



# Design Knowledge as Attention Emphasizer in Large Language Model-Based Sentiment Analysis

Yi Han

Department of Mechanical and Industrial Engineering,  
Northeastern University,  
Boston, MA 02115  
e-mail: han.yi1@northeastern.edu

Mohsen Moghaddam<sup>1</sup>

Department of Mechanical and Industrial Engineering,  
Northeastern University,  
Boston, MA 02115;  
Khoury College of Computer Sciences,  
Northeastern University,  
Boston, MA 02115  
e-mail: m.moghaddam@northeastern.edu

*Aspect-based sentiment analysis (ABSA) enables a systematic identification of user opinions on particular aspects, thus improving the idea creation process in the initial stages of a product/service design. Large language models (LLMs) such as T5 and GPT have proven powerful in ABSA tasks due to their inherent attention mechanism. However, some key limitations remain. First, existing research mainly focuses on relatively simpler ABSA tasks such as aspect-based sentiment analysis, while the task of extracting aspects, opinions, and sentiment in a unified model remains largely unaddressed. Second, current ABSA tasks overlook implicit opinions and sentiments. Third, most attention-based LLMs use position encoding in a linear projected manner or through split-position relations in word distance schemes, which could lead to relation biases during the training process. This paper incorporates domain knowledge into LLMs by introducing a new position encoding strategy for the transformer model. This paper addresses these gaps by (1) introducing the ACOSI (aspect, category, opinion, sentiment, implicit indicator) analysis task, developing a unified model capable of extracting all five types of labels in the ACOSI analysis task simultaneously in a generative manner; (2) designing a new position encoding method in the attention-based model; and (3) introducing a new benchmark based on ROUGE score that incorporates design domain knowledge inside. The numerical experiments on manually labeled data from three major e-Commerce retail stores for apparel and footwear products showcase the domain knowledge inserted transformer method's performance, scalability, and potential. [DOI: 10.1115/1.4067212]*

*Keywords: position encoding, design knowledge, multi-instance inference, large language model, transformers*

## 1 Introduction

Extracting user opinions and completing implicit knowledge from online product reviews have become increasingly vital for successful innovative product design. Utilizing design domain expertise synergistically with large language models (LLMs) to augment the design process and facilitate implicit knowledge completion has emerged as a prevalent prospective methodology in recent years [1–6]. With the exponential growth of online purchasing platforms, a vast amount of user-generated information has accumulated on customer experiences with various products and services. According to recent market surveys, reviews significantly impact customer purchase decisions. Aspect-based sentiment analysis (ABSA) and fine-grained specified LLMs have emerged as crucial facilitators in finding widespread user opinions through e-Commerce and social media platforms. However, due to the noisy and unstructured nature of review data and the lack of efficient methods to insert domain knowledge into LLMs for specified models, extracting valuable information and completing implicit knowledge are often hindered by the limitations of state-of-the-art natural language processing (NLP) methods.

Moreover, there is a large deficiency of automated methods for the completion of implicit knowledge on a large scale from reviews, which can provide valuable information for product designers and improve the probability of successful product development [7]. Effortless completion of implicit knowledge is expected to improve both the quantity and the quality of ideas in the design generation process [8,9]. To address this, we propose a novel approach to implicit knowledge completion. Our dataset, named ACOSI, consists of five labels: aspect (A), category (C), opinion (O), sentiment (S), and implicit indicator (I). The latter represents implicit sentiment-related information from various perspectives, crucial for comprehensive implicit knowledge completion. The encoding algorithm is a novel methodology based on transformer models [10,11], that is, the design-knowledge-guided position design algorithm (DKG), designed to enhance implicit knowledge completion. The validity of the methodology is examined using a sizable dataset collected from notable e-Commerce platforms in the apparel and footwear industry.<sup>2</sup> This section outlines the rationale behind the research, the selection of the dataset and methodology, as well as the goals and contributions of the article in the context of implicit knowledge completion for product design.

The significance of implicit knowledge completion in product design is immense. Unlike explicit knowledge that customers can

<sup>1</sup>Corresponding author.

Manuscript received April 4, 2024; final manuscript received November 5, 2024; published online December 13, 2024. Assoc. Editor: Jianxi Luo.

<sup>2</sup>[www.finishline.com](http://www.finishline.com), [www.newbalance.com](http://www.newbalance.com), and [www.asics.com](http://www.asics.com)

easily articulate, implicit knowledge represents unexpressed desires or problems that customers may not even be aware of [12]. This hidden knowledge often leads to breakthrough innovations and can provide a competitive advantage in the market [13]. By completing implicit knowledge, designers can create products that not only meet current expectations but also anticipate future demands, potentially revolutionizing entire product categories [14]. This proactive approach to design aligns with the concept of “design-driven innovation” proposed by Verganti [15], where companies lead users rather than simply following their expressed needs. Furthermore, the completion of implicit knowledge can result in higher customer satisfaction and loyalty, as users often experience a sense of delight when a product solves problems they had not previously recognized [16]. In the context of our research, the extraction of implicit knowledge from online reviews represents a critical step toward systematically completing this knowledge, thus improving the potential for truly innovative and user-centered product design [17]. Despite extensive research, a universally accepted definition of implicit knowledge remains elusive. Many researchers approach the concept by focusing on edge users and edge use cases [18–22]. Another perspective defines implicit knowledge as unexpected delighters—features or attributes that, once revealed, significantly enhance user satisfaction [23]. While this latter definition captures the essence of the impact of implicit knowledge, it also makes the concept more challenging to extrapolate and apply consistently across different contexts.

**1.1 Knowledge Gaps.** Exploration of user needs is a preliminary step in early-stage new product development processes [24]. Existing need-finding approaches could be divided into two categories, empirical study and data-driven study, and the attention-based model is mostly utilized in recent data-driven research [25]:

- *Gaps in empirical studies.* The main basis for these approaches is the analysis of previous designs [26,27], surveys, and focus group studies [28], and Web-based configurators [29,30]. However, these methods have inherent biases because they target only specific portions of the user population and product instances and are limited to structured inquiries. The lack of direct methods for customers to articulate their requirements [31] and the influence of prior knowledge [32] exacerbate this shortcoming. These restrictions have hindered the widespread adoption of mass customization approaches in industry, given the considerable economic and operational gaps involved [33].
- *Gaps in data-driven studies.* In this category, some researchers have attempted to integrate the information from the image and text to evaluate the generative design in the design concept generation process [34–37]. When using the pure text-based database, sentiment analysis has become a key enabler for the “large-scale” needs to find and allow the extraction of opinions from myriad users of e-Commerce and social media platforms [38]. Sentiment analysis is the process of identifying the subjective opinion of an opinion holder (e.g., user) for a target (e.g., product attribute) from an unstructured text (e.g., product review) [24,27,39–51]. Among the three levels of sentiment analysis (document level, sentence level, and word level sentiment analysis), ABSA could provide the most fine-grained information from the raw text, namely, aspect, opinion, and sentiment. With the increasing demand for unified analysis, the extraction of triplets of opinion sentiments on aspects attracts much attention from the community [38,39,42,45, 51–54]. Some researchers are expanding the task with a new label “category” that makes the task become a quadruple extraction problem, ACOS (aspect, category, opinion term, sentiment) [55–60]. However, the ABSA and ACOS quadruples cannot elicit implicit opinions and aspects. Among the proposed methods, the implicit opinion has been ignored or simply denoted as a “Null” label. Even the ACOS task is only capable of predicting the four labels. When the review

does not mention explicit aspects or opinions, the model will output “Null.” This peculiarity prevents the model from extracting information that implies or describes the aspect indirectly.

- *Gaps in position encoding in the attention-based language model.* In pre-trained language models such as T5 (Text-To-Text Transfer Transformer) [61] and Bidirectional Encoder Representations from Transformers (BERT) [62], to integrate position information into the transformer, position encoding is processed along with input. There are two directions of position encoding in a previous study, fixed position encoding and relative position encoding. In terms of fixed position encoding, BERT uses the same method as in the original transformer model, which is the cosine projection of the fixed position. T5 chooses the relative position encoding strategy in their model, they choose the prefixed length 128 as a relation range, and assume that the distance of words further than this range is not related. However, both position encodings only take the word position as the index, and in domain-specific problems like the needs identification problem, some words are more important than others, even the word appears in the edge area. Researcher tried various methods to reduce position bias [63–66] in sequence-to-sequence tasks, especially specific position-sensitive tasks such as machine translation. Yet, the position bias was treated as the redundant part and not being used as a part to carry information.
- *Gaps in incorporating domain knowledge in LLMs.* LLMs have shown power in several NLP tasks [11,62,67–72]. The success of ChatGPT [73] invokes more resources and expectations for LLMs’ ability. However, even with a great text generation ability, ChatGPT could not answer questions that require domain knowledge well. Moreover, in transformer-based models, there is a lack of systematic way to include domain knowledge in the training process [25], most domain knowledge incorporation is processed by the fine-tuning process or few shot learning process [68,74–78], yet this type of method did not consider domain knowledge in the training process; the model will work purely on the fixed structure without prior knowledge.

**1.2 Objectives.** This research paper aims to address the absence of systematic approaches in the implicit knowledge completion from online reviews using NLP by developing a new position encoding algorithm in standard attention-based models [55,79] and also introduces a new annotated dataset and an efficient position encoding algorithm that facilitate the automated extraction of implicit user needs from online reviews with design domain knowledge integrated with a transformer. Currently, no NLP model is capable of identifying aspects or opinions that have not been explicitly mentioned in a review or those that do not correspond to a specific aspect, category, or sentiment. To overcome this limitation, the paper proposes a new NLP task called ACOSI, which has a structure similar to ACOS [55], but with a crucial distinction. In ACOSI, the text that implies the aspect will be identified and labeled as the indirect opinion when the user does not provide explicit descriptions of the aspect. This new task covers aspects, categories, opinions, sentiments, and implicit opinion extraction, hence the acronym ACOSI.

