scientific reports

Check for updates

OPEN

A multigrained preference analysis method for product iterative design incorporating AI-generated review detection

Zhaojing Su[1,2], Mei Yang[1]✉, Qingbo Zhai[3]✉, Kaiyuan Guo[1], Yuexin Huang[2] & Yangfan Cong[2,4]

Online reviews significantly influence consumer purchasing decisions and serve as a vital reference for product improvement. With the surge of generative artificial intelligence (AI) technologies such as ChatGPT, some merchants might exploit them to fabricate deceptive positive reviews, and competitors may also fabricate negative reviews to influence the opinions of consumers and designers. Attention must be paid to the trustworthiness of online reviews. In addition, the opinions expressed by users are limited, and design details hidden behind reviews also affect the product usage experience. Therefore, on the basis of integrated AI-generated review detection, a multigrained user preference analysis method is proposed in this work. The proposed method utilizes pre-trained language models and designs an authenticity detection model for online reviews. Subsequently, attribute-grained preference analysis is considered a text-filling problem and uses the text-infilling objective for domain-adaptive pretraining, facilitating knowledge transfer. On the basis of the feature selection algorithm, a calculation method for the importance of product design features is proposed by introducing a random idea. The proposed method analyzes user preferences at the granularity of product attributes and design features, enabling targeted cost control and optimization in product development and guiding design decisions. Rigorous comparative and few-shot experiments substantiate the superiority of the proposed method.

Keywords Product iterative design, User preference analysis, AI-generated review detection, User-generated content, Pretrained language model, Text filling

Design decisions, as important links that affect product quality and manufacturing costs, largely determine the development cycle and market competitiveness of a product[1]. With the growth of e-commerce and continuous innovation in information technology, product improvement is no longer solely based on the knowledge and experience accumulated by designers but rather focuses on the use of more reasonable and comprehensive big data analysis results[2]. On the one hand, product designers can grasp the advantages and disadvantages of a product by analyzing praise and complaints from online customer reviews. On the other hand, consumers can obtain product usage experience through reviews[3]. However, the veracity of reviews needs to be taken seriously, because not all reviews are written by genuine purchasers. Generative AI tools can quickly produce a large volume of high-quality review content, making the production of fraudulent content more covert and efficient. This interferes with the product design process and may lead designers astray because of false indicators. Table 1 shows the reviews of camera buyers generated by AI. As shown in Table 1, we encounter three meticulously crafted reviews, each demonstrating an astonishingly lifelike effect.

Both genuine reviews and AI-generated reviews often involve multiple attributes (i.e., aspects) of the product. For example, in Table 1, the reviews discuss image quality, functional range, and physical stability. However, they overlook other design features, such as body material, weight, energy efficiency, and operational convenience. These less obvious design elements can impact product performance and user satisfaction. Traditional review analysis methods often fail to capture these hidden dimensions, leaving designers without the comprehensive insights necessary for targeted product improvements.

[1]Department of Industrial Design, College of Arts, Shandong University of Science and Technology, Tsingtao 266590, China. [2]Key Laboratory of Ministry of Industrial Design and Ergonomics, Northwestern Polytechnical University, Xi'an 710072, China. [3]College of Ocean Science and Engineering, Shandong University of Science and Technology, Tsingtao 266590, China. [4]School of Design, The Hong Kong Polytechnic University, Hung Hom, Hong Kong 999077, SAR, China. ✉email: skdyangmei@163.com; Zhai0526@126.com

Prompt	AI-generated reviews
Response	Review by Sarah: ★★★★★ I recently purchased this camera for my photography hobby, and I couldn't be happier with it! The image quality is outstanding, even in low light conditions. The autofocus is lightning-fast, which makes capturing fast-moving subjects a breeze. The build quality feels solid, and the controls are intuitive. Plus, the battery life is impressive, allowing me to shoot for hours without needing a recharge. Overall, I highly recommend this camera to any enthusiast or professional photographer! Review by David: ★★★☆☆ I bought this camera hoping for exceptional video performance, and while it does deliver decent quality, I found the lack of in-body stabilization disappointing. It's a bit frustrating when shooting handheld footage, especially in dynamic environments. However, the image quality for photos is fantastic, and the range of features is impressive. If you primarily shoot stills, this camera is a solid choice, but if video is your focus, you might want to consider other options. Review by Emily: ★★★★☆ As someone new to photography, I wanted a camera that would help me learn and grow without overwhelming me with complexity. This camera has been perfect for that. The user interface is straightforward, and the beginner-friendly modes have been invaluable as I've been experimenting with different settings. The included kit lens is versatile, and the image quality exceeds my expectations. I'm excited to continue honing my skills with this camera by my side!

Table 1. Fabricated reviews via an intelligent AI chatbox model, ChatGPT3.5.

Driven by these observations, the primary objective of this paper is to provide a comprehensive solution that effectively filters out AI-generated fake reviews and leverages attribute-grained and feature-grained preference analysis to guide product iterative design. The remainder of this paper is organized as follows: Sect. 2 reviews the related literature, highlights research gaps, and delineates the innovative contributions of this study. Section 3 outlines the task description and presents the proposed multigranularity preference analysis method that integrates AI-generated review detection. In Sect. 4, experiments are conducted to evaluate the proposed method and its performance metrics. Section 5 carried out extensive comparative experiments, followed by a discussion of the limitations of the experimental results and future research prospects. Section 6 provides conclusions and recommendations for future work.

Related work

The continuous emergence of a large amount of user-generated content (UGC) provides rich resources for researchers to extract preferences from UGC. Many related works are ongoing to explore and improve existing methods. From the two aspects of fake review detection (FRD) and user preference analysis, we reviewed the main methods, techniques, and associated challenges in the literature.

Fake review detection

Research shows that the accuracy of manually identifying fake reviews is only 53.1–61.9% [4]. Researchers generally classify fake reviews into the following three types[5–7]. The first category is untrue opinions, i.e., deliberately posting negative reviews to damage the image of the product/brand or posting positive reviews to promote the product/brand. The second category is reviews that focus on the product brand rather than the product itself. The third category is nonreviews, which do not provide real opinions or are just advertisements. Jindal and Liu[6] are considered pioneers in the automatic detection of fake reviews. Inspired by Jindal and Liu, many heuristic methods have been proposed. The features widely used in the FRD can be divided into (1) review centric, (2) reviewer centric, (3) metadata centric, and (4) product centric[7].

Review-centric features are the most commonly used features in FRD models and involve the classification and summarization of opinions via psychological analysis, natural language processing (NLP) and data mining. Reference[8] is one of the first empirical studies on the relationship between psychological suggestion and fake reviews and identifies fake reviews on the basis of psychological cues, time distance, and reviewer location. Wang et al.[9] further proposed using the interaction of affective cues, cognitive cues, and review titers to detect fake reviews. Hajek[10] proposed two neural network models that integrate traditional bag-of-words as well as the word context and user sentiment. Bathla[11] extracted aspects and sentiments from reviews and fed them to deep learning models, including convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, to detect fake reviews. Luo[12] developed a supervised probabilistic method that uses differences in the distributions of genuine and fake reviews to enable review monitoring. In addition, the frequency of adverbs, verbs and pronouns can be used to detect fake reviews[13–15].

The reviewer-centric category detects fake reviews by evaluating the basic behavioral attributes of the publisher rather than the content of the review itself, for example, the social background[16,17], personal characteristics[18] and behavioral habits[19,20] of the review publisher. Recent reviewer-centric research has also focused on detecting reviews written by the same reviewer under different usernames[21].

Metadata can describe relevant information of the data, including time features, ID, length, rating, data, IP address at the time of publication, and date variance of reviews[5]. Metadata can be used to detect unusual reviews. When a user is identified as an untrusted reviewer, all reviews linked to that user are classified as fake reviews.

The fourth category is product-centric detection methods, such as price or the average product rating[22,23].

User preference analysis

A deep understanding of user preferences is at the core of product design and business practice. To obtain user preferences, several classic techniques are widely used in the product design process, such as questionnaires[24], interviews and focus groups[25], prototyping[26], and brainstorming[27]. However, the aforementioned approach requires constant interaction between designers and users, and the process of the operations team gathering and analyzing user preferences can take weeks or even months. Therefore, these methods are not suitable for product development projects that are expected to be completed quickly.

The digitization of online user activities, as well as advances in natural language processing, sentiment recognition, and intelligent analytics technology, enables quick identification of user preferences. The task of coarse-grained preference analysis is to treat entire reviews as a whole sentimental expression and assign them a unique affective polarity label. Fu et al.[28] developed a phrase recursive autoencoder (PRAE) model to detect the sentiment of sentences. Zhang et al.[29] used a support vector machine with naive Bayes features to optimize CNNs and improve the accuracy of sentiment classification. Phan et al.[30] proposed a fuzzy-enhanced deep neural network (FeDN2), which deepens deep neural networks by adding fuzzy and defuzzy classes to improve sentence-level sentiment analysis. Recently, coarse-grained sentiment analysis has undergone a shift toward large pretrained language models (e.g., BERT, RoBERTa and XLNet)[31].

Fine-grained sentiment analysis focuses on specific aspects or targets in sentences to determine sentiment polarity, which is considered more challenging than coarse-grained tasks[32]. Previous studies have addressed several tasks at the aspect level, such as aspect term extraction, aspect term classification, aspect term polarity classification, and aspect term classification with polarity[33]. Fine-grained sentiment analysis methods can be roughly divided into knowledge-based methods and learning-based methods[34]. Knowledge-based methods focus on the construction or utilization of knowledge bases, and several typical examples are those in [35,36]. Learning-based methods typically utilize statistical learning methods (such as latent Dirichlet assignment[37]) and neural network models[38]. Sentiment analysis using attentional mechanisms[39], pretrained language models[40] and hybrid neural networks[41] has become popular recently.

