NATURAL LANGUAGE PROCESSING IN MARKETING

Jochen Hartmann and Oded Netzer

ABSTRACT

The increasing importance and proliferation of text data provide a unique opportunity and novel lens to study human communication across a myriad of business and marketing applications. For example, consumers compare and review products online, individuals interact with their voice assistants to search, shop, and express their needs, investors seek to extract signals from firms' press releases to improve their investment decisions, and firms analyze sales call transcripts to increase customer satisfaction and conversions. However, extracting meaningful information from unstructured text data is a nontrivial task. In this chapter, we review established natural language processing (NLP) methods for traditional tasks (e.g., LDA for topic modeling and lexicons for sentiment analysis and writing style extraction) and provide an outlook into the future of NLP in marketing, covering recent embedding-based approaches, pretrained language models, and transfer learning for novel tasks such as automated text generation and multi-modal representation learning. These emerging approaches allow the field to improve its ability to perform certain tasks that we have been using for more than a decade (e.g., text classification). But more importantly, they unlock entirely new types of tasks that bring about novel research opportunities (e.g., text summarization, and generative question answering). We conclude with a roadmap and research agenda for promising NLP applications in marketing and provide supplementary code examples to help interested scholars to explore opportunities related to NLP in marketing.

Keywords: Natural language processing; text mining; text analytics; deep learning; topic modeling; sentiment analysis; word embeddings; language models; transformers; computational linguistics

---

Artificial Intelligence in Marketing
Review of Marketing Research, Volume 20, 191–215

Published under exclusive licence by Emerald Publishing Limited
ISSN: 1548-6435/doi:10.1108/S1548-643520230000020011

1. INTRODUCTION

The availability of and ability to extract meaningful information from unstructured data (e.g., image, video, voice, and text) continues to rise and has attracted much interest in the marketing community (e.g., Balducci & Marinova, 2018; Grewal, Gupta, & Hamilton, 2021). As a result, there are numerous applications leveraging unstructured data in marketing, ranging from assessing the effectiveness of marketing video content (Liu, Shi, Teixeira, & Wedel, 2018), automatically generating SEO content (Reisenbichler, Reutterer, Schweidel, & Dan, 2022) and wine reviews (Carlson, Kopalle, Riddell, Rockmore, & Vana, 2022), to analyzing the success of multimodal branded social media content (Hartmann, Heitmann, Schamp, & Netzer, 2021).

Among the unstructured data types, text has been the most frequently analyzed modality in marketing to date, presumably due to the relatively higher accessibility of textual data compared to images, voice, and video, and the relative approachability of numerous text mining methods. While the first applications of modern natural language processing (NLP) in marketing appeared at the end of the first and start of the second decade of the new millennium (e.g., Eliashberg, Hui, & Zhang, 2007; Lee & Bradlow, 2011; Netzer, Feldman, Goldenberg, & Fresko, 2012), the development of NLP tools for content analysis dates back to the 1960s (e.g., Stone, 1966). However, the automatic extraction of information from textual data did not gain popularity until the late 1990s (see Humphreys, 2019 for a historical review).

Applications of NLP outside of marketing are broad and range from the prediction of election outcomes (Bovet, Morone, & Makse, 2018) and disease spread (Yang, Santillana, & Kou, 2015) to monitoring public wellbeing from user-generated social media data (Jaidka et al., 2020) and tracking social stereotypes over time (Charlesworth, Caliskan, & Banaji, 2022). Deep learning has been particularly helpful in analyzing these types of unstructured data (LeCun, Bengio, & Hinton, 2015; Sejnowski, 2020).

In this chapter, we review established NLP methods for traditional tasks (e.g., LDA for topic modeling and lexicons for sentiment analysis) and provide an outlook on the future of NLP in marketing, covering recent embedding-based approaches for novel tasks such as automated text generation and multimodal representation learning. We elaborate on the opportunity provided by the rich information embedded in textual information and argue that – while the excitement about new modes of unstructured data such as images or videos is understandable (Grewal et al., 2021) – NLP applications in marketing are here to stay. The opportunity is still vast and with the additional development of methods to extract meaningful information from textual data, additional applications will arise, leading to a second "NLP spring" in marketing.

While previous reviews of textual analysis in marketing primarily focused on the application of textual analysis, uniting the tribes of the marketing field (Berger, Humphreys, et al., 2020; Humphreys & Wang, 2018), our unique focus in this chapter is on the exciting and novel methods in the field and the opportunity they pose for emerging applications in marketing. The chapter is structured as follows. First, we discuss the dual role of language, either revealing information about the producer or affecting the recipient (Berger, Humphreys, et al., 2020). In this context, we also discuss the role of text in subsequent analyses, either acting as the independent or dependent variable in the investigation of marketing phenomena, and address the importance of leveraging text to study causal effects. Second, we review the current state of NLP, focusing on established methods for established tasks like concept and topic extraction or sentiment and writing style extraction. Third, we provide an outlook on the future, enabled by emerging methods like embedding-based approaches, pretrained language models, and transfer learning. We discuss the opportunities from leveraging both static and contextual embeddings, whose main strengths lie in their ability to model the relationship across words and sentences and to understand the meaning of words in the context in which the appear. These emerging approaches allow the field to improve its ability to perform certain tasks that we have been using for more than a decade (e.g., text classification). But more importantly, they open the opportunity to perform entirely new types of tasks that bring about novel research opportunities (e.g., text generation, text summarization, question answering). We provide code examples for several of the new tasks to facilitate their application in marketing.[1] We conclude with a roadmap and research agenda for promising NLP applications in marketing.