Despite the similarity of their acronyms, ACOS [55] and ACOSI address completely different problems in terms of their outputs. In ACOS, researchers labeled implicit opinions as “Null,” while in ACOSI, the model can identify and extract opinion text related to aspects, regardless of whether they are explicit or implicit. This crucial difference means that the ACOSI analysis task has the capability to output opinions in association with aspects, whereas the ACOS task does not. Design-knowledge-guided position encoding is generated to achieve the proposed goal. The key contributions of this paper are summarized below.

- A newly created annotated and curated dataset is available to address ACOSI analysis tasks in the product design domain through NLP techniques. This dataset can also be utilized to solve ABSA and ACOS tasks.
- A novel NLP model has been developed and trained on the annotated dataset, with the potential to solve problems related to extracting implicit sentiments, aspects, and opinions.
- A DKG position encoding algorithm has been developed to overcome the context unawareness problem.

The remainder of this article is organized as follows. Section 2 provides a summary of background work related to the main research topics, including ACOSI analysis task and position encoding in large language models. Section 3 discusses the details of the proposed ACOSI analysis task, including the DKG position encoding, the model building process, new loss, the benchmark building process, and the model training process. Section 4 presents the experimental results, analyses, and implications of the developed methodology. Section 6 provides concluding remarks and several directions for future research.

## 2 Background

This section provides an overview of related and background work on ACOSI analysis task, ABSA task in the unified model, and position encoding in the attention-based deep learning model.

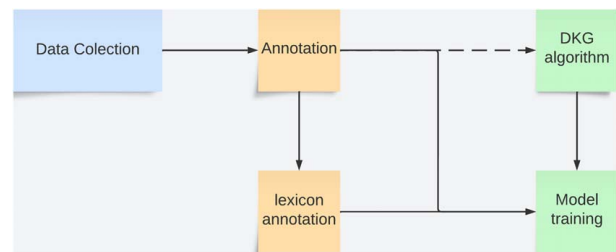
**2.1 ACOSI Analysis Task.** A standard sentiment analysis task involves annotating the user review language into two basic subtasks: aspect extraction and aspect-based sentiment classification. When the two subtasks are integrated, an aspect–sentiment pair of [A–S] is produced where A is the aspect term and S is the sentiment [42,44,45,51]. Expanding on this approach, Qiu et al. [80] produced a triplet of aspect A, sentiment S, and opinion O. With these two approaches, researchers were able to perform A–O pair extraction [81–83] and Aspect, Opinion, Sentiment (AOS) triple extraction [84].

Recently, Cai et al. [55] found that previous research focused solely on extracting explicit aspects and opinion terms from user product reviews, ignoring the fact that 44% of the time these reviews also contained implicit aspects or opinion terms. They introduced a comprehensive task called ACOS that combines explicit and implicit aspects and opinion terms. For example, consider the review: “I like the look and the velvet is great, but the quality of the velvet is not strong.” This review can be broken down into two segments:

“I like the look”: Here, the aspect term is implicit (Null), the category is “Appearance,” the opinion term is “like,” and the sentiment is “Positive.” Thus, the corresponding ACOS label is [Null-Appearance-like-Positive]. “The velvet is great, but the quality of the velvet is not strong”: This segment contains two aspects: (a) “The velvet is great”: The aspect term is “velvet,” the category is “Material,” the opinion term is “great,” and the sentiment is “Positive.” The ACOS label is [velvet-Material-great-Positive]. (b) “The quality of the velvet is not strong”: The aspect term is “velvet,” the category is “Material,” the opinion term is implicit (NULL), and the sentiment is “Negative.” The ACOS label is [velvet-Material-NULL-Negative].

This example demonstrates how the ACOS task can capture both explicit and implicit aspects and opinions within a single review.

While offering advantages, the quadruple extraction task fails to efficiently integrate the four subtasks and capture implicit aspects and opinions. This limits the extraction of useful information, such as the concern raised about the durability of velvet when an opinion is labeled “Null.” To enhance the effectiveness of the ACOS quadruple, we incorporated the implicit “I” tag, resulting in the ACOSI quintet task. It is important to note that the order of these elements (A, C, O, S, I) in the annotation is flexible and can be adapted to specific needs or preferences. In this task, annotators



**Fig. 1 The ACOSI (aspect, category, opinion, sentiment, implicit indicator) analysis task**

mark the span of the opinion text and identify whether it is a direct or indirect opinion. For instance, in the phrase “the velvet quality does not hold up,” we might annotate it as [velvet-Material, quality does not hold up, Negative, Indirect Opinion] or [velvet-Material, Negative, quality does not hold up, Indirect Opinion]. This approach enables us to maintain the information contained within the opinion text span while indicating its implicit nature. The flexibility in the order of ACOSI elements allows for consistency with various annotation formats and model outputs, addressing the reviewer’s concern about apparent discrepancies between sections. The key is that all five elements (A, C, O, S, I) are present in the annotation, regardless of their sequence. The task is shown in Fig. 1.

**2.2 ABSA in a Unified Model.** The ACOSI analysis task can be separated into three classification tasks, namely, category classification, sentiment classification, and implicit indicator classification, as well as two sequence generation tasks, aspect extraction and opinion extraction. Previous research focused on specific subsets of these labels as an expansion of ABSA [56,59,85]. For example, aspect extraction, opinion extraction, sentiment extraction [81], aspect–opinion extraction [84], aspect–sentiment extraction, and opinion–sentiment extraction [86]. However, combining these subtasks is a time-consuming process and does not capture the mutual dependence of sentiments on both opinions and aspects. To achieve a unified approach that accomplishes multiple tasks within a single model, previous studies have been divided into three directions.

**2.3 Large Language Model for Implicit Knowledge Completion.** The use of LLMs for implicit knowledge completion has become an attractive topic in recent years. Researchers have explored incorporating named entity recognition tasks to extract entities and reassemble the results according to user needs [60,87]. One approach involves extracting and summarizing information from user reviews [88]. In sentiment analysis, researchers have used LLMs to analyze user feedback and reviews, extracting sentiment and identifying specific pain points or desired features [89,90]. For requirements elicitation, LLMs have been employed to process and summarize large volumes of user interviews and surveys, helping to identify common themes and requirements [91]. In trend prediction, by analyzing social media posts and online discussions, LLMs have been used to predict emerging user needs and preferences [92]. Regarding persona generation, some studies have explored the use of LLMs to create more accurate and diverse user personas based on aggregated user data [93]. In the natural language processing of user feedback, LLMs have been used to process and categorize open-ended feedback, enabling more efficient analysis of large-scale user studies [94–96]. Additionally, researchers have experimented with LLM-powered chatbots to conduct initial user interviews and gather preliminary insights [97,98]. While LLMs have shown significant promise in these areas, researchers also caution about possible biases and emphasize the need for human oversight in interpreting results [99]. As the technology continues to evolve, it is expected that LLMs will play an increasingly

important role in implicit knowledge completion, complementing traditional research methods.

**2.4 Position Encoding.** In the classic encoder–decoder structured model [25], the position information of the input and output should also be included in the training process. The direct position information is linearly increasing and aligns with the work rank, but the information for each word should be considered as the same instead of linearly increasing; to solve this problem, the original transformer model chooses the trigonometric function to scale down the position to range  $(-1, 1)$ . But the trigonometric transformation still feeds fixed information to the model; the relationship between words has been neglected. To capture the relationship information between words inside input and output, a relative position encoding method was generated [100], and in relative position encoding, the position information was no longer represented by a single vector, instead, it is a square matrix whose dimension was equal to the length of input/output. The relative position encoding shows its advance in the T5 [61] model.

- (1) Developing heuristic models based on syntax rules and algorithms based on lexicons [101].
- (2) Building deep learning models using an extraction approach based on named entity recognition that assigns a tag to each word within the text-based dataset [60,102–104].
- (3) Generating all necessary labels using a predictive format, where the model outputs labels iteratively based on the previous ones. In other words, the model operates sequentially [58,105,106].

The review text can be categorized into explicit and implicit aspects and opinions. However, previous research has not given sufficient attention to the implicit category, with researchers often denoting implicit aspects and opinions as “Null.” Additionally, there has been a lack of detailed analysis regarding the “implicit” category.

### 3 Methodology

This section introduces the dataset along with the proposed model structure and the DKG algorithm.

**3.1 The Dataset.** The unprocessed review dataset was extracted from three websites, namely, Finish Line, ASICS, and New Balance, resulting in a total of 145,430 reviews. Within this dataset, there were 10,700 lengthy reviews that exceeded 60 words, 22,458 concise reviews with less than 10 words, and 1636 reviews that referenced specific product names. To eliminate any innate bias, lengthy and concise reviews were discarded, resulting in a dataset of 59,184 reviews. Within this filtered dataset, 75.89% were rated 5-star, 12.85% 4-star, 4.96% 3-star, 2.64% 2-star, and 4.92% 1-star. In the referenced annotated dataset [107], we utilized basic length heuristics to filter reviews, preserving those with 2–5 sentences, and conducted a uniform sampling based on the star rating of the reviews. The annotation task was performed by native-English-speaking product design students that we hired. Each example was assigned two annotators and any conflicts that arise were resolved by a project lead. We used Cohen’s Kappa [108] as the metric to measure inter-annotator agreement, and we observed a 52.8% chance-adjusted agreement on the categorical components and 40.8% on the span extraction tasks, which include aspect and opinion, between the annotators. These agreement scores reflect the inherent complexity of multi-label annotation in product reviews, where multiple valid interpretations often exist for the same text segment. For instance, when a reviewer comments on a shoe’s “uncomfortable fit,” one annotator might focus on the general comfort category while another might specifically note the sizing aspect, both being equally valid interpretations. Each sample was annotated following the ACOSI format, resulting in a

set of ground-truth quintuples, which include aspect term, aspect category, opinion, sentiment polarity, and an implicit/explicit opinion flag. To ensure annotation quality despite the inherent subjectivity, we implemented a rigorous review process where any significant disagreements were discussed and resolved by the project lead, who had extensive experience in both product design and annotation tasks. Annotators were instructed to annotate implicit opinion spans by labeling the minimal supporting span that covered the implicit sentiment expression for the current tuple. Future research will expand the dataset with additional annotators and implement more detailed annotation guidelines to further improve consistency while maintaining the richness of multi-perspective interpretation.