Research gaps

Despite notable achievements in both fake review detection and user preference analysis, several critical gaps remain unaddressed.

(1) Lack of detection for AI-generated user reviews.

Traditional fake review detection methods focus primarily on human-written fake reviews and rely on linguistic or behavioral patterns that are less effective when they are applied to AI-generated content. AI-generated reviews exhibit high fluency and contextual coherence, making their detection by existing approaches more challenging.

(2) Limited integration of fake review detection and preference analysis.

Most existing studies treat fake review detection and user preference analysis as independent tasks. However, integrating these two tasks could improve the reliability of insights derived from user feedback by ensuring that preference analysis is based on authentic reviews. This integration remains unexplored in the current literature.

(3) Insufficient consideration of the hidden design features.

Current user preference analysis methods often focus on explicit product attributes mentioned in reviews, such as appearance or functionality. However, less obvious design features, such as material durability, energy efficiency, or ergonomic convenience, can also significantly impact user satisfaction. Methods capable of capturing both explicit and implicit preferences are still underdeveloped.

To address these research gaps, the original contribution of this paper is as follows.

(1) A method leveraging pretrained language models is proposed to distinguish genuine user reviews from AI-generated reviews, mitigating the potentially misleading effects of deceptive content.
(2) On the basis of detecting fake reviews, a user preference analysis method that encompasses both attribute-level and feature-level analysis to ensure a comprehensive understanding of user preferences is introduced. By integrating fake review detection, this approach provides more reliable insights for product design.
(3) The method is capable of addressing the challenges of data scarcity. Extensive experiments, including comparisons with baseline models, demonstrate the effectiveness of the proposed method. Additionally, the robustness of the method is verified under low-resource conditions, highlighting its practicality in scenarios with limited data availability.

Methodology

Task description

Generative AI systems have demonstrated remarkable capabilities in areas of cognition that were believed to be exclusive to humans[42]. However, AI-generated fake reviews pose significant challenges to the reliability of user feedback data. To avoid the potential impact of misleading insights on product design and consumer trust, the proposed method integrates AI-generated review detection with user preference analysis, forming a comprehensive framework, as shown in Fig. 1. Detecting AI-generated reviews serves as a critical preliminary step to filter out potentially deceptive content, improving the accuracy and reliability of subsequent preference analysis.

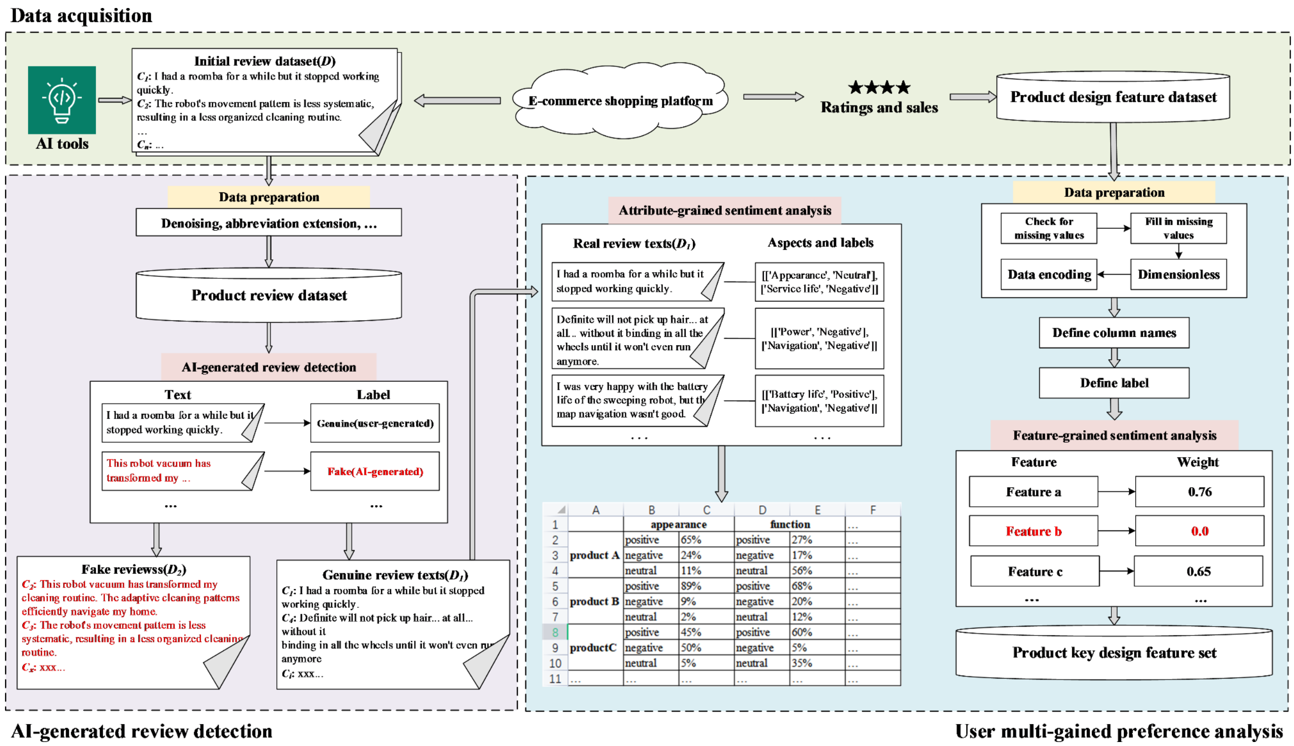


Fig. 1. Overall structure of the proposed architecture.

As shown in Fig. 1, the initial dataset $\boldsymbol{D} = [C_1, C_2, \cdots, C_i, \cdots, C_n]$ contains *n* user reviews. Some of them are posted by users, whereas others are generated by AI. In other words, the AI-generated review detection task can be considered a text classification task. A classifier that can identify fake reviews in $\boldsymbol{D}$ needs to be designed. Then, $\boldsymbol{D}$ is divided into two subsets, where $\boldsymbol{D}_1$ consists of the reviews generated by users and $\boldsymbol{D}_2$ consists of the reviews generated by AI.

The subsequent multigrained preference analysis task is described in $\boldsymbol{D}_1$. In this study, product attributes refer to the inherent properties of a product. Using robotic vacuum cleaners as an example, attributes such as lifespan, suction power, and noise level are considered product attributes. The attribute-grained preference analysis task is aimed at identifying aspects that users like or dislike and comparing them with competing products. Moreover, one filtered genuine review $\boldsymbol{C} = [w_1, w_2, \cdots, w_i, \cdots w_m]$ consists of *m* words, and some of the words$\{w_{i1}, \cdots, w_{ik}\}$ are preidentified targets $\boldsymbol{T} = [t_1, \cdots, t_k]$. The review contains *k* words, and $\boldsymbol{T}$ is a subsequence of $\boldsymbol{C}$. The sentiment polarity $\boldsymbol{Y}$ of review $\boldsymbol{C}$ toward the target $\boldsymbol{T}$needs to be determined, where $\boldsymbol{Y} \in \{positive, negative, neutral, none\}$. In our research, $\boldsymbol{P} = \{P_1, P_2, \cdots, P_n\}$ is assumed to be a set of competing products of different brands. For each product P_i, the attribute-grained preference analysis task is described as follows: The sentiment polarity $\boldsymbol{Y}$ is predicted over the full set of target-aspect pairs$\{(t, a) : t \in \boldsymbol{T}, a \in \boldsymbol{A}\}$ given the reviews in $\boldsymbol{D}_1$, a set of target entities $\boldsymbol{T}$ and a fixed aspect set $\boldsymbol{A}$.

Design features refer to the tangible and measurable characteristics of a product. These include quantifiable features such as length, width, height, and weight, which can be expressed directly through numerical values. Additionally, design features may also encompass qualitative characteristics that cannot be quantified, such as color (e.g., black or white). To comprehensively evaluate the importance of design features, the data structure of product design features is described below. Let $\boldsymbol{P}$ be the collected product design feature dataset and let $\boldsymbol{P}$ contain M similar products of different brands. Each product contains N design features and a label that can reflect both sales volume and user satisfaction. The original design feature set is expressed as $\mathbf{DF} = \{DF_1, DF_2, \cdots DF_N\}$. The design features of product $i\,(i = 1, \cdots, N)$ is represented as an N-dimensional vector $\mathbf{X_i} = (X_{i1}, X_{i2}, \cdots, X_{iN})$, where X_{ij} represents the specific value of the design feature and the label is represented by y_i. Then, the product design feature dataset $\boldsymbol{P}$ is represented as a matrix $(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X} = (X_1, X_2, \cdots, X_M)^T$, $\mathbf{Y} = (y_1, y_2, \cdots, y_M)^T$. The task of feature-grained preference analysis is defined as follows: a key design feature subset $\mathbf{KDF}$ is identified from the original design feature set $\mathbf{DF}$ so that $\mathbf{KDF}$ contains product design features that are positively correlated with the label y_i, and irrelevant and redundant design features are eliminated as much as possible.

AI-generated review detection

Data acquisition

Scrapy is an open-source web crawler framework developed by Scrapinghub[43]. The framework is written in Python and is widely used in various data mining problems. At the beginning of this work, the Scrapy framework was used to collect a dataset of genuine reviews, as shown in Fig. 2. ChatGPT3.5 (also known as ChatGPT in its first version) and ERNIE Bot, developed by Baidu, Inc., are used to generate fake reviews.

