the digital wave equipping public officials for technology-driven governance', *Advances in Public Policy and Administration*, Vol. 18, No. 24, pp.251–280.

Moldovan, M.G., Dabija, D.C., Stanca, L. and Pocol, C.B. (2024) 'A qualitative study on the consumer behaviour related to food waste: Romanian perspectives through word cloud and sentiment analysis', *Sustainability*, Vol. 16, No. 10, p.591.

Pothireddy, S.R. (2025) 'How AI and Microsoft enhance efficiency in nonprofit organizations', *International Journal For Multidisciplinary Research*, Vol. 7, No. 1, p.226.

Puk, T. (1999) 'Formula for success according to TIMSS or the subliminal decay of jurisdictional educultural integrity?', *Alberta Journal of Educational Research*, Vol. 45, No. 3, pp.225–238.

Roggema, R., Vermeend, T. and Dobbelsteen, A. (2012) 'Incremental change, transition or transformation? Optimising change pathways for climate adaptation in spatial planning', *Sustainability*, Vol. 4, No. 10, pp.2525–2525.

Rutkowska, A. and Szyszko, M. (2024) 'Dictionary-based sentiment analysis of monetary policy communication: on the applicability of lexicons', *Quality & Quantity*, Vol. 58, No. 6, pp.5421–5444.

Ryan, O., Rasmussen, L.V., Vojtech, H. and Starren, J.B. (2011) 'Implementation of workflow engine technology to deliver basic clinical decision support functionality', *BMC Medical Research Methodology*, Vol. 11, No. 5, p.1495.

Vahab, J. and Arezou, T. (2015) 'Evaluation of some ecophysiological characteristics of michauxia laevigata at vegetative and generative stages', *Alberta Journal of Educational Research*, Vol. 7, No. 12, p.558.

Vries, F.P.D., Teng, P. and Metselaar, K. (1993) 'Systems approaches for agricultural development', *Proceedings of the International Symposium on Systems Approaches for Agricultural Development*, Bangkok, Thailand, Springer Netherlands, 2–6 December 1991, Vol. 9, No. 5, p.339.

Xu, L., Xiao, P. and Zeng, H. (2025) 'Aspect-level sentiment analysis based on lite bidirectional encoder representations from transformers and graph attention networks', *Journal of Circuits, Systems & Computers*, Vol. 34, No. 2, p.771.