**3.2 Unified Model Structure and Implicit Knowledge Completion.** Pre-trained LLMs utilize the likelihood of single words or word spans in real text to encode natural language into numerical values. The concept of a language model, originally introduced in the early 1900s with neural networks, has since been integrated into many text analysis tasks. With the advent of deep learning, researchers have developed LLMs that are pre-trained in large corpora. T5 [61] is a pre-trained LLM developed by Google AI Language that uses transformers as a benchmark model structure and trains on massive datasets, specifically the full version of the C4 dataset [109]. These powerful LLMs can be fine-tuned for specific NLP tasks, such as adding a single layer, resulting in state-of-the-art performance in several NLP tasks. In this study, we use T5 as our base LLM and fine-tune it for the ACOSI label prediction task. The identification of implicit knowledge is still in its infancy. Some research defined those implicit knowledge into three categories [23]: unexpected delighted, lead user needs, and extraordinary user needs, while others defined them as needs from edge users/latent users [18,20,21,110,111]. Systematic methods for implicit knowledge completion still lack the ability to identify indirect opinions that the user has in text, do not have large enough databases, leading to an increase in bias, and it is a challenge to obtain or hierarchically categorize product attributes without human involvement.

In this article, inspired by the work of Cai et al. [55], we further expanded the quadruple task. Their work, for example, in the review “I like the look and the velvet is great, but the quality of the velvet is not strong,” the aspect term is “cushion,” the category is “Material,” the opinion term is “great,” and the sentiment is “Positive.” Thus, the corresponding ACOS labels are [velvet-Material-great-Positive]. Additionally, we can extract [Null-Appearance-like-Positive] where the aspect term is implicit in the second part of the review and [velvet-material-NULL-Negative] where the opinion term is implicit in the third part.

While offering advantages, the quadruple extraction task fails to efficiently integrate the four subtasks and capture implicit aspects and opinions. This limits the extraction of useful information, such as the concern raised about the durability of velvet when an opinion is labeled “Null.” To enhance the effectiveness of the ACOS quadruple, we incorporated the implicit “I” tag, which results in the ACOSI quintet task. In this task, the annotators mark the span of the opinion text and identify whether it is a direct or indirect opinion. For instance, in the phrase “the velvet quality does not hold up,” we annotate [velvet-Material quality does not hold up (Indirect Opinion)-Negative]. This approach enables us to maintain the information contained within the opinion text span while indicating its implicit nature. The ACOSI analysis task is completed using a consolidated model that generates output. The T5 model processes all tasks related to text in a sequence-to-sequence fashion, where tasks like sentiment analysis produce outputs such as “positive” and “negative.” Even regression tasks are processed in this way, with predictions in the form of string output such as “five.” In the T5 model, a standard encoder-decoder transformer architecture [25] is used, which consists of three parts; in addition to the structure of the standard transformer

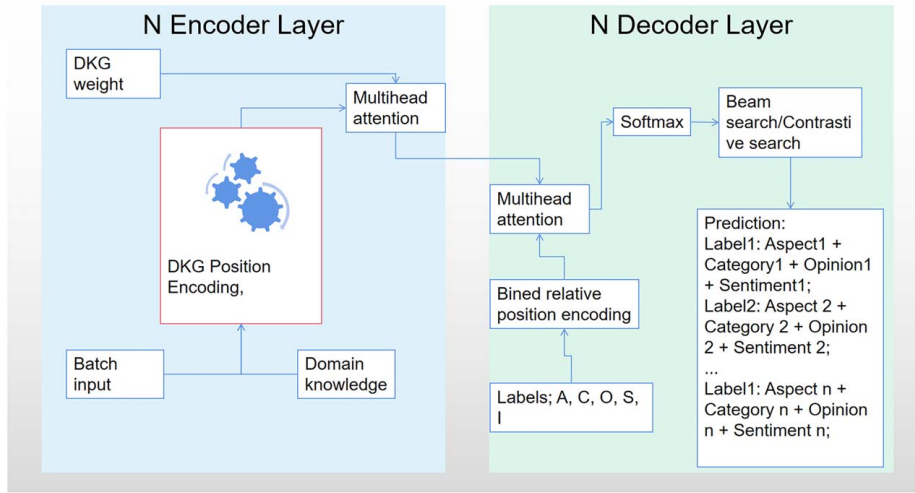


Fig. 2 DKG transformer structure

model, the knowledge-guided position algorithm was tested in contrast to the same experimental settings. Figure 2 provides a visual representation of the entire structure of the model.

- **Tokenization.** To provide text input to the model, it is necessary to encode the entire text into input IDs. In this experiment, the T5 tokenizer is used in the first step, which consists of a vocabulary with 32,128 elements. Given that this is a problem that involves multiple tasks and labels, five special tokens are included in the vocabulary: “{label},” “{A},” “{C},” “{O},” “{S},” “{I}”. These tokens denote the initiation of a label and the different types of labels, respectively.
- **Encoding process.** The input data are encoded using the T5 large model encoder, which comprises a conventional transformer encoder with 12 attention layers, where each layer has a hidden dimension of 768.
- **Knowledge guided encoding.** The position encoding of the input was performed before being fed into the model; the details of the algorithm are introduced in the next paragraph.
- **Decoding process.** Using a generative approach, the encoder output is decoded. The decoder employs a six-layer attention mechanism and generates each prediction through an autoregressive process by which the next prediction is based on the previous output and the current decoder output.

**3.2.1 Design-Knowledge-Guided Position Encoding.** In a standard attention-based LLM, for each self-attention head  $h$ , the model operates on an input sequence of  $n$  elements,  $(x_0, x_1, \dots, x_n)$  that  $x_i \in \mathbb{R}^{d_x}$ , then computes a new sequence of the same length  $(z_0, z_1, \dots, z_n)$  that  $z_i \in \mathbb{R}^{d_z}$ . Here,  $d_x$  indicates the dimension of input  $x$  for each output element  $z_i$ , and it is calculated as the weighted sum of each linear project of input elements, Eqs. (1)–(5), adopted from Shaw’s [100] work:

$$z_i = \sum_{j=0}^n \alpha_{i,j} (x_j W^V) \quad (1)$$

The attention weight coefficient is calculated as

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=0}^n \exp(e_{ik})} \quad (2)$$

each  $e_{ij}$  is computed through the weighted product of two input elements:

$$e_{ij} = \frac{(x_i W^Q)(x_j W^K)^T}{\sqrt{d_z}} \quad (3)$$

These three weight matrices are learned during the training process,  $W^Q, W^K, W^V \in \mathbb{R}^{d_x \times d_z}$ . In relation-aware self-attention models, the position relationship between two input elements  $x_i$  and  $x_j$  is represented by two vectors  $a_{ij}^V$  and  $a_{ij}^K$ . The learned coefficients from Eqs. (1) and (3) could then be used to calculate the following output:

$$z_i = \sum_{j=0}^n \alpha_{ij} (x_j W^V + b_{ij}^V) \quad (4)$$

where

$$e_{ij} = \frac{x_i W^Q (x_j W^K + b_{ij}^K)^T}{\sqrt{d_z}} \quad (5)$$

During the experiment, we found that the improvement in model performance with  $b_{ij}$  inserted did not show pertinent results during the training process; thus, in addition to the domain knowledge carrier of position encoding, we also want to incorporate some linguistic information into the training process. To be more specific, inside each bin,  $b_{ij}$  will be times the weight  $\beta$  that is calculated as (Fig. 3):

$$\beta = \text{cosine similarity}(\text{current bin, attribute lexicon}) \quad (6)$$

Then, Eq. (4) becomes

$$z_i = \sum_{j=0}^n \alpha_{ij} (x_j W^V + \beta * b_{ij}^V) \quad (7)$$

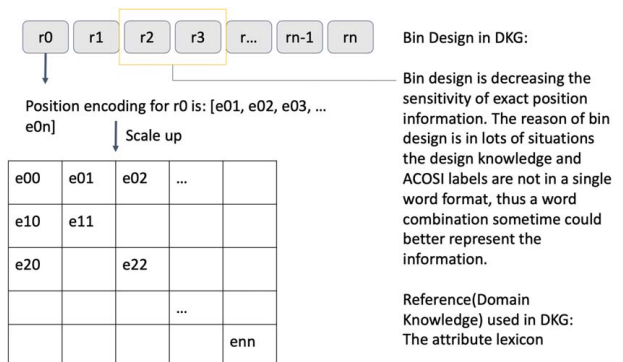


Fig. 3 Bin design in DKG position encoding

When using the relative position representations, previous studies assume that two inputs will be irrelevant beyond a certain distance, e.g., T5 assumes that this distance is 128. Suppose that the maximum distance here is  $k$ , then the unique labels  $2k + 1$  should be considered, then the  $a_{ij}$  in Eqs. (4) and (5) could be calculated as

$$b_{ij}^K = W_{\text{chunk}(j-k,k)}^K \quad \text{and} \quad b_{ij}^V = W_{\text{chunk}(j-k,k)}^V \quad (8)$$

where  $\text{chunk}(x, k) = \max\{-k, \min(k, x)\}$  and the relative position representations  $W^K$  and  $W^V$  could be learned in the training process. On top of the relative position encoding, we implemented the DKG position encoding as shown below.