AI-generated review detection

The advent of BERT[44] marked a significant advancement in the field of NLP, because it better understands the context and semantic relationships of language. DeBERTa[45] improves BERT with two new components: an enhanced mask decoder and disentangled attention. DeBERTaV3 builds on the improvements of DeBERTa by incorporating a more efficient pretraining strategy and optimizing the MLM objective, resulting in better training efficiency and superior performance on NLP tasks[46].

Pretrained language models are generally trained on vast text corpora to acquire contextual word representations through a self-supervised objective, known as mask language modeling (MLM). Given a sequence $\mathbf{X} = \{x_i\}$, 15% of the words in the corpus are randomly masked with the [mask] token, creating the new sequence $\tilde{\mathbf{X}}$. This text is then fed to a multilayered stack of transformer encoders to train a language model parameterized by θ, which reconstructs $\mathbf{X}$ by predicting the masked tokens $\tilde{x}$, conditioned on $\tilde{\mathbf{X}}$:

$$\max_{\theta} \log p_{\theta}\left(\mathbf{X} \left| \tilde{\mathbf{X}}\right.\right) = \max_{\theta} \sum_{i \in \mathcal{C}} \log p_{\theta}\left(\tilde{x}_i = x_i \left| \tilde{\mathbf{X}}\right.\right), \tag{1}$$

where C is the index set of the masked tokens in the sequence.

Unlike BERT, which uses a single transformer encoder and is trained through MLM, the feature extractor for the proposed AI-generated review detection model employs ELECTRA-style[47] style pretraining and two transformer encoders. The generator is trained via the MLM, whereas the discriminator is trained via a token-level binary classification approach. We denote the parameters of the generator as θ_G and those of the discriminator as θ_D. The training goal for the discriminator is referred to as replaced token detection (RTD). The loss function for the generator is expressed as follows:

$$L_{MLM} = \mathbb{E}\left(-\sum_{i \in \mathcal{C}} \log p_{\theta_G}\left(\tilde{x}_{i,G} = x_i \left| \tilde{\mathbf{X}}_G\right.\right)\right), \tag{2}$$

where $\tilde{\mathbf{X}}_G$ is derived by randomly masking 15% of the tokens in X.

The input for the discriminator is created by substituting masked tokens with new tokens sampled on the basis of the output probabilities from the generator:

$$\tilde{x}_{i,D} = \begin{cases} \tilde{x}_i \sim p_{\theta_G}\left(\tilde{x}_{i,G} = x_i \left| \tilde{\mathbf{X}}_G\right.\right), & i \in \mathcal{C} \\ x_i, & i \notin \mathcal{C} \end{cases}, \tag{3}$$

The loss function of the discriminator is expressed as follows:

$$L_{RTD} = \mathbb{E}\left(-\sum_{i} \log p_{\theta_D}\left(1\left(\tilde{x}_{i,D} = x_i\right) \left| \tilde{\mathbf{X}}_D, i\right.\right)\right), \tag{4}$$

where 1 is the indicator function and $\tilde{\mathbf{X}}_D$ is the input to the discriminator formed via Eq. 3. The losses L_{MLM} and L_{RTD} are optimized jointly, with the total loss $L = L_{MLM} + \lambda L_{RTD}$, and λ is the weight assigned to the discriminator's loss L_{RTD}, which is 50.

Additionally, the discriminator implements a new embedding-sharing method referred to as gradient disentangled embedding sharing (GDES), which shares the same token embedding for the discriminators and generators in ELECTRA, as shown in Fig. 3. $\boldsymbol{E}$ and $g_{\mathbf{E}}$ are token embeddings and their gradients, respectively, and sg is the stop gradient operator that prevents the discriminator from updating $\boldsymbol{E}_G$.

AI-generated review detection is considered a classic binary text classification task. The given reviews are separated into tokens and then sent to the pretrained model as the input. To obtain the complete meaning of the entire review, the output embeddings of each token are averaged through an average pooling layer. The binary classification task is designed by feeding the averaged encoding into the fully connected layer with dropout.

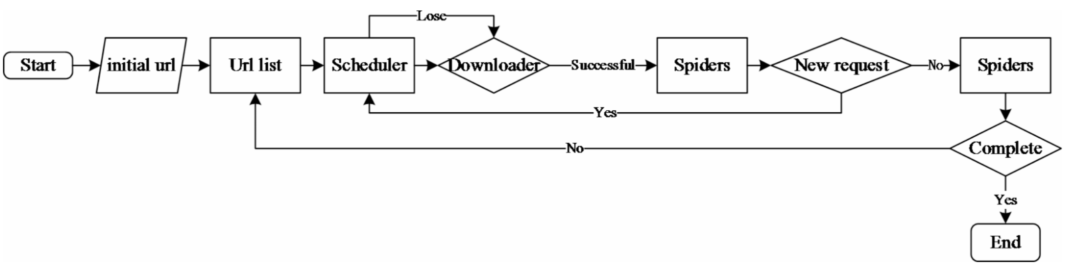


Fig. 2. Acquisition process of genuine reviews.

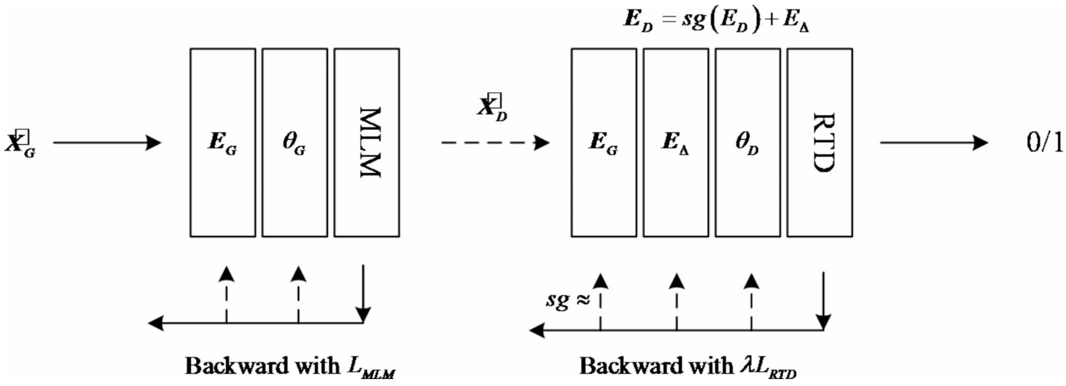


Fig. 3. GDES.

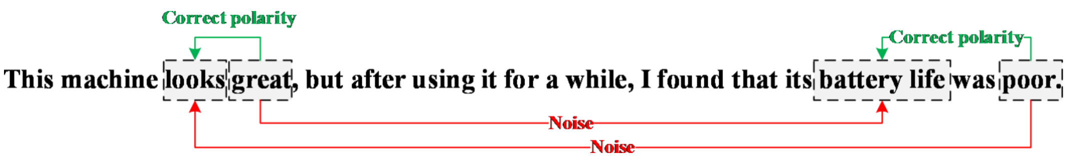


Fig. 4. Local context influence in fine-grained sentiment analysis.

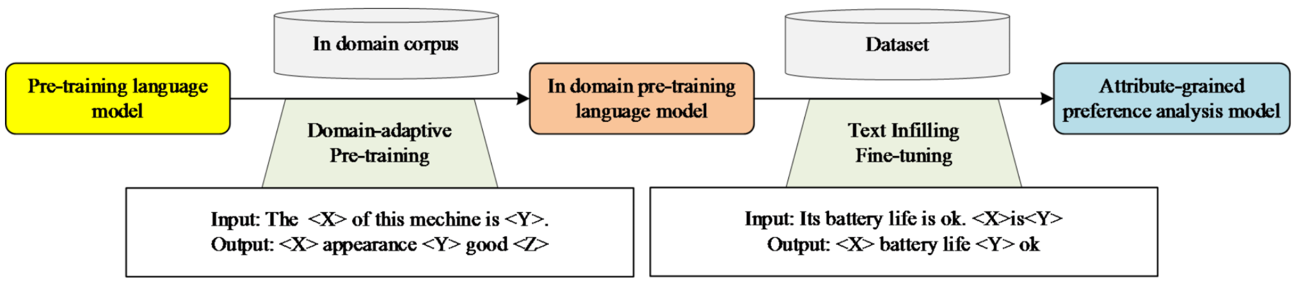


Fig. 5. Overview of the attribute-grained preference analysis task.

Multigrained preference analysis

Attribute-grained preference analysis

Many user reviews involve multiple aspects. Even in one review, users may have different opinions and sentiment tendencies on different aspects, which makes the process of attribute-grained user sentiment analysis complicated. As shown in Fig. 4, when there are multiple aspects, the local context also affects the polarity classification of sentiment. Some sentiment words exist in the context corresponding to different aspects, resulting in noise interference to identify emotional polarity. Therefore, the correspondence between words needs to be distinguished correctly.

Motivated by the work of Wang et al.[48], the attribute-grained preference analysis task was converted to a text-filling task. An example is as follows.

- Input: I have been using this robot vacuum for six months, and its battery life is ok. < X > is < Y >.
- Output: < X > battery life < Y > ok.

As shown in the above example, < X > and < Y > can be used as placeholders for aspect and sentiment polarity, respectively. In other words, the aspects and sentiments to be predicted can be considered corrupted tokens in the input. To make the modified input more natural, the sentiment labels are mapped to the corresponding label words, i.e., neutral to ok, positive to great and negative to bad.

Domain adaptive pretraining (DAPT) was proposed by Gururangan et al.[49]. DAPT pretrains on a large-scale unlabeled corpus related to the domain and then fine-tunes specific tasks, which can bridge the domain gap. In the attribute-grained preference analysis task, the language model is pretrained via text-filling targets on an in-domain corpus, as shown in Fig. 5.

Both text-infilling fine-tuning and domain-adaptive pretraining are optimized with the text-infilling objective. The text-infilling loss is defined as:

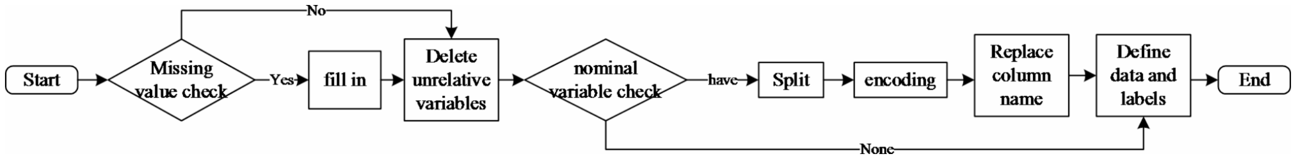


Fig. 6. Feature preprocessing.

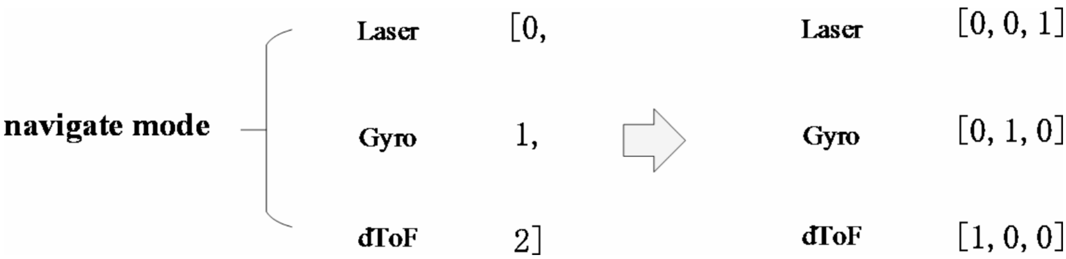


Fig. 7. One-hot encoding.

$$Loss\left(\theta\right) = \sum_{t=1}^{|x|} - \log P_\theta \left(x_t^{mask} \left| x^{\backslash mask}, x_{<t}^{mask} \right. \right), \tag{1}$$

where θ denotes the trainable model parameters; $x^{\backslash mask}$ and x_t^{mask} are the corrupted input and corresponding target spans, respectively; $|x|$ denotes the length of the target sequence; and $x_{<t}^{mask}$ represents the target sequence that has been generated thus far.

A correction strategy based on fuzzy matching prediction is adopted to avoid the possibility that the predicted aspect may differ from the basic fact in terms of tense and capitalization. More information about this strategy is provided in[48].

Feature-grained preference analysis

The initial step in implementing the feature-grained preference analysis outlined in Sect. 3.1 is to collect datasets of product design features. The dataset includes appearance design parameters, feature design parameters, functional design parameters, and the user satisfaction label. The collected features may have missing values, but these features cannot be abandoned directly. In addition, most algorithms can handle only numerical data, so it is necessary to encode text features and convert them to numerical variables. Figure 6 shows the feature preprocessing procedure.

As shown in Fig. 6, feature preprocessing is divided into four steps. The first step is to check and fill in missing values. Simple padding, manual padding, and hot deck imputation are used to fill in missing data. The second step is dimensionless. To avoid the adverse effects of one or several design features with a large range of values on modeling, min–max normalization (also known as min–max scaling) is used for the dimensionless processing of data. The normalized data range will be between [0, 1]. The third step is data encoding. Many product design feature data are not in numerical form. For example, the navigation method of a sweeping robot can be "laser", "gyroscope", or "dToF", and the layout of buttons may include "up and down layout" or "left and right layout". Textual data must be converted to numerical data, which is also known as data encoding. In traditional encoding methods, the three classification features are converted to numerical data [0,1,2]. Regarding the algorithm, these three numbers are not equal but are continuous and computable. This process disregards the mathematical properties contained in the numbers and transmits incorrect information to the algorithm. Therefore, one-hot encoding is used to convert nominal variables to dummy variables, as shown in Fig. 7. The fourth step is to define the column names, labels and data.

Lasso[50] is a widely used and effective feature selection method that induces sparsity in the feature coefficients by introducing L_1 regularization. One limitation of Lasso is that when variables are selected, it can select only one or a few highly correlated important variables rather than all important variables. Therefore, random Lasso is used to calculate the importance of each design feature[51]. The specific steps are as follows:

(1) Calculation of importance measures for all features.

Step 1: The bootstrap[52] is used to extract B samples from the original dataset.
Step 2: For the i-th bootstrap sample b_i, $i = 1, 2, \cdots, B$, q_1 product design features are selected from p variables($q_1 \leqslant p$). Then, the Lasso algorithm is used to obtain the estimate $\hat{\beta}_j^{(b_i)}$ of these q_1 features, where $(j = 1, 2, \cdots, p)$. The coefficients of the remaining $q - q_1$ features that were not selected are set to 0.

Step 3: The importance measure I_j of the design feature X_i is calculated according to Eq. (2), where $j \in \{1, 2, \cdots, p\}$. C_j represents the number of times feature X_i is selected in B bootstrap samples.

$$I_j = \left| C_j^{-1} \sum\nolimits_{i=1}^{B} \hat{\beta}_j^{(b_i)} \right| \tag{2}$$

(2) Weight calculation.

Step 4: As in step 1, B Bootstrap samples were extracted from the original dataset.
Step 5: Similarly, for the i-th bootstrap sample b_i, $i = 1, 2, \cdots, B$, q_2 $(q_2 \leqslant p)$ candidate features are selected on the basis of the feature importance calculated in Step 3, and then the Lasso algorithm is used to obtain the estimate $\hat{\beta}_j^{(b_i)}$ of these q_2 variables, where $(j = 1, 2, \cdots, p)$. The estimates of the $p - q_2$ features are set to 0.
Step 6: The final weight coefficient estimate $\hat{\beta}_j$ of each feature is calculated according to Eq. (3), where C'_j is the number of times feature X_i is selected in B bootstrap samples.

$$\hat{\beta}_j = C_j'^{-1} \sum_{i=1}^{B} \hat{\beta}_j^{(b_i)} \tag{3}$$

The algorithm flow is shown in Fig. 8.

Experiment and results

Data preparation

To develop a dataset for AI-generated review detection, user reviews of sweeping robots were collected from the Amazon platform. To ensure authenticity, genuine reviews were selected from June 2021 to June 2022, prior to the public release of ChatGPT in November 2022. These raw reviews contained noise such as symbols, usernames, website links, and misleading tags. Preprocessing included replacing usernames and links with special tokens and expanding common abbreviations (e.g., "U," "idk," and "omg") into fully spelled words. Elements such as spelling errors and repeated punctuation were retained to preserve the expressive nature of user feedback. AI-generated reviews were created via ChatGPT 3.5 and ERNIE Bot. These tools were instructed to mimic real-world users, generate reviews of varying lengths and sentiments, and focus on aspects such as cleaning quality and obstacle navigation. The dataset for the AI-generated text detection module comprises a total of 768 user reviews, including 312 genuine user reviews and 456 AI-generated reviews. Table 2 shows a selection of reviews and their corresponding labels.

In the attribute-grained preference analysis task, 20,000 unlabeled reviews of robotic vacuum cleaners were collected from the Amazon platform for domain-adaptive pretraining. Three experimental setups were designed: (1) Individual experiment: Reviews from a single brand (iLife) were analyzed to assess user satisfaction with specific product attributes. (2) Comparative experiment: Reviews from three brands (iLife, iRobot, and Tesvor) were utilized to identify competitive strengths and weaknesses across product attributes. (3) Migration experiment: The model was trained on reviews from the three aforementioned brands and tested on reviews from three additional brands (Ecovacs, Shark, and Coredy) to evaluate its generalizability. All selected products were priced between \$150 and \$250. The aspect-level corpora were annotated following the SemEval 2014 Task 4 guidelines and processed into two forms, as shown in Fig. 9. The training sets for these experiments consisted of 768, 2200, and 2200 reviews, whereas the validation and test sets each contained 256 reviews.

In feature-grained preference analysis, the dataset consists of 31 design features from 32 products: 22 numerical features of the product, such as dimension, charging power, net weight, battery capacity, and suction, and 9 nominal variable features, such as navigation mode, sterilization function, drying function, virtual wall function, and dust collection function. Several design features are shown in Table 3.

Experimental settings

The deep learning programming framework PyTorch[53] is adopted. The model training is completed on an NVIDIA 3070 GPU (12 GB of RAM). The number of samples and hardware conditions were comprehensively considered during the training process. The hyperparameter settings of the proposed method were meticulously optimized, as detailed in Table 4. The optimizers are Adam[54] and AdamW[55].

Experimental results

We conducted 10-fold cross-validation to evaluate the performance of the proposed AI-generated review detection model. The evaluation metrics were composed of accuracy, precision, recall, and the F1 measure to assess the classification capabilities of the model comprehensively[56]. Table 5 presents the results of each fold, as well as the mean and standard deviation across all folds.

In addition, Fig. 10 shows the results of the 10-fold cross-validation with standard deviations, where each bar represents the mean score across all folds and the error bars indicate the standard deviation. The low standard deviations across all the metrics emphasize the robustness and reliability of the model in distinguishing between AI-generated reviews and genuine reviews.

Figure 11 shows the predicted outcome of the model when it is presented with a sentence of a user review. In Fig. 11, the user reviews "I had a roomba for a while but it stopped working quickly (it no longer picks up trash)." is considered "genuine".