---

**Algorithm 1** Design knowledge-guided position encoding

---

A pre-defined design knowledge lexicon  $D$

$x_i$  in length  $L$  input

**if**  $\{x_i \text{ or } x_j \in D\}$  **then**

**if**  $x_i \text{ and } x_j \in D$  **then**

**if**  $|j - i| < k$  **then**

$\text{pick}(i, j, k, D) = j - i$

**else**

$\text{pick}(i, j, k, D) = \max(-k, \min(j-i, k))$

**end if**

**else**

**if**  $|j - i| < k$  **then**

$\text{pick}(i, j, k, D) = \min(D \text{ in } D(i))$

**else**

$\text{pick}(i, j, k, D) = k$

**end if**

**end if**

**else**

$\text{pick}(i, j, k, D) = j - i$

**end if**

---

**3.2.2 Loss Function.** The classic transformer model generates the output word by word, left-to-right, and then the weights will be updated in each iteration during the training process. The standard cross-entropy treats each class as they share the same weight 1. In many real-world applications, the label in the dataset is not evenly distributed. To solve the unbalancing problem, the weighted cross-entropy has been introduced, and the customized weight shows a significant improvement in the training process [60]. Typically, the weighted cross-entropy only emphasizes the weight of a specific class or a group of sequential classes; it will be intuitive thoughts to give higher weights to the attribute words inside the attribute lexicon. However, only highlighting the attribute will be incorrect and will affect the ultimate purpose, and only highlighting the attributes inside the lexicon will lure the model to predict words directly from the lexicon, where all words cannot be counted as implicit or latent in this context. To incorporate the lexicon information in the loss function, we choose to weight the entire vocabulary in the tokenizer instead of only a subset of this vocabulary, the massive weighted cross-entropy function, each weight for every single token looks like

$$W_i = \text{Mean}(\text{Cosine Similarity}(\text{token}(i), \text{attribute} : \text{lexicon})) \quad (9)$$

Through massive cross-entropy, the entire vocabulary was weighted with the lexicon, and the loss in each iteration will be directly affected.

**3.2.3 The Design Knowledge Benchmark.** In most classification tasks, accuracy, recall, and F1 score serve as standard benchmarks. However, these three classification benchmarks may not fully address the complexities of the ACOSI analysis task. This inadequacy arises because the primary objective of the ACOSI analysis task is to identify implicit opinions from reviews, often requiring identification of aspects that are similar to, but not exactly

matching, the provided labels. Consequently, standard classification benchmarks, which predominantly rely on exact matches, may not suffice.

To address this challenge, we opt for the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) score, typically used in text summarization tasks, as a benchmark. However, the ROUGE score only assesses the overlap between predictions and the original text, lacking the capacity to gauge the extent to which domain knowledge information has been captured. To remedy this limitation, we propose a modification to the original structure of the ROUGE score, enabling it to incorporate domain knowledge information in predictions (DKG-ROUGE). We utilize the attribute lexicon established in prior work as a reference. Notably, the attribute lexicon comprises only words or terminology specific to sneakers. Consequently, for the DKG-ROUGE, we calculate the single-word ROUGE score [112]. The DKG-ROUGE is calculated as

$$\text{DKG - ROUGE} = \text{ROUGE1}(\text{Prediction}, \text{attributelexicon}) \quad (10)$$

## 4 Experiments and Results

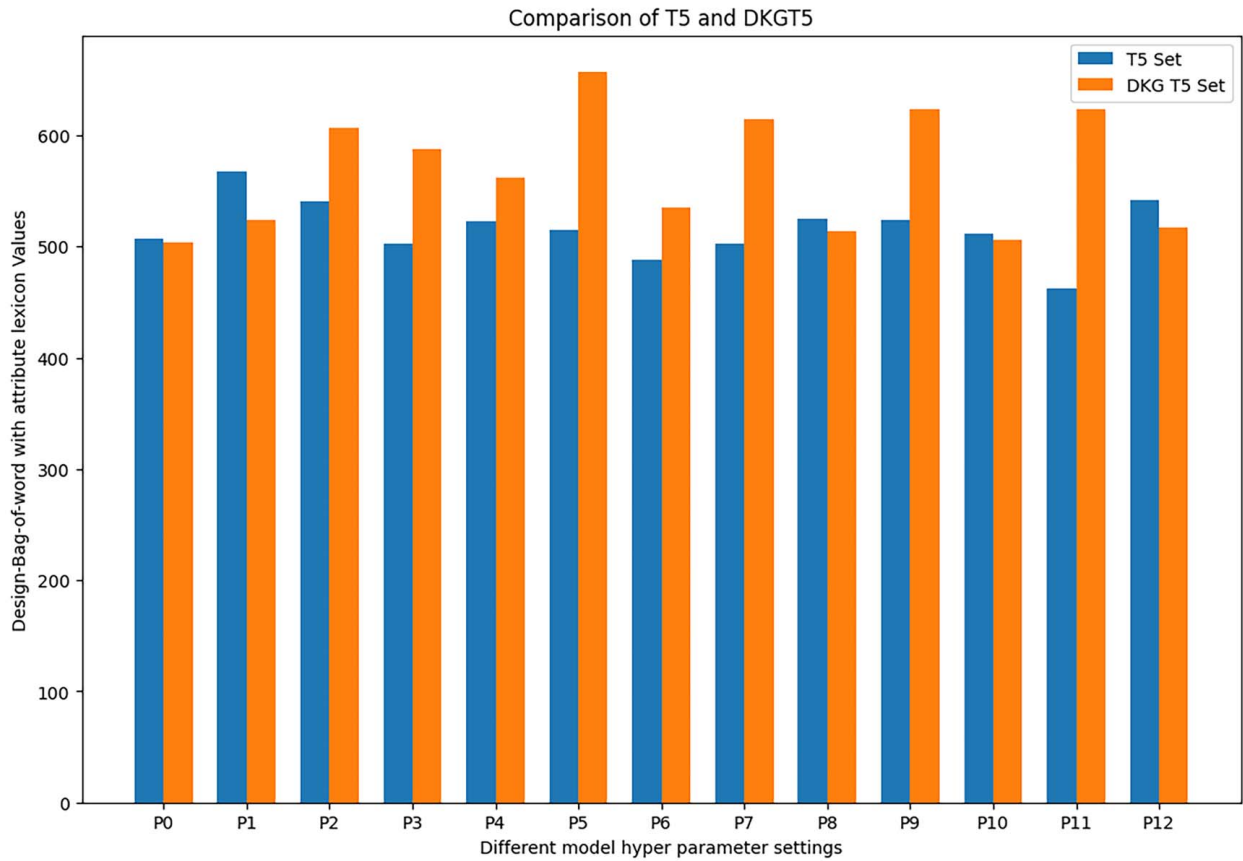
In this work, we choose the T5 model [11] as our baseline, which has the ability to solve all types of NLP tasks in a unified generative manner.

**4.1 Data Preparation and Tokenization.** Regarding the references to attributes, the most frequently mentioned were “Exterior” and “Fit.” Specifically, within all reviews that referenced attributes, 58.25% referred to “exterior,” 76.88% to “fit,” 12.21% to “shoe parts,” 32.11% to “Durability,” 7.33% to “Permeability,” 15.48% to “Stability,” and 16.59% to “Impact absorption.” The attribute data were scrutinized with a specialized sneaker attribute lexicon [85]. After filtering, 2000 reviews were randomly selected for annotation, with 400 reviews from each rating to create a more balanced training dataset.

Since there are five different labels in the same task, we want to introduce some special tokens to indicate the difference among tokens; to be more specific, we introduced six new tokens into the tokenizer we used in the experiment, namely “ $\langle \text{labels} \rangle$ ,” “ $\langle A \rangle$ ,” “ $\langle C \rangle$ ,” “ $\langle O \rangle$ ,” “ $\langle S \rangle$ ,” and “ $\langle I \rangle$ .” The first token “ $\langle \text{labels} \rangle$ ” is a separator that indicates the start of a new set of labels.

**4.2 Model Performance.** In the experiment, we chose the pure T5 model as our baseline comparison model; for fairness of the comparison, we use the exact same parameters settings and benchmarks to measure performance. In each parameter setting, we choose the best model as the final model. The benchmark used includes the following: ROUGE1, ROUGE2, ROUGEL (ROUGE long), and DKG-ROUGE. Since the nuance of the DKG-ROUGE performance is hard to interpret the difference between the baseline model and the DKG model, we provided another metric design-bag-of-words (DBW) technique which uses bag-of-words technique to give a more direct comparison. The DBW used in the experiments is the words that appear in both the test prediction and the attribute lexicon; the results show that in 13 parameter settings, the DKG algorithm robustly has higher chances of predicting design-related words in the test set; the result is shown in Fig. 4; the statistical results of DKG-ROUGE are shown in Figs. 5 and 6. In the following figures and charts, T5 is the based model used in the article and DKG-T5 is the new model we developed (Table 1).

In parallel experiments, if choosing the regular ROUGE score as the benchmark, we find that the T5 universally has around 0.1 higher ROUGE score than the DKG algorithm, in the mean time, on the 1k training dataset, we noticed that the T5 is converging faster than the DKG algorithm, we think that both results are because the T5 has a much larger training dataset, the prior



**Fig. 4 The design-bag-of-words results of T5 and DKG-T5**

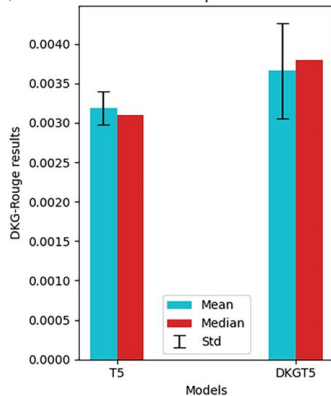
knowledge in T5 makes the model generate predictions faster, and the results will automatically more correct in terms of syntax, which also lead to a higher ROUGE score. Worth to mention, DKG attention forces the multi-head mechanism to pay more attention to a specific domain; it will inevitably make the results get further from the general context; this also explains that the pure T5 model has a better plain ROUGE performance than the DKG model.

In terms of domain knowledge performance, our observations indicate that our model outperforms all best T5 models. Specifically, all best T5 models exhibit a DKG-ROUGE score near 0.003 (rounded to 4 decimal places), while in our experiment, of 13 different parameter settings, 7 sets demonstrate superior domain knowledge performance compared to T5.

Conducting a hypothesis test using a Student's *t*-test with a 95% confidence level, we obtain a *p*-value of 0.01. This result suggests that our DKG model performs better than the pure T5 model in domain knowledge extraction. Furthermore, we present the statistics report of the bag-of-words in Fig. 4. From this report, we can readily conclude that the DKG position encoding algorithm consistently predicts more words related to the domain of sneaker design attributes.

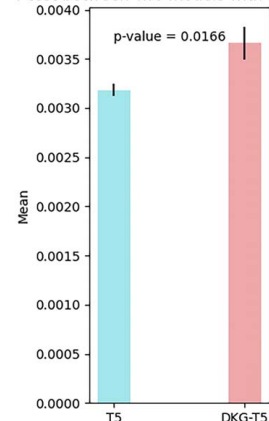
Our findings have significant implications for extracting user knowledge, particularly implicit user knowledge, through the ACOSI analysis task. The superior performance of the DKG model in domain knowledge extraction suggests that it is more skilled at identifying and categorizing specific aspects of user

Mean, Median, and 25 std deviation Comparison between Two models' DKG Rouge



**Fig. 5 The design-bag-of-words statistic results of T5 and DKG-T5**

T-test between Two models with Error Bars



**Fig. 6 The t-test results of SKG-ROUGE between T5 and DKG-T5**

**Table 1 Performance of the DKG model with different parameters compared to the basal T5 model [11]**

Metric	Learning rate	Max. prediction length	Number of beams	Number of repeat ngram	ROUGE1	DKG-ROUGE	DBW
Pure T5	$2 \times 10^{-4}$	64	3	2	0.658	0.0030	507
	$2 \times 10^{-4}$	128	3	2	0.686	0.0036	568
	$2 \times 10^{-4}$	256	3	2	0.673	0.0031	541
	$2 \times 10^{-5}$	128	3	2	0.688	0.0031	502
	$2 \times 10^{-6}$	128	3	2	0.689	0.0034	523
	$2 \times 10^{-4}$	128	4	2	0.686	0.0033	515
	$2 \times 10^{-4}$	128	5	2	0.686	0.0029	488
	$2 \times 10^{-4}$	128	3	3	0.692	0.0031	502
	$2 \times 10^{-4}$	64	4	4	0.687	0.0034	525
	$2 \times 10^{-4}$	64	3	5	0.684	0.0032	524
	$2 \times 10^{-4}$	64	5	4	0.683	0.0030	512
	$2 \times 10^{-5}$	128	3	4	0.695	0.0029	462
	$2 \times 10^{-5}$	128	3	5	0.694	0.0034	542
	DKG transformer	$2 \times 10^{-4}$	64	3	2	0.566	0.0037
$2 \times 10^{-4}$		128	3	2	0.561	0.0033	524
$2 \times 10^{-4}$		256	3	2	0.561	0.0042	607
$2 \times 10^{-5}$		128	3	2	0.557	0.0038	588
$2 \times 10^{-6}$		128	3	2	0.552	0.0039	562
$2 \times 10^{-4}$		128	4	2	0.515	0.0043	657
$2 \times 10^{-4}$		128	5	2	0.519	0.0034	535
$2 \times 10^{-4}$		128	3	3	0.523	0.0044	615
$2 \times 10^{-4}$		64	4	4	0.561	0.0027	514
$2 \times 10^{-4}$		64	3	5	0.550	0.0044	624
$2 \times 10^{-4}$		64	5	4	0.555	0.0030	506
$2 \times 10^{-5}$		128	3	4	0.557	0.0041	623
$2 \times 10^{-5}$		128	3	5	0.539	0.0026	517

feedback that relate to product design and user experience. This aligns with our core aim of developing a more nuanced and context-aware method for completing implicit knowledge from online reviews. The improved ability to capture domain-specific language and concepts is crucial for identifying implicit knowledge, which is often embedded in the subtleties of user expressions and may not be immediately apparent in standard language models.

The ACOSI (aspect, category, opinion, sentiment, and implicit indicator) task benefits from this enhanced domain knowledge in several ways. First, the more accurate identification of aspects and categories ensures that we are capturing the relevant product features and usage contexts. Second, the improved extraction of opinions and sentiments allows for a more nuanced understanding of user experiences. Crucially, the ability to better identify implicit indicators is vital for completing implicit knowledge, as these are often not explicitly stated but inferred from context and domain-specific language. The benchmark is advanced in completing implicit knowledge because it goes beyond simple language understanding to incorporate domain-specific knowledge and context. This allows for a more sophisticated interpretation of user feedback, potentially revealing knowledge that users themselves may not have directly articulated. By combining the strengths of large language models (like T5) with domain-specific guidance (through DKG), we create a system that is both linguistically competent and contextually aware, making it particularly well-suited for the complex task of completing implicit knowledge in specific product domains.

This analysis reinforces our initial research background and core aim to develop a more effective method for completing implicit user knowledge from online reviews in specific product domains. The results demonstrate that our approach offers a significant advance in this direction, providing a more nuanced and context-aware tool for designers and researchers in the field of user-centered product development.

**4.3 Example Model Predictions.** The results of the post-analysis with the model output associated with an example user review are presented in this section.

*Review*

*My ankles tend toward supination, so a well-cushioned heel is crucial. Every other pair of NBs I own give great even support, but the outer heels on this pair collapsed after less than a week ! The shoes now slant outward in a very unsafe and totally unacceptable way that makes my supination way worse. They are completely unusable - huge waste of money. Don't buy this pair unless you have a perfectly even foot strike!*

*Elicited Labels*

- Label 1: (“NULL,” “ContextOfUse#Purchase\_Context,” “neutral,” “ankles tend toward supination, so a well-cushioned heel is crucial,” “direct”)
- Label 2: (“outer,” “Appearance#Shoe Component,” “negative,” “outer,” “direct”)
- Label 3: (“outer,” “ContextOfUse#Usage frequency,” “negative,” “IMPLICIT,” “direct”)
- Label 4: (“shoes,” “Performance#Sizing/Fit,” “negative,” “makes my supination way worse,” “indirect”)
- Label 5: (“NULL,” “Cost/Value,” “negative,” “huge waste of money,” “indirect”)
- Label 6: (“this,” “Performance#Sizing/Fit,” “negative,” “unless you have a perfectly even foot strike,” “indirect”)

*User Needs*

The user has specific needs due to their tendency toward supination (where the foot rolls outward). They require a shoe with a well-cushioned heel that provides even support to counteract their supination tendencies. Labels 4 and 6 indicate that sneaker designers may put more effort for users with special needs; even a normal user could have uneven foot strikes in daily life, a “smart” insole could perhaps help this situation.

*Review*

*I purchased these less than a month ago and the pattern is so cute but . . . It is so worn that it starts to look trashy. I would not recommend any of the tie dye crocs.  
I love the shoe itself, just not how delicate they are with the pattern.*



### Elicited Labels

- Label 1: (“NULL,” “ContextOfUse#Review\_Temporality,” “neutral,” “purchased these less than a month ago,” “direct”)
- Label 2: (“tie dye crocs,” “Appearance#Color,” “negative,” “I would not recommend any,” “direct”)
- Label 3: (“shoe,” “General,” “positive,” “I love the shoe,” “direct”)
- Label 4: (“shoe,” “Appearance#Material,” “negative,” “just not how delicate they are with the pattern,” “direct”)
- Label 5: (“NULL,” “Appearance#Color,” “negative,” “pattern is so cute,” “direct”)

### User Needs

The user shows a pronounced fondness for the ornamental configurations of this particular sneaker. Labels 4 and 5 serve as indicators that sneaker designers should seriously consider durability even in the creation of a seemingly straightforward pattern. The user need in this example is a sneaker that blends esthetic appeal with durability, emphasizing the desire for ornamental patterns that maintain their visual quality over time. They prefer products that offer both style and longevity, suggesting a need for durable or replaceable decorative elements that do not detract from the shoe’s overall quality and appearance. It is important to note that while the above labels are direct outputs from our model performing implicit knowledge completion, the user needs described here represent human designers’ interpretation and analysis of these model-generated labels. The model itself does not directly extract or identify user needs; rather, it provides structured information that human experts can then analyze to identify and articulate specific user needs and design implications.

## 5 Discussion and Limitations

While our proposed ACOSI analysis task demonstrates promising results in analyzing customer reviews for athletic footwear, it is important to acknowledge several limitations that may affect its broader applicability and generalizability.