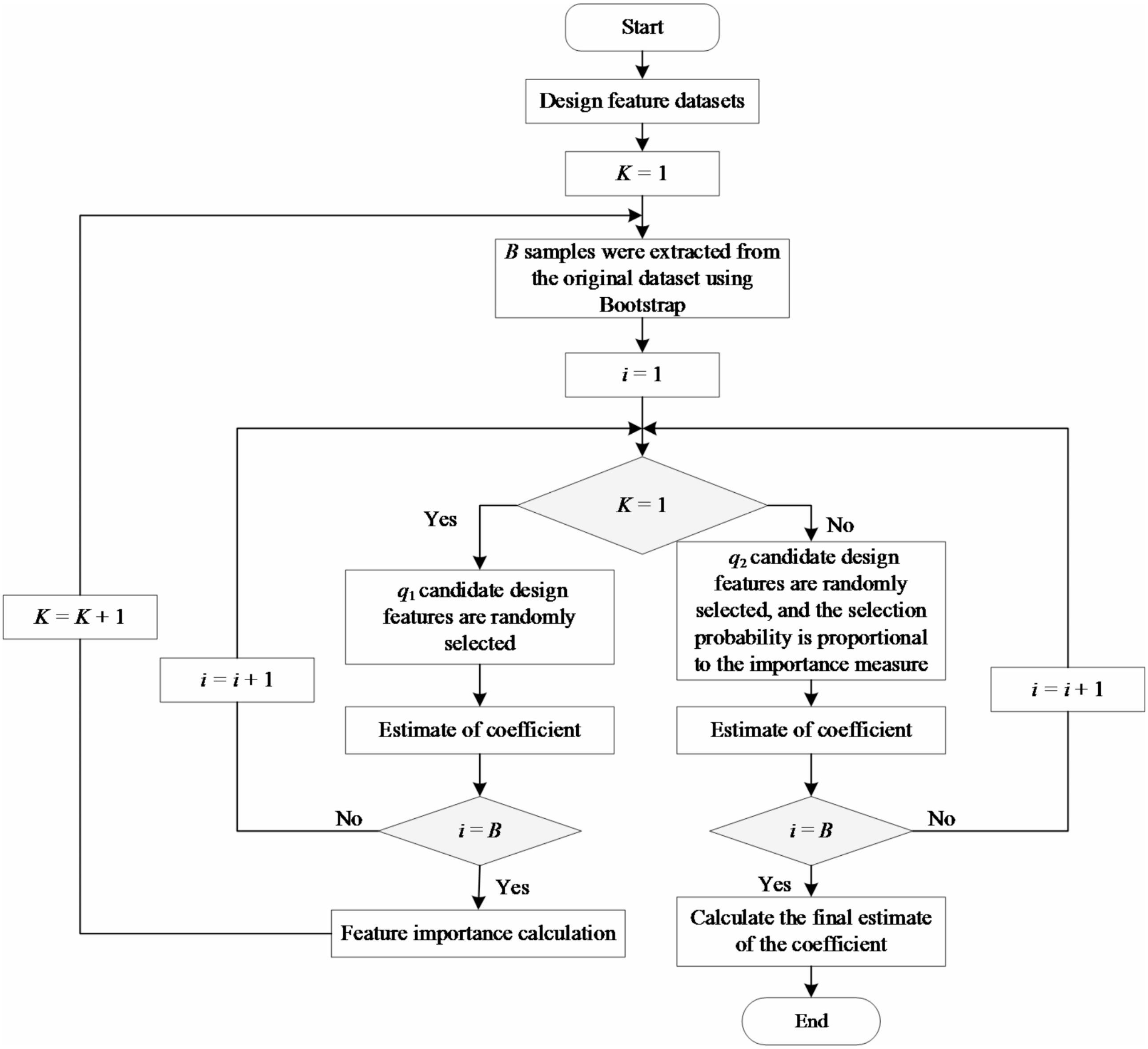


Fig. 8. Feature importance calculation.

User review	Label
Worth every penny! My floors have never looked cleaner. The scheduling feature is a lifesaver, and I love coming home to a pristine house. I cannot imagine my life without it now.	Fake
As a busy professional, time is of the essence. This robot vacuum has given me back valuable time that I used to spend cleaning. It is smart, effective, and a true time-saver.	Fake
If it is left to its own devices in my apartment, it will get stuck on the tassels on my rug, climb up my shoes and push around lightweight objects or knock them over.	Genuine
However, a couple of weeks ago, the robot vacuum just stopped working. It would start on a cleaning cycle then just stop even though nothing was in its way.	Genuine

Table 2. Sample dataset for the AI-generated review detection model.

In the Attribute-Grained Preference Analysis Task, three random numbers of seeds—29, 33, 153, 42 and 757—were used in the experiment. For each experiment, we recorded the results and computed the mean and standard deviation of the evaluation metrics. These results are presented in Table 6.

During the experiment, the weights and biases service (Weights & Biases software, wandb.com, CA, USA) was used to track all the different configurations of the parameters. The changes in the F1 measure on the validation set across training epochs for seeds 27, 42, and 757 are shown in Fig. 12. The horizontal axis in Fig. 12 represents the training epochs, and the vertical axis represents the average F1 measure of all steps in each epoch.

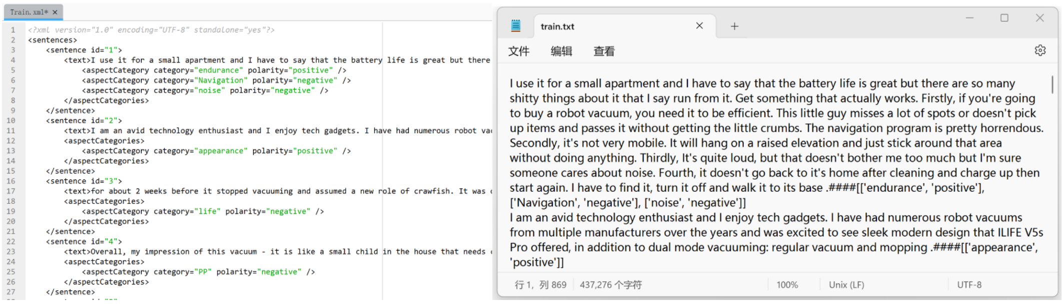


Fig. 9. Datasets in attribute-granular preference analysis tasks.

Product number	Maximum noise (dB)	Charging power (W)	Dust box capacity (ml)	Endurance (min)	Suction (Pa)	Navigation mode	…	Water tank capacity (ml)	Product height (mm)
1	72	67	470	150	2500	Laser	…	400	96.5
2	75	45	400	240	5000	Tof radar	…	800	103.5
3	65	45	430	180	2500	Laser	…	500	108
4	66	40	430	90	2300	Laser	…	240	93
5	75	40	300	150	2200	Visual	…	400	
6	72	68	470	150	2500	Laser	…	300	96
7	77	33	450	120	2800	Laser	…	190	96
8	72	40	550	110	2700	Visual	…	250	81
9	72	40	450	120	3000	Laser	…	450	106
10	63	40	550	180	3000	Laser	…	250	81.5

Table 3. Product design feature dataset.

	Value	
Hyperparameter	AI-generated review detection	Attribute-grained preference analysis
Maximum sequence length	512	128
Batch size	8	16
Learning rate	0.00001	0.0003
Epochs	6.0	20.0
Warmup_rate	–	0.06
Weight_decay	–	0
Adam_epslon	–	0.00000001

Table 4. Hyperparameter settings.

The training process is completed after 20 epochs. The model and parameters with the highest F1 measures on the validation set are retained.

Furthermore, we obtained user reviews of five sweeping robots on the Amazon platform, with 300 reviews for each product. The selected sweeping robots represent five different brands, and their prices range from US$179 to US$209. When consumers want to buy a sweeping robot within the range of approximately US$200, all five brands may be considered. The weight parameters obtained after the seed 42 training is completed are used to analyze the user comments. The percentage of user satisfaction for each aspect was calculated on the basis of the output results (see Table 7). As shown in Table 7, user satisfaction with product appearance and suction is generally high. In terms of noise, navigation, and battery life, user satisfaction differs greatly between different products. Therefore, when the budget is limited, users can choose products on the basis of their satisfaction with the aspects they care more about, and R&D personnel can also clarify the direction of the next improvement on the basis of the results of competitive product analysis.

In feature-grained preference analysis, the evaluation score and sales volume are comprehensively considered, and the label is $Y = 10\left[S - \text{int}\left(S\right)\right] + 1e^{-5}V$, where S is the consumer rating and V is the product sales volume. The hyperparameter alpha was 0.0035. The design features and their importance weights were obtained, as shown in Table 8.

Fold and statistics	Accuracy	Precision	Recall	F1 measure
1	0.9610	0.9655	0.9333	0.9492
2	0.9870	1.0000	0.9730	0.9863
3	0.9870	1.0000	0.9667	0.9831
4	0.9870	1.0000	0.9714	0.9855
5	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000
7	0.9870	1.0000	0.9688	0.9841
8	0.9870	1.0000	0.9756	0.9877
9	0.9737	1.0000	0.9487	0.9737
10	1.0000	1.0000	1.0000	1.0000
Mean	0.9870	0.9966	0.9737	0.9849
Std Dev	0.0108	0.0140	0.0271	0.0130

Table 5. Results of 10-fold cross-validation.