- *Transferability.* The current study focused exclusively on athletic footwear reviews from three specific websites (Finish Line, ASICS, and New Balance). This narrow product focus raises questions about the transferability of our approach to other product categories, particularly more complex or technically sophisticated items. For instance: (a) *Complex engineering products:* The applicability of our ACOSI analysis task to products like electric vehicles, which involve numerous intricate components and systems, remains untested. Such products may require a more nuanced approach to identify and categorize customer needs accurately. (b) *Consumer electronics:* High-tech consumer products often have rapidly evolving features and capabilities, which may present challenges for our current model in capturing and interpreting user needs effectively. (c) *Service-based products:* Our ACOSI analysis task’s effectiveness in analyzing reviews for intangible products or services, which may have different evaluation criteria compared to physical goods, is yet to be explored.
- *Scalability.* While our dataset of 59,184 reviews provided a substantial basis for analysis, there are potential scalability concerns: (a) *Data volume:* As the volume of reviews increases, particularly for popular products or diverse product lines, the computational resources required for processing and analysis may become a limiting factor. (b) *Language diversity:* Our current model is trained on English-language reviews. Scaling to accommodate multiple languages would require significant additional resources and may introduce new challenges in maintaining consistency across languages. (c) *Real-time analysis:* The ability of our ACOSI analysis task to handle real-time or near-real-time analysis of incoming reviews, which would be valuable for timely product improvements, has not been assessed.
- *Implicit knowledge completion.* Our approach to identifying implicit knowledge, while innovative, has limitations: (a) *Subjectivity:* The identification and categorization of implicit knowledge may still be influenced by subjective interpretations, despite our efforts to maintain objectivity. (b) *Cultural context:* The model’s ability to discern culturally specific implicit knowledge may be limited, potentially missing important nuances in global markets. (c) *Temporal dynamics:* Customer needs and expectations evolve over time. Our current ACOSI analysis task may not adequately capture these temporal dynamics, potentially leading to outdated insights if not regularly updated.
- *Lexicon dependency and evaluation metrics.* Our method’s reliance on domain-specific lexicons for positional encoding presents both strengths and limitations. Although this approach allows for precise, context-aware analysis in the sneaker domain, it raises questions about generalizability to other product categories. The creation of high-quality lexicons for diverse domains may be challenging and resource-intensive, potentially limiting the method’s broad applicability. For instance: (a) *Diverse product categories:* Developing comprehensive lexicons for products with rapidly evolving features (e.g., smartphones) or highly specialized technical components (e.g., industrial machinery) may require frequent updates and extensive domain expertise. (b) *Abstract concepts:* For services or products dealing with intangible qualities, creating an exhaustive and accurate lexicon might be particularly challenging. Additionally, while the ROUGE-based evaluation metric provides a standardized measure of performance, it may not fully capture the nuanced aspects of latent need identification. Future work should explore complementary evaluation methods that can better assess the quality and relevance of extracted implicit knowledge across various domains. This could include human expert evaluations, user studies, or more sophisticated natural language understanding metrics that go beyond surface-level text similarity. It is important to note that while the lexicon used in DKG-ROUGE and DBW metrics was utilized in the DKG algorithm, it was not directly used in the training process of the DKG-T5 model. The gradient descent during the training process did not optimize to align the results with the lexicon. Therefore, while the lexicon played a role in informing the model with some design-related knowledge, it did not inherently bias the model’s prediction performance. This separation ensured that the evaluation of the model remained robust and unbiased. Future research could explore alternative methods for knowledge integration and evaluation to further validate this approach.
- *Label order in auto-regressive models.* One important consideration in our approach is the potential impact of the sequence order in text generation when using auto-regressive models like T5. As pointed out by a reviewer, the order in which we present the ACOSI elements during training and inference (e.g., ACOSI, CSOIA, SOAIC, etc.) may influence the model’s predictions. While our current study did not explicitly investigate this aspect, it is a valuable point for future research. The sequence dependency in auto-regressive models could potentially affect the reliability and consistency of predictions across different element orders. This limitation opens up an interesting avenue for future work, where experiments could be conducted to train and evaluate the model using various ACOSI element sequences. Such studies could provide insights into the robustness of the approach and potentially lead to ensemble methods that leverage multiple sequence orders to enhance prediction reliability. As we continue to refine and expand this research, exploring the impact of element order on model performance will be a priority to ensure the most accurate and dependable results in aspect-based sentiment analysis tasks.

Addressing these limitations will be crucial to enhance the robustness and wide-scale applicability of our proposed ACOSI analysis task. Future research should focus on validating the approach across diverse product categories, improving scalability for larger and more diverse datasets, and refining the identification of implicit knowledge to consider cultural and temporal factors.

## 6 Conclusions and Future Research Directions

This article introduces a ground breaking dataset designed to address a novel natural language processing task known as ACOSI (Comprehensive Procurement of Implied Requirements from Virtual Critiques). It proposes a pioneering approach that combines a consolidated T5 paradigm with a DKG positional coding algorithm to automate the generation of implied views and facets on a wide scale. By leveraging advanced research in natural language processing and linguistic models, this paradigm aims to significantly reduce the time and effort required for data preparation. Moreover, it seeks to decrease reliance on manually crafted expert systems for extracting aspect-opinion-sentiment from reviews. The strengths of this paradigm for ACOSI extraction on a large scale include:

- *Ability to incorporate domain knowledge.* In current transformer-based language models, the integration of domain knowledge is usually done by fine-tuning with a smaller domain dataset. Our work provided a novel, effective, and efficient method; through only changing the position encoding, we emphasize the prior knowledge in the network.
- *Transformation.* The method of operation culminates in the extraction of a comprehensive register of prospective aspects along with their associated opinions. This paradigm represents a noteworthy advance in facilitating automatic and wide-ranging elicitation of implicit aspect opinion, thereby transcending user-centric approaches and potentially uncovering more informative and revolutionary revelations to underpin the design process.
- *Scalability.* Leveraging pre-trained linguistic models, such as T5, alleviates the requirement for extensive manually annotated data. The entire spectrum of methodological constituents is systematized in a streamlined manner and is amenable to rapid adaptation and application to novel datasets.

The emergence of multi-agent LLMs has garnered significant attention in recent years [113,114]. Unlike traditional single input/output format systems [115], multi-agent systems enable models to collaborate or compete with each other within a more intricate environment. This setup closely mirrors real-world systems, offering the potential to simulate human thinking processes more accurately. In many complex tasks, such as the ultimate goal proposed in this paper, addressing implicit knowledge completion proves challenging within a single model setup. Implicit knowledge completion pertain to extracting requirements not explicitly articulated in review texts, a task even challenging for human annotators. Consequently, benchmarking the model's output evaluation becomes exceedingly difficult.

However, by simulating the identification of implicit knowledge completion akin to human thinking or brainstorming processes, each sub-step of the model can be validated more accurately. For instance, breaking down implicit knowledge completion into basic functionalities, emotional support, and fashion allure, and further subdividing each into more granular needs, facilitates a finer evaluation. Through such breakdowns of the ultimate goal, multi-agent systems can leverage diverse types of prior knowledge and agents with distinct "personalities." This diversity enhances the validity and robustness of the system.

*Potential implementation:* The completion of implicit knowledge from online reviews has the potential to significantly enhance the process of user needs finding in the field of human-centered design. Our approach can serve as a valuable tool for designers

and researchers, uncovering insights not immediately apparent or explicitly stated in user feedback, and revealing deeper user needs and desires. This paves the way for more innovative and user-centered product designs. Furthermore, the ability of our approach to accurately identify aspects, categories, opinions, sentiments, and implicit indicators in user feedback provides a comprehensive and nuanced understanding of user experiences. This can aid designers in understanding the context and subtleties of user needs, allowing them to create products that meet current expectations and anticipate future demands. It is important to note that while our approach aids in the completion of implicit knowledge, it should not replace direct user engagement in the design process, but rather serve as a complementary tool enhancing traditional user research methods.

In conclusion, our ACOSI analysis task and the associated T5 model with the DKG positional coding algorithm constitute a significant advancement in the field of natural language processing, specifically for the automated extraction of implicit aspect-opinion-sentiment from reviews. The proposed approach has demonstrated promising results in the athletic footwear domain, and it is expected to make substantial contributions to other domains as well, given its scalability and adaptability. Future research will focus on addressing the limitations identified in this study and further refining the ACOSI analysis task and the associated models for wider and more effective applications.

## Acknowledgment

This material is based on work supported by the National Science Foundation (NSF) under the Engineering Design and System Engineering (EDSE) Grant No. 2050052. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

## Conflict of Interest

There are no conflicts of interest.

## Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

## References

- [1] Emelin, D., Bonadiman, D., Alqahtani, S., Zhang, Y., and Mansour, S., 2022, "Injecting Domain Knowledge in Language Models for Task-Oriented Dialogue Systems," EMNLP, Abu Dhabi, United Arab Emirates.
- [2] Liu, K., Jiang, J., and Lyu, F., 2022, "A Domain Knowledge Enhanced Pre-Trained Language Model for Vertical Search: Case Study on Medicinal Products International Conference on Computational Linguistics," ACL Anthology, Gyeongju, South Korea, pp. 1014–1023.
- [3] Jia, Q., Liu, Y., Wu, D., Xu, S., Liu, H., Fu, J., Vollgraf, R., and Wang, B., 2023, "KG-FLIP: Knowledge-Guided Fashion-Domain Language-Image Pre-training for E-Commerce," ACL Anthology, Toronto, Canada, pp. 81–88.
- [4] Yu, J., Zha, Z.-J., Wang, M., Wang, K., and Chua, T.-S., 2011, "Domain-Assisted Product Aspect Hierarchy Generation: Towards Hierarchical Organization of Unstructured Consumer Reviews," ACL Anthology, Portland, OR, pp. 140–150.
- [5] Xu, S., Li, H., Yuan, P., Wang, Y., Wu, Y., He, X., Liu, Y., and Zhou, B., 2021, "K-PLUG: Knowledge-Injected Pre-trained Language Model for Natural Language Understanding and Generation in E-Commerce," ACL Anthology, Bangkok, Thailand, pp. 1–17.
- [6] Zhang, S., Jia, W., Xia, Y., Meng, Y., and Yu, H., 2010, "Extracting Product Features and Sentiments from Chinese Customer Reviews," ACL Anthology, Uppsala University, Sweden.
- [7] Cooper, R. G., Edgett, S. J., and Kleinschmidt, E. J., 2004, "Benchmarking Best NPJ Practices—III," Research-Technology Management.
- [8] Marion, T. J., and Fixson, S. K., 2018, *The Innovation Navigator: Transforming Your Organization in the Era of Digital Design and Collaborative Culture*, University of Toronto Press.
- [9] Füller, J., and Matzler, K., 2007, "Virtual Product Experience and Customer Participation—A Chance for Customer-Centred, Really New Products," *Technovation*, 27(6), pp. 378–387.