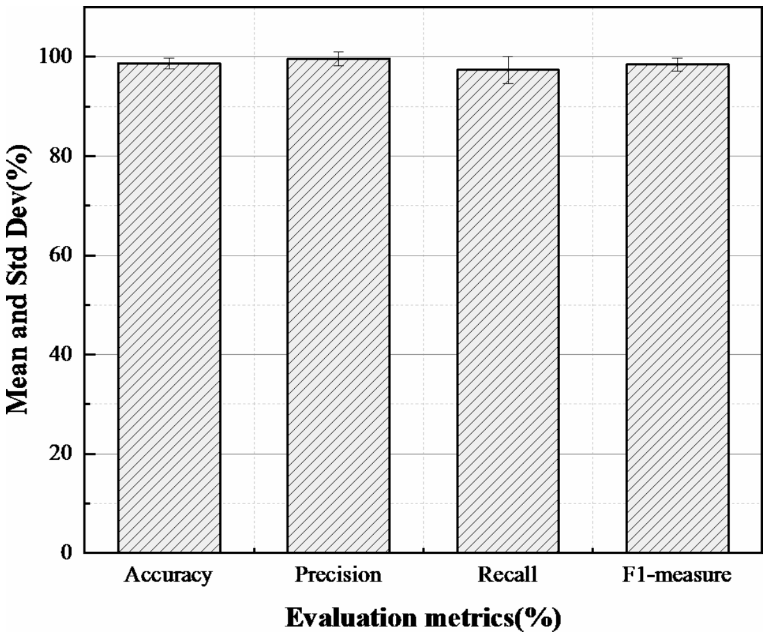


Fig. 10. 10-fold cross-validation with standard deviations.

```
1]:  model.eval()
     with torch.no_grad():
         text = "I had a roomba for a while but it stopped working quickly (no longer picking up trash)."
         token = [cls_token] + tokenizer.tokenize(text)[1:-1][0:config.max_seq_length - 2] + [sep_token]
         ids =  tokenizer.convert_tokens_to_ids(token)
         ids = torch.tensor([ids]).to(device)
         attention_mask = ids.ne(0)
         logits,_ = model(ids, attention_mask)
         pred = torch.argmax(logits, dim=1)
         for p in pred:
             print(labels_list[p])

     genuine
```

Fig. 11. Test results of the AI-generated review detection model.

Seed values and statistics	Evaluation metrics	Individual experiment	Comparative experiment	Migration experiment
Seed 29	Precision	0.8645	**0.9317**	0.9331
	Recall	0.8247	**0.9479**	0.9208
	F1-measure	0.8441	**0.9398**	0.9269
Seed 33	Precision	0.8488	0.9247	**0.9336**
	Recall	0.7902	0.9375	0.9274
	F1-measure	0.8185	0.9310	0.9305
Seed 42	Precision	0.8433	0.9024	0.9251
	Recall	**0.8506**	0.9306	**0.9372**
	F1-measure	0.8469	0.9162	**0.9312**
Seed 153	Precision	**0.8728**	0.9285	0.9331
	Recall	0.8477	0.9201	0.9208
	F1-measure	**0.8601**	0.9250	0.9269
Seed 757	Precision	0.8077	0.9189	0.9276
	Recall	0.8448	0.9444	0.9307
	F1-measure	0.8258	0.9315	0.9292
Mean	Precision	0.8474	0.9212	0.9305
	Recall	0.8316	0.9361	0.9274
	F1-measure	0.8391	0.9287	0.9289
Std Dev	Precision	0.0225	0.0103	0.0035
	Recall	0.0226	0.0100	0.0062
	F1 measure	0.0150	0.0078	0.0018

Table 6. Training results of the attribute-grained preference analysis model. In each experimental module, the highest values of each evaluation metric are in bold.

The information that ordinary consumers can express is limited. Obtaining their preferences comprehensively is especially important. In the current big data environment, the use of intelligent algorithms to construct a model for discovering key design features of products, considering the issues of the priority of design features, design costs, and efficiency, can help designers focus on key design parameters and information during the design process. On the basis of their close relationship with product functionality and consumer satisfaction, designers can quickly obtain design solutions.

Comparisons and few-shot experiments

In a previous study, a method for AI-generated review detection and consumer preference analysis was proposed and experimentally validated. In this section, the experimental results are compared with those of other existing methods. Moreover, the number of training samples was reduced to explore the experimental results of the proposed method under the few-shot setting. The limitations and suggestions for future research directions are discussed.

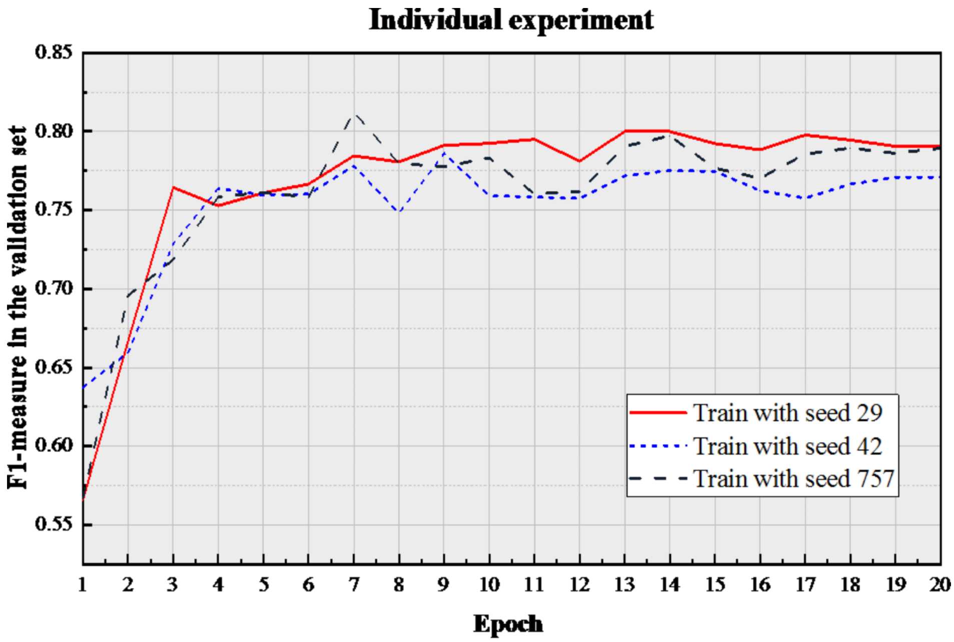
Comparisons

The proposed AI-generated review detection model is compared with a range of baseline methods, including BERT, ALBERT, RoBERTa, support vector machines (SVMs), naive Bayes, and gradient boosting decision trees (GBDTs). Similarly, all algorithms were evaluated via 10-fold cross-validation, with the metrics outlined in Sect. 4.3. The average performance metrics across all the folds are summarized in Table 9 The results in Table 9 indicate that the proposed model outperforms all the baseline methods in terms of accuracy, precision, recall, and F1 measure. ALBERT achieves an F1 score of 0.9825, and BERT reaches 0.9719, making them the closest competitors.

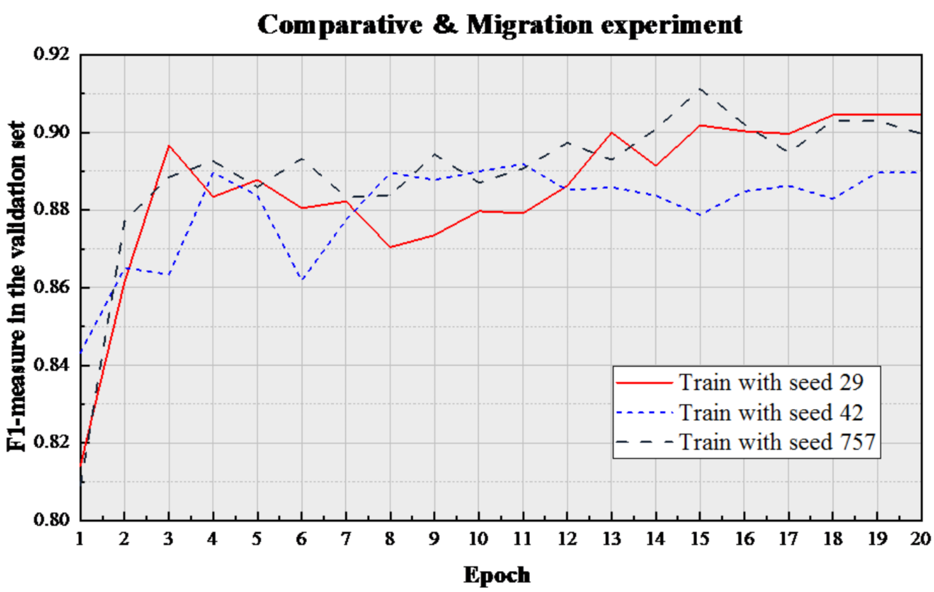
BERT, BERT-PT, RoBERTa, attention-based LSTM, and SVMs combined with TF-IDF were used as baseline models for comparison with the proposed attribute-grained preference analysis method. The training results when the datasets of individual experiments, comparison experiments and migration experiments remained unchanged are shown in Table 10. The F1 measures of our model in the three experimental modules are 2.57%, 2.43% and 1.7% higher, respectively. This improvement reflects the superiority of the proposed model in understanding and analyzing consumer preferences. Compared with the baseline model, treating the attribute-grained preference analysis task as a text-filling task and conducting domain adaptive pretraining can more accurately capture the nuances of consumer preferences, providing enterprises with more precise market positioning and product improvement strategies.

Few-shot experiments

Compared with the previously mentioned baseline model, in scenarios with fewer training set samples, the performance gap between the baseline model and the proposed model is more significant. In practical applications, data scarcity is challenging, especially when new products are just launched. Therefore, a few-shot experiment was designed to explore the applicability of the method in more challenging environments.



(a) Individual experiment



(b) Comparative and migration experiments

Fig. 12. Changes in the F1 measure with increasing number of epochs in the validation set.

In the AI-generated review detection module, the performance of the model is verified with only 32 and 64 training samples. In the attribute-grained preference analysis module, the training set sample consists of 32, 128 and 256 reviews. The validation set and the test set have the same sample number as the training set. Table 11 shows the results of the few-shot experiments.

In the AI-generated review detection experiments, the proposed method still performed well with only 32 training samples, obtaining an average F1 measure of 0.8732. This finding shows that the proposed method still has good stability when a small number of samples is used and provides strong support for the actual application

Brands	Lefant	Mamnv	Honiture	Tikom	Tesvor
Images					
Price (USD)	199	179	189	199	209
Appearance					
Positive	91.3%	100%	96.2%	88.5%	86.4%
Negative	0	0	0	7.7%	4.5%
Neutral	8.7%	0	3.8%	3.8%	9.1%
Function					
Positive	58.1%	67.8%	60.3%	86.8%	61.2%
Negative	7%	15.8%	19.8%	4.1%	18.4%
Neutral	34.9%	16.4%	19.9%	91%	20.4%
Use-life					
Positive	0	0	16%	7.7%	0
Negative	81.8%	100%	80%	92.3%	85.7%
Neutral	18.2%	0%	4%	0	14.3%
User experience					
Positive	94.7%	72.5%	87.4%	76.2%	64.3%
Negative	3.3%	5.9%	2.9%	9.5%	22.4%
Neutral	2.0%	21.6%	9.7%	14.3%	13.3%
Navigation					
Positive	67.2%	43.0%	42.0%	86.2%	28.2%
Negative	26.9%	54.8%	36.2%	10.3%	59.8%
Neutral	5.9%	2.2%	21.8%	3.5%	12%
Suction					
Positive	96.5%	94.2%	97.6%	89.2%	91.9%
Negative	1.5%	0	0	6.8%	0%
Neutral	2.0%	5.8%	2.4%	4.0%	8.1%
Noise					
Positive	56.2%	88.6%	83.6%	91.8%	17.7%
Negative	33.7%	2.8%	9.8%	6.1%	61.3%
Neutral	10.1%	8.6%	6.6%	2.1%	21.0%
Battery life					
Positive	34.3%	21.1%	73.1%	85.2%	89.3%
Negative	57.1%	45.7%	19.2%	11.1%	3.6%
Neutral	8.6%	34.2%	7.7%	3.7%	7.1%

Table 7. Aspect-based user satisfaction.

scenario of data scarcity. According to the experimental results of the attribute-grained preference analysis, when the training set consists of 32 and 128 samples, the performance of the model is acceptable. As the number of training samples increases to 256, the F1 measure of the model significantly increases to 0.8125. On the basis of this finding, the minimum number of samples in the training set should be 256, which can balance the performance of the model with the cost of training.

Limitations and prospects

Although this study has made contributions to AI-generated review detection and multigrained sentiment analysis, there are still several limitations that need to be noted in future research. First, the proposed method only applies to UGC in the form of text. While the method can capture unusual features such as capitalization, misspellings, and repeated punctuation in statements, it has not taken into account other forms, such as memes and graphics. These forms of UGC can contain some information and sentiment that is not fully utilized owing to the limitations of the model. Thus, the proposed method may not be able to identify and analyze these nontextual forms of content effectively, thus limiting the full understanding of consumer preferences. Second, there is a progressive relationship between the AI-generated review detection task and the multigrained preference analysis task. This progressive relationship may result in several limitations, such as the possibility of longer processing times in practical applications and the inability to meet real-time requirements from the

Feature	Weight coefficients	Feature	Weight coefficients
Dust box capacity	0.72	Voice control	0.17
Navigation mode	0.92	Water tank volume	0.12
Hot air drying	0.745	Battery capacity	0.105
Weight	0.45	Sterilization function	0.075
Width	0.43	Maximum height	0.05
Automatic dust collection	0.52	Minimum noise	0.05
Height	0.35	Endurance	0.045
Automatic mop washing	0.42	Charging power	0.04
Endurance	0.295	Number of maps	0.03
Water tank gear	0.275	Virtual wall function	0.005
Suction gear	0.235	Obstacle crossing height	0
Maximum noise	0.195	Color quantity	0
Warranty time	0.175	Suction	0
3D human recognition	0.22	Applicable area	0
Identifying carpet	0.185		

Table 8. Design features and importance weight coefficients.

Method	Accuracy	Precision	Recall	F1 measure
SVM	0.9596	0.9637	0.9481	0.9556
Naive Bayes	0.9453	0.9121	0.9774	0.9422
GBDT	0.9154	0.9171	0.8976	0.9060
BERT	0.9753	0.9909	0.9542	0.9719
ALBERT	0.9844	0.9858	0.9797	0.9825
RoBERTa	0.9740	0.9883	0.9536	0.9710
Ours	0.9870	0.9966	0.9737	0.9849

Table 9. Comparative experimental results of AI-generated review detection.

Method	Evaluation metrics	Individual experiment	Comparative experiment	Migration experiment
Attention-based LSTM	Precision	0.7702	0.8524	0.8693
	Recall	0.7126	0.8021	0.8218
	F1 measure	0.7403	0.8265	0.8458
SVM combined with TF-IDF	Precision	0.7817	0.8282	0.8215
	Recall	0.6379	0.7535	0.8019
	F1 measure	0.7025	0.7891	0.8112
BERT	Precision	0.8265	0.8751	0.8953
	Recall	0.8012	0.8969	0.8987
	F1 measure	0.8136	0.8857	0.8970
RoBERTa	Precision	0.8343	0.8805	0.9079
	Recall	0.7960	0.8958	0.9109
	F1 measure	0.8147	0.8881	0.9093
BERT-PT	Precision	**0.8686**	0.8814	0.9142
	Recall	0.7787	0.9028	0.9142
	F1 measure	0.8212	0.8919	0.9142
Ours (seed 42)	Precision	0.8433	**0.9024**	**0.9251**
	Recall	**0.8506**	**0.9306**	**0.9372**
	F1 measure	**0.8469**	**0.9162**	**0.9312**

Table 10. Comparative experimental results of the attribute-based preference analysis. In each experimental module, the highest values of each evaluation metric are in bold.

Evaluation metrics	AI-generated review detection		Attribute-grained preference analysis		
	32	64	32	128	256
Precision	1.0000	0.8571	0.5089	0.7338	0.8014
Recall	0.7770	0.9375	0.4821	0.7635	0.8239
F1 measure	0.8732	0.8955	0.4947	0.7483	0.8125

Table 11. Results of few-shot experiments. When the number of samples in the training set is 32, three subsets are randomly selected from the dataset for training, and the results of these three training sessions are averaged.

demand side. Therefore, in future research, a more efficient one-step method needs to be designed to improve processing efficiency and real-time performance to better meet the needs of practical applications.

Conclusion
With the rapid development of large language models, identifying and preventing the spread of deceptive information and accurately capturing consumer preferences is particularly important. In response to this challenge, we focused on two core tasks: AI-generated review detection and consumer preference analysis. Building on the pretrained language model, we have proven the effectiveness of treating AI-generated review detection tasks as text classification tasks. The proposed method reduces the negative impact of deceptive information on society. Through the multigrained preference analysis method, the satisfaction of consumers with different aspects and the importance of design features can be obtained, providing more targeted strategies for product design and marketing. The proposed method achieves the mean F1 measures of 98.49% and 92.89% in AI-generated review detection and attribute-grained consumer preference analysis experiments, respectively, showing the robustness and potential of our model as a cornerstone for maintaining trustworthiness in product improvement. The advantages of the proposed method are demonstrated by comparing it with baselines. While the digital age offers unparalleled opportunities for innovation and efficiency, it also demands a heightened sense of responsibility from all the actors involved in product design and e-commerce.

Data availability
The code supporting the findings of this study is available from the corresponding author [Yang] and [Zhai] or the first author [Su] upon reasonable request.

Received: 8 October 2024; Accepted: 13 January 2025
Published online: 20 January 2025