- [10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I., 2017, "Attention is All You Need," *NeurIPS*, Long Beach, CA.
- [11] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J., 2019, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *J. Mach. Learn. Res.*
- [12] Narver, J. C., Slater, S. F., and MacLachlan, D. L., 2004, "Responsive and Proactive Market Orientation and New-Product Success," *J. Prod. Innov. Manage.*, **21**(5), pp. 334–347.
- [13] Slater, S. F., and Narver, J. C., 1995, "Market Orientation and the Learning Organization," *J. Market.*, **59**(3), pp. 63–74.
- [14] Lilien, G. L., Morrison, P. D., Searls, K., Sonnack, M., and von Hippel, E., 2002, "Performance Assessment of the Lead User Idea-Generation Process for New Product Development," *Manage. Sci.*, **48**(8), pp. 1042–1059.
- [15] Verganti, R., 2009, *Design Driven Innovation: Changing the Rules of Competition by Radically Innovating What Things Mean*, Harvard Business Press.
- [16] Kano, N., Seraku, N., Takahashi, F., and Tsuji, S., 1984, "Attractive Quality and Must-Be Quality," *J. Jap. Soc. Qual. Control.*, **14**(2), pp. 147–156.
- [17] Ulwick, A. W., 2005, *What Customers Want: Using Outcome-Driven Innovation to Create Breakthrough Products and Services*, McGraw-Hill, New York.
- [18] Timoshenko, A., and Hauser, J. R., 2019, "Identifying Customer Needs From User-Generated Content," *Market. Sci.*, **38**(1), pp. 1–20.
- [19] Armstrong, J., Stone, R., and Cox, L., 2015, "Customer Needs Extraction Using Disability Simulation for Purposes of Inclusive Design," International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE), Boston, MA, ASME Digital Collection
- [20] Raviselvam, S., Anderson, D., Hölltä-Otto, K., and Wood, K. L., 2018, "Systematic Framework to Apply Extraordinary User Perspective to Capture Latent Needs Among Ordinary Users," IDETC-CIE, Quebec City, ASME Digital Collection
- [21] Raviselvam, S., Subburaj, K., Wood, K. L., and Hölltä-Otto, K., 2019, "An Extreme User Approach to Identify Latent Needs: Adaptation and Application in Medical Device Design," IDETC-CIE, Anaheim, CA, ASME Digital Collection
- [22] Pui-yuk King, A., 2019, "Participatory Design with Older Adults: Exploring the Latent Needs of Young-Old and Middle-Old in Daily Living Using a Universal Design Approach," International Conference on Applied Human Factors and Ergonomics, Washington, DC, July 24–28.
- [23] Zhou, F., Jiao, R. J., and Linsey, J. S., 2015, "Latent Customer Needs Elicitation by Use Case Analogical Reasoning From Sentiment Analysis of Online Product Reviews," *ASME J. Mech. Des.*, **137**(7), p. 071401.
- [24] Schaffhausen, C. R., and Kowalewski, T. M., 2015, "Large-Scale Needfinding: Methods of Increasing User-Generated Needs From Large Populations," *ASME J. Mech. Des.*, **137**(7), p. 071403.
- [25] Vaswani, A., Noam, S., Niki, P., Jakob, U., Llion, J., Aidan N., G., Lukasz, K., and Illia, P., 2017, "Attention is all you need," *Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA.
- [26] Stacey, M., and Sketch, J., 1997, "Managing Effective Communication in Knitwear Design," Technical Report 3.
- [27] Rasoulifar, G., Eckert, C., and Prudhomme, G., 2015, "Communicating Consumer Needs in the Design Process of Branded Products," *ASME J. Mech. Des.*, **137**(7), p. 071404.
- [28] Fogliatto, F. S., and da Silveira, G. J. C., 2008, "Mass Customization: A Method for Market Segmentation and Choice Menu Design," *Int. J. Prod. Econ.*, **111**(2), pp. 606–622.
- [29] Felfernig, A., 2007, "Standardized Configuration Knowledge Representations as Technological Foundation for Mass Customization," *IEEE Trans. Eng. Manage.*, **54**(1), pp. 41–56.
- [30] Franke, N., Schreier, M., and Kaiser, U., 2010, "The 'I Designed It Myself' Effect in Mass Customization," *Manage. Sci.*, **56**(1), pp. 125–140.
- [31] Franke, N., Keinz, P., and Steger, C. J., 2009, "Testing the Value of Customization: When Do Customers Really Prefer Products Tailored to Their Preferences?," *J. Market.*, **73**(5), pp. 103–121.
- [32] Lord, C. G., Ross, L., and Lepper, M. R., 1979, "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence," *J. Pers. Soc. Psychol.*, **37**(11), pp. 2098–2109.
- [33] Fogliatto, F. S., Da Silveira, G. J. C., and Borenstein, D., 2012, "The Mass Customization Decade: An Updated Review of the Literature," *Int. J. Prod. Econ.*, **138**(1), pp. 14–25.
- [34] Yuan, C., Marion, T., and Moghaddam, M., 2022, "Leveraging End-User Data for Enhanced Design Concept Evaluation: A Multimodal Deep Regression Model," *ASME J. Mech. Des.*, **144**(2), p. 021403.
- [35] Yuan, C., and Moghaddam, M., 2020, "Attribute-Aware Generative Design With Generative Adversarial Networks," *IEEE Access*, **8**(1), p. 190710.
- [36] Yuan, C., Marion, T., and Moghaddam, M., 2023, "DDE-GAN: Integrating a Data-Driven Design Evaluator Into Generative Adversarial Networks for Desirable and Diverse Concept Generation," *ASME J. Mech. Des.*, **145**(4), p. 041407.
- [37] Ghasemi, P., Yuan, C., Marion, T., and Moghaddam, M., 2023, "Are Generative Adversarial Networks Capable of Generating Novel and Diverse Design Concepts? An Experimental Analysis of Performance," ICED, University of Bordeaux, France.
- [38] Ravi, K., and Ravi, V., 2015, "A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications," *Knowledge-Based Syst.*, **89**(11), pp. 14–46.
- [39] Tang, H., Tan, S., and Cheng, X., 2009, "A Survey on Sentiment Detection of Reviews," *Expert Syst. Appl.*, **36**(7), pp. 10760–10773.
- [40] Hoyle, C., Chen, W., Wang, N., and Koppelman, F. S., 2010, "Integrated Bayesian Hierarchical Choice Modeling to Capture Heterogeneous Consumer Preferences in Engineering Design," *ASME J. Mech. Des.*, **132**(12), p. 121010.
- [41] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A., 2010, "Sentiment Strength Detection in Short Informal Text," *J. Am. Soc. Inform. Sci. Technol.*, **61**(12), pp. 2544–2558.
- [42] El Dehaibi, N., Goodman, N. D., and MacDonald, E. F., 2019, "Extracting Customer Perceptions of Product Sustainability From Online Reviews," *ASME J. Mech. Des.*, **141**(12), p. 121103.
- [43] He, Y., Camburn, B., Liu, H., Luo, J., Yang, M., and Wood, K., 2019, "Mining and Representing the Concept Space of Existing Ideas for Directed Ideation," *ASME J. Mech. Des.*, **141**(12), p. 121101.
- [44] Bing, L., 2015, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Vol. 38, The Cambridge University Press, pp. 41–51.
- [45] Ireland, R., and Liu, A., 2018, "Application of Data Analytics for Product Design: Sentiment Analysis of Online Product Reviews," *CIRP J. Manuf. Sci. Technol.*, **23**(1), pp. 128–144.
- [46] Decker, R., and Trusov, M., 2010, "Estimating Aggregate Consumer Preferences From Online Product Reviews," *Int. J. Res. Market.*, **27**(4), pp. 293–307.
- [47] Rai, R., 2012, "Identifying Key Product Attributes and Their Importance Levels From Online Customer Reviews," IDETC-CIE, Chicago, IL, Aug. 12–15, Vol. 3, pp. 533–540.
- [48] Stone, T., and Kyum Choi, S., 2013, "Extracting Consumer Preference From User-Generated Content Sources Using Classification," IDETC-CIE, Portland, OR, Aug. 4–7.
- [49] Qi, J., Zhang, Z., Jeon, S., and Zhou, Y., 2016, "Mining Customer Requirements From Online Reviews: A Product Improvement Perspective," *Inf. Manage.*, **53**(8), pp. 951–963.
- [50] Lim, S., and Tucker, C. S., 2017, "Mitigating Online Product Rating Biases Through the Discovery of Optimistic, Pessimistic, and Realistic Reviewers," *ASME J. Mech. Des.*, **139**(11), p. 111409.
- [51] Wang, Y., Mo, D. Y., and Tseng, M. M., 2018, "Mapping Customer Needs to Design Parameters in the Front End of Product Design by Applying Deep Learning," *CIRP Ann.*, **67**(1), pp. 145–148.
- [52] Lee, T. Y., and Bradlow, E. T., 2011, "Automated Marketing Research Using Online Customer Reviews," *J. Market. Res.*, **48**(5), pp. 881–894.
- [53] Pang, B., and Lee, L., 2006, "Opinion Mining and Sentiment Analysis," *Found. Trends Inf. Retrieval*, **1**(2), pp. 91–231.
- [54] Zhang, L., Wang, S., and Liu, B., 2018, "Deep Learning for Sentiment Analysis : {A} Survey," CoRR, abs/1801.0.
- [55] Cai, H., Xia, R., and Yu, J., 2021, "Aspect-Category-Opinion-Sentiment Quadruple Extraction With Implicit Aspects and Opinions," *ACL Anthology, Bangkok, Thailand*, pp. 340–350.
- [56] Mao, Y., Shen, Y., Yu, C., and Cai, L., 2021, "A Joint Training Dual-MRC Framework for Aspect Based Sentiment Analysis," *AAAI Conference on Artificial Intelligence, Vancouver, Canada*.
- [57] Yu, G., Ao, X., Luo, L., Yang, M., Sun, X., Li, J., and He, Q., 2021, "Making Flexible Use of Sub-tasks: A Multiplex Interaction Network for Unified Aspect-based Sentiment Analysis," *ACL-IJCNLP, Bangkok, Thailand*.
- [58] Yan, H., Dai, J., Ji, T., Qiu, X., and Zhang, Z., 2021, "A Unified Generative Framework for Aspect-Based Sentiment Analysis," *ACL IJCNLP, Bangkok, Thailand*.
- [59] Xu, L., Chia, Y. K., and Bing, L., 2021, "Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction," *ACL-IJCNLP, Bangkok, Thailand*.
- [60] Han, Y., and Moghaddam, M., 2021, "Eliciting Attribute-Level User Needs From Online Reviews With Deep Language Models and Information Extraction," *ASME J. Mech. Des.*, **143**(6), p. 061403.
- [61] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J., 2019, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *J. Mach. Learn. Res.*
- [62] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., 2018, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL-HLT, New Orleans, LA*.
- [63] Zhang, S., and Feng, Y., 2022, "Reducing Position Bias in Simultaneous Machine Translation with Length-Aware Framework," *ACL, Dublin, Ireland*.
- [64] Ko, M., Lee, J., Kim, H., Kim, G., and Kang, J., 2020, "Look At the First Sentence: Position Bias in Question Answering," *ACL Anthology, Online*, pp. 1109–1121.
- [65] Yan, H., Gui, L., Pergola, G., and He, Y., 2021, "Position Bias Mitigation: A Knowledge-Aware Graph Model for Emotion Cause Extraction," *ACL Anthology, Bangkok, Thailand*, pp. 3364–3375.
- [66] Zhang, S., Feng, Y., and Li, L., 2021, "Future-Guided Incremental Transformer for Simultaneous Translation," *AAAI*, **35**(16), pp. 14428–14436.
- [67] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V., 2019, "Roberta: A Robustly Optimized BERT Pretraining Approach," *China National Conference on Computational Linguistics, Kunming, China*.
- [68] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., and Neelakantan, A., 2020, "Language Models are Few-Shot Learners," *NeurIPS, British Columbia, Canada*.
- [69] Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., and Zhang, Y., 2021, "TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification," *NeurIPS, Virtual Online*.