References
1. Jing, L. et al. An integrated implicit user preference mining approach for uncertain conceptual design decision-making: a pipeline inspection trolley design case study. *Knowl. Based Syst.* **270**, 110524. https://doi.org/10.1016/j.knosys.2023.110524 (2023).
2. Lee, C., Tsang, Y., Chong, W., Au, Y. & Liang, J. Achieving eco-innovative smart glass design with the integration of opinion mining, QFD and TRIZ. *Sci. Rep.* **14**, 9822 (2024).
3. Wang, Q., Zhang, W., Li, J., Mai, F. & Ma, Z. Effect of online review sentiment on product sales: the moderating role of review credibility perception. *Comput. Hum. Behav.* **133**, 107272. https://doi.org/10.1016/j.chb.2022.107272 (2022).
4. Shunxiang, Z., Aoqiang, Z., Guangli, Z., Zhongliang, W. & KuanChing, L. Building fake review detection model based on sentiment intensity and PU Learning. *IEEE Trans. Neural Networks Learn. Syst.* **34**, 6926–6939. https://doi.org/10.1109/TNNLS.2023.3234427 (2023).
5. Mohawesh, R. et al. Fake reviews detection: a survey. *IEEE Access.* **9**, 65771–65802 (2021).
6. Jindal, N. & Liu, B. in *Proceedings of the 16th international conference on World Wide Web.* 1189–1190.
7. Walther, M., Jakobi, T., Watson, S. J. & Stevens, G. A systematic literature review about the consumers' side of fake review detection – which cues do consumers use to determine the veracity of online user reviews? *Computers Hum. Behav. Rep.* **10**, 100278. https://doi.org/10.1016/j.chbr.2023.100278 (2023).
8. Li, L., Lee, K. Y., Lee, M. & Yang, S. B. Unveiling the cloak of deviance: linguistic cues for psychological processes in fake online reviews. *Int. J. Hospitality Manage.* **87**, 102468. https://doi.org/10.1016/j.ijhm.2020.102468 (2020).
9. Wang, E. Y., Fong, L. H. N. & Law, R. Detecting fake hospitality reviews through the interplay of emotional cues, cognitive cues and review valence. *Int. J. Contemp. Hospitality Manage.* **34**, 184–200 (2022).
10. Hajek, P., Barushka, A. & Munk, M. Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. *Neural Comput. Appl.* **32**, 17259–17274 (2020).
11. Bathla, G., Singh, P., Singh, R. K., Cambria, E. & Tiwari, R. Intelligent fake reviews detection based on aspect extraction and analysis using deep learning. *Neural Comput. Appl.* **34**, 20213–20229 (2022).
12. Luo, J., Luo, J., Nan, G. & Li, D. Fake review detection system for online E-commerce platforms: a supervised general mixed probability approach. *Decis. Support Syst.* **175**, 114045. https://doi.org/10.1016/j.dss.2023.114045 (2023).
13. Li, J., Ott, M., Cardie, C. & Hovy, E. in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* 1566–1576.
14. Shukla, A., Wang, W., Gao, G. G. & Agarwal, R. Catch me if you can—detecting fraudulent online reviews of doctors using deep learning. *Ritu, Catch Me If You Can—Detecting Fraudulent Online Reviews of Doctors Using Deep Learning (January 14*, (2019). (2019).
15. Shan, G., Zhou, L. & Zhang, D. From conflicts and confusion to doubts: examining review inconsistency for fake review detection. *Decis. Support Syst.* **144**, 113513 (2021).
16. Zhao, J. et al. Fake reviewer detection based on reinforced heterogeneous graph neural networks. *Knowl. Based Syst.* **280**, 111029. https://doi.org/10.1016/j.knosys.2023.111029 (2023).

17. Cheng, L. C., Wu, Y. T., Chao, C. T. & Wang, J. H. Detecting fake reviewers from the social context with a graph neural network method. *Decis. Support Syst.* **179**, 114150. https://doi.org/10.1016/j.dss.2023.114150 (2024).
18. Barbado, R., Araque, O. & Iglesias, C. A. A framework for fake review detection in online consumer electronics retailers. *Inf. Process. Manag.* **56**, 1234–1244 (2019).
19. Manaskasemsak, B., Tantisuwankul, J. & Rungsawang, A. Fake review and reviewer detection through behavioral graph partitioning integrating deep neural network. *Neural Comput. Appl.*, 1–14 (2021).
20. Pv, S. & Bhanu, S. M. S. UbCadet: detection of compromised accounts in twitter based on user behavioural profiling. *Multimedia Tools Appl.* **79**, 19349–19385 (2020).
21. Li, Y. Z. & Li, S. T. Do reviewers' words and behaviors help detect fake online reviews and spammers? Evidence from a hierarchical model. *IEEE Access.* **10**, 42181–42197 (2022).
22. Wang, X., Liu, K., He, S. & Zhao, J. in *Proceedings of the conference on empirical methods in natural language processing.* 866–875. (2016).
23. Holla, L. & KS, K. in *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India.*
24. Violante, M. G. & Vezzetti, E. Kano qualitative vs quantitative approaches: an assessment framework for products attributes analysis. *Comput. Ind.* **86**, 15–25 (2017).
25. Engelbrektsson, P. Effects of product experience and product representations in focus group interviews. *J. Eng. Des.* **13**, 215–221 (2002).
26. AlMousa, M., Al-Khalifa, H. S. & AlSobayel, H. Requirements elicitation and prototyping of a fully immersive virtual reality gaming system for upper limb stroke rehabilitation in Saudi Arabia. *Mob. Inform. Syst.* **2017** (2017).
27. Sakhnini, V., Mich, L. & Berry, D. M. Group versus individual use of power-only EPMcreate as a creativity enhancement technique for requirements elicitation. *Empir. Softw. Eng.* **22**, 2001–2049 (2017).
28. Fu, X., Liu, W., Xu, Y. & Cui, L. Combine HowNet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis. *Neurocomputing* **241**, 18–27. https://doi.org/10.1016/j.neucom.2017.01.079 (2017).
29. Zhang, Y., Zhang, Z., Miao, D. & Wang, J. Three-way enhanced convolutional neural networks for sentence-level sentiment classification. *Inf. Sci.* **477**, 55–64. https://doi.org/10.1016/j.ins.2018.10.030 (2019).
30. Phan, H. T., Pham, D. T. & Nguyen, N. T. FeDN2: fuzzy-enhanced deep neural networks for improvement of sentence-level sentiment analysis. *Cybernetics Syst.*, 1–17. https://doi.org/10.1080/01969722.2023.2296252
31. Su, J. et al. Sentence-level sentiment analysis based on supervised gradual machine learning. *Sci. Rep.* **13**, 14500 (2023).
32. Liu, Y., Shi, J., Huang, F., Hou, J. & Zhang, C. Unveiling consumer preferences in automotive reviews through aspect-based opinion generation. *J. Retailing Consumer Serv.* **77**, 103605. https://doi.org/10.1016/j.jretconser.2023.103605 (2024).
33. Zhang, W., Li, X., Deng, Y., Bing, L. & Lam, W. A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *IEEE Trans. Knowl. Data Eng.* (2022).
34. Zhang, J., Zhang, A., Liu, D. & Bian, Y. Customer preferences extraction for air purifiers based on fine-grained sentiment analysis of online reviews. *Knowl. Based Syst.* **228**, 107259. https://doi.org/10.1016/j.knosys.2021.107259 (2021).
35. Cambria, E., Poria, S., Hazarika, D. & Kwok, K. in *Proceedings of the AAAI conference on artificial intelligence.*
36. Du, K., Xing, F. & Cambria, E. Incorporating multiple knowledge sources for targeted aspect-based financial sentiment analysis. *ACM Trans. Manage. Inform. Syst.* (2023).
37. Rajput, L. & Gupta, S. Sentiment analysis using latent Dirichlet allocation for aspect term extraction. *J. Computers Mech. Manage.* **1**, 30–35 (2022).
38. Bian, Y., Ye, R., Zhang, J. & Yan, X. Customer preference identification from hotel online reviews: a neural network based fine-grained sentiment analysis. *Comput. Ind. Eng.* **172**, 108648. https://doi.org/10.1016/j.cie.2022.108648 (2022).
39. Sindhu, C. & Vadivu, G. Fine grained sentiment polarity classification using augmented knowledge sequence-attention mechanism. *Microprocess. Microsyst.* **81**, 103365 (2021).
40. Talaat, A. S. Sentiment analysis classification system using hybrid BERT models. *J. Big Data*. **10**, 1–18 (2023).
41. Xu, L., Pang, X., Wu, J., Cai, M. & Peng, J. Learn from structural scope: improving aspect-level sentiment analysis with hybrid graph convolutional networks. *Neurocomputing* **518**, 373–383 (2023).
42. Pezzulo, G., Parr, T., Cisek, P., Clark, A. & Friston, K. Generating meaning: active inference and the scope and limits of passive AI. *Trends Cogn. Sci.* https://doi.org/10.1016/j.tics.2023.10.002 (2023).
43. Kouzis-Loukas, D. *Learning Scrapy* (Packt Publishing Livery Place, 2016).
44. Devlin, J., Chang, M. W., Lee, K., Toutanova, K. & Bert Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
45. He, P., Liu, X., Gao, J. & Chen, W. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:03654* (2020). (2006).
46. He, P., Gao, J. & Chen, W. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *arXiv preprint arXiv:2111.09543* (2021).
47. Clark, K., Luong, M. T., Le, Q. V., Manning, C. D. & Electra Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:10555* (2020). (2003).
48. Wang, Z., Xie, Q. & Xia, R. in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval.* 1765–1770.
49. Gururangan, S. et al. Don't stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:10964* (2020). (2004).
50. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. Ser. B: Stat. Methodol.* **58**, 267–288 (1996).
51. Wang, S., Nan, B., Rosset, S. & Zhu, J. Random Lasso. *Annals Appl. Stat.* **5**, 468 (2011).
52. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (CRC, 1994).
53. Paszke, A. et al. Pytorch: an imperative style, high-performance deep learning library. *Adv. Neural. Inf. Process. Syst.* **32** (2019).
54. Kingma, D. P., Ba, J. & Adam A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
55. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
56. Aburass, S., Dorgham, O. & Al Shaqsi, J. A hybrid machine learning model for classifying gene mutations in cancer using LSTM, BiLSTM, CNN, GRU, and GloVe. *Syst. Soft Comput.* **6**, 200110 (2024).

Acknowledgements
This research was supported by the Natural Science Foundation of Shandong Province (ZR2024QG216), the Humanities and Social Science Fund of the Ministry of Education of China (24YJCZH260), the Social Science Planning Project of Shandong Province (23BLYJ04) and the Natural Science Foundation of Shandong Province (ZR2024QE211).We thank the editor and reviewers for their hard work and valuable feedback.

Author contributions
Zhaojing Su, Mei Yang, and Qingbo Zhai conceived the study. Zhaojing Su and Kaiyuan Guo performed the experiments. Yuexin Huang and Yangfan Cong analyzed the experimental results. Zhaojing Su and Mei Yang

wrote the main manuscript text. Qingbo Zhai proofread the manuscript. All authors reviewed and approved the final manuscript and declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.Y. or Q.Z.

Reprints and permissions information is available at www.nature.com/reprints.