- [70] Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Jozefowicz, R., and Bengio, S., 2015, "Generating Sentences from a Continuous Space," CoNLL, Berlin, Germany.
- [71] See, A., Liu, P. J., and Manning, C. D., 2017, "Get To The Point: Summarization with Pointer-Generator Networks," ACL, Vancouver, Canada.
- [72] Bražinskas, A., Lapata, M., and Titov, I., 2019, "Unsupervised Opinion Summarization as Copycat-Review Generation," ACL, online.
- [73] ChatGPT, 2023, Website.
- [74] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., and Bashlykov, N., 2023, Llama 2: Open Foundation and Fine-Tuned Chat Models, arXiv.
- [75] Dai, Z., and Huang, R., 2021, "A Joint Model for Structure-based News Genre Classification with Application to Text Summarization," ACL-IJCNLP, Bangkok, Thailand.
- [76] Sun, E., Hou, Y., Wang, D., Zhang, Y., and Wang, N. X. R., 2021, "D2S: Document-to-Slide Generation Via Query-Based Text Summarization," NAACL, Mexico City, Mexico.
- [77] Liu, Y., Shen, S., and Lapata, M., 2020, "Noisy Self-Knowledge Distillation for Text Summarization," NAACL, Mexico City, Mexico.
- [78] Liu, Y., and Lapata, M., 2019, "Text Summarization with Pretrained Encoders," EMNLP, Hong Kong, China.
- [79] Han, Y., and Moghaddam, M., 2019, "Analysis of Sentiment Expressions for Customer-Centric Design," *Expert Syst. Appl.*, **1**(617), pp. 1–12.
- [80] Qiu, G., Liu, B., Bu, J., and Chen, C., 2011, "Opinion Word Expansion and Target Extraction Through Double Propagation," *Comput. Linguist.*, **37**(1), pp. 9–27.
- [81] Chen, S., Liu, J., Wang, Y., Zhang, W., and Chi, Z., 2020, "Synchronous Double-Channel Recurrent Network for Aspect-Opinion Pair Extraction," ACL Anthology, Online, pp. 6515–6524.
- [82] Nayeem, M. T., Fuad, T. A., and Chali, Y., 2018, "Abstractive Unsupervised Multi-document Summarization Using Paraphrastic Sentence Fusion," ACL, Santa Fe, NM, August, pp. 1191–1204.
- [83] Qiu, Y., and Jin, Y., 2021, "Engineering Document Summarization Using Sentence Representations Generated by Bidirectional Language Model," IDETC-CIE, Online, Aug. 17–20.
- [84] Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., and Si, L., 2020, "Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis," *AAAI*, **34**(5), pp. 8600–8607.
- [85] Han, Y., and Moghaddam, M., 2021, "Analysis of Sentiment Expressions for User-Centered Design," *Expert Syst. Appl.*
- [86] Ahmed, M., Chen, Q., Wang, Y., Nafa, Y., Li, Z., and Duan, T., 2021, "DNN-Driven Gradual Machine Learning for Aspect-Term Sentiment Analysis," ACL Anthology, Bangkok, Thailand, pp. 488–497.
- [87] Han, X., and Eisenstein, J., 2019, "Unsupervised Domain Adaptation of Contextualized Embeddings for Sequence Labeling," EMNLP-IJCNLP, Hong Kong, China, Nov. 3–7, pp. 4237–4247.
- [88] Han, Y., Nanda, G., and Moghaddam, M., 2023, "Attribute-Sentiment-Guided Summarization of User Opinions From Online Reviews," *ASME J. Mech. Des.*, **145**(4), p. 041402.
- [89] Liu, J., 2023, "Leveraging Large Language Models for Enhanced Sentiment Analysis in User Feedback," *J. Exp. Res.*, **15**(3), pp. 245–260.
- [90] Garcia, M., and Lopez, F., 2023, "Combining Traditional and LLM-Based Methods for Comprehensive User Needs Analysis," *J. Prod. Innov. Manage.*, **40**(5), pp. 589–605.
- [91] Zhang, Y., and Chen, H., 2022, "LLM-Assisted Requirements Elicitation: A Comparative Study," *Requir. Eng.*, **27**(4), pp. 501–518.
- [92] Wang, L., 2024, "Predictive Analytics of User Needs Using LLMS on Social Media Data," *IEEE Trans. Affect. Comput.*, **15**(2), pp. 178–191.
- [93] Johnson, A., and Smith, B., 2023, "Generating Dynamic User Personas With GPT-Based Models," *Proceedings of the International Conference on Human-Computer Interaction*, Copenhagen, Denmark, July, pp. 89–103.
- [94] Brown, T., 2022, "Automated Categorization of User Feedback Using Transformer-Based Models," *ACM Trans. Comput.-Hum. Interact.*, **29**(4), pp. 1–28.
- [95] Mueller, E., and Schmidt, A., 2022, "Enhancing Focus Group Analysis With LLM-Based Text Summarization," *Qual. Market. Res.: Int. J.*, **25**(4), pp. 521–537.
- [96] Chen, X., 2023, "Temporal Analysis of Evolving User Needs Using Historical Data and LLMS," *MIS Q.*, **47**(3), pp. 1125–1142.
- [97] Lee, S., and Park, J., 2023, "Chatresearch: Exploring the Potential of LLM-Powered Chatbots in User Research," *Int. J. Hum.-Comput. Stud.*, **169**, p. 102930.
- [98] Patel, K., 2022, "LLMs in Competitive Analysis: Extracting User Needs From Competitor Product Reviews," *Strateg. Manage. J.*, **43**(13), pp. 2456–2474.
- [99] Smith, R., 2024, "Ethical Considerations in AI-Assisted User Needs Identification," *AI & Soc.*, **39**(2), pp. 411–425.
- [100] Shaw, P., Uszkoreit, J., and Vaswani, A., 2018, "Self-Attention with Relative Position Representations," NAACL, New Orleans, LA.
- [101] Han, Y., and Moghaddam, M., 2021, "Analysis of Sentiment Expressions for User-Centered Design," *Expert Syst. Appl.*, **171**(4), p. 114604.
- [102] Wan, H., Yufei, Y., Jianfeng, D., Yanan, L., Kunxun, Q., and Jeff Z., P., 2022, "Target-Aspect-Sentiment Joint Detection for Aspect-Based Sentiment Analysis," *Proceedings of the AAAI Conference on Artificial Intelligence*, New York.
- [103] Xu, L., Li, H., Lu, W., and Bing, L., 2020, "Position-Aware Tagging for Aspect Sentiment Triplet Extraction," EMNLP, Online.
- [104] Han, Y., and Moghaddam, M., 2023, "A Design Knowledge Guided Position Encoding Methodology for Implicit Need Identification From User Reviews," ASME Digital Collection.
- [105] Peper, J. J., and Wang, L., 2022, "EMNLP 2022," EMNLP, Abu Dhabi.
- [106] Han, Y., Bruggeman, R., Peper, J., Chehade, E. C., Marion, T., Ciuccarelli, P., and Moghaddam, M., 2023, "Extracting Latent Needs From Online Reviews Through Deep Learning Based Language Model," ICED, Bordeaux France.
- [107] Peper, J. J., Qiu, W., Bruggeman, R., Han, Y., Chehade, E. C., and Wang, L., 2024, "Shoes-ACOSI: A Dataset for Aspect-Based Sentiment Analysis with Implicit Opinion Extraction," EMNLP, Miami, FL.
- [108] Cohen, W. W., 1996, "Learning Trees and Rules With Set-Valued Features," AAAI, Portland, OR, Aug. 4–8, AAAI Press, pp. 709–716.
- [109] c4 TensorFlow Datasets, Oct. 2022, Public dataset.
- [110] Wang, K., Shen, W., Yang, Y., Quan, X., and Wang, R., 2020, "Relational Graph Attention Network for Aspect-based Sentiment Analysis," ACL, Online.
- [111] Lee, J. S., and Taxman, F. S., 2020, "Using Latent Class Analysis to Identify the Complex Needs of Youth on Probation," *Child. Youth Serv. Rev.*, **115**, p. 105087.
- [112] Lin, C.-Y., 2004, "Rouge: A Package for Automatic Evaluation of Summaries," ACL, Barcelona, Spain.
- [113] Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., and Chen, Z., 2023, "A Survey on Large Language Model based Autonomous Agents," *Front. Comput. Sci.*
- [114] Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., and Zhang, M., 2023, The Rise and Potential of Large Language Model Based Agents: A Survey, arXiv.
- [115] Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., and Narasimhan, K., 2023, "Neural Information Processing Systems (NeurIPS 2023)," NeurIPS, New Orleans, LA.