2. THE ROLE OF TEXT IN MARKETING

2.1 The Dual Role of Language

Understanding the dual role of language is crucial when working with textual data. Text can reflect information about the producer (e.g., a social media user posting or commenting on Twitter) or affect the recipients (e.g., investors reacting to firms' press releases). Consumers, firms, investors, institutions/society can act both as text receivers or producers in different constellations (see Berger, Humphreys, et al., 2020 for a comprehensive summary of the dual role of language).

Marketing scholars have explored text production and what it reveals about the writer of the text in various contexts. Netzer, Lemaire, and Herzenstein (2019) investigate what the text may signal about the borrower of a loan and their likelihood of defaulting on their loan. Chung, Johar, Li, Netzer, and Pearson (2022) extract from textual data the motivation of people to engage in the sharing economy. Hartmann et al. (2021) infer expressed purchase intentions from unstructured social media comments. On the other hand, other studies have directly focused on how language affects the reader. For example, Toubia, Berger, and Eliashberg (2021) assess how narrative arc in movies or academic papers affects their success. All these studies have in common, which is often unique to NLP applications in marketing, that they not only extract meaningful constructs from the textual data but also seek to relate them to relevant marketing, business, or societal outcomes.

2.2 Text as Independent Variables

In terms of its role in marketing applications, text is most frequently used as independent variables. That is, text features are used to predict or explain some outcome variable of interest. For example, Berger and Milkman (2012) predict the virality of news articles based on the emotion conveyed by the texts. Packard and Berger (2021) analyze how the concreteness of language shapes customer satisfaction. Rocklage, Rucker, and Nordgren (2021) predict marketplace success based on mass-scale emotionality. Tirunillai and Tellis (2012) relate user-generated content to abnormal returns on stock markets. Liu, Lee, and Srinivasan (2019) analyze the effect of consumer review content on sales conversion.

In econometric analyses, text is often combined with nontextual data. Nontextual data can be the focal variable that the researcher is trying to predict or understand (e.g., sales or stock prices), they may be predictors that the researcher includes in addition to the text variables in the model, or they can be data that are used to validate the textual approach. Nontextual data may be fully external to the textual data or may include quantifiable summaries of the textual information such as the average document length.

2.3 Text as the Dependent Variable

While NLP methods have predominantly been used to generate independent variables, few studies extract the outcome variable of interest from text. Melumad, Inman, and Pham (2019) explore how smartphone use changes the linguistic characteristics of user-generated content. Woolley and Sharif (2021) study how incentives increase the relative positivity of review content. Hartmann et al. (2021) train a text classifier to identify expressed purchase intentions in user-generated social media comments.

The opportunity to extract strong marketing-relevant dependent variables from textual data can increase the potential impact of marketing research. One of the reasons for limited research using dependent variables extracted from text is that such extracted outcomes tend to be a noisy measure of the true outcome due to possible limited accuracy of the extraction method. As NLP tools advance, the extraction of such variables will become more accurate. We believe there is an opportunity for the field to look for outcomes that may be hidden in unstructured data.

2.4 Establishing Causality

NLP methods are often used in the context of observational data and, due to textual data's high dimensionality, often involve machine learning in the extraction phase. This has likely led to the perception that NLP methods are primarily suitable for descriptive or predictive tasks. However, marketers are often interested in causal effects (Goldfarb, Tucker, & Wang, 2022). If the data come from a randomized controlled trial experiment, or if one can use causal methodology for observational studies like instrumental variables, or regression discontinuity, causal inference is possible, and researchers can use such textual data to test social science theories (Egami, Fong, Grimmer, Roberts, & Stewart, 2018). Noteworthy examples of studies that have leveraged text to draw causal conclusions include Puranam, Narayan, and Kadiyali (2017), who use a difference-in-difference estimation approach to study the causal effect of calorie posting regulation on the proportion of health-related discussions in consumer reviews. Simonov and Rao (2022) estimate a structural model of demand for news. Puranam, Kadiyali, and Narayan (2021) exploit a natural experiment to estimate the causal effect of minimum wages on consumer perceptions of service. An emerging area of research in computer science and statistics focuses on causal effects using text data (e.g., Keith, Jensen, & O'Connor, 2020; Veitch, Sridhar, & Blei, 2020). We expect that these advances will percolate also into the marketing field.

3. THE CURRENT STATE OF NLP IN MARKETING

The current use of NLP methods in marketing can be classified into three categories based on the level and type of information that the researcher is trying to extract from the text: These are *concept and topic extraction, relationship extraction, and sentiment and writing style extraction*. Table 1 describes the methods available to extract the corresponding three information types and examples of their application in marketing.

3.1 Concept and Topic Extraction

The objective of concept and topic extraction is to identify single words, n-grams, or entire topics in text. For many marketing applications, it is required to identify certain concepts that are the focus of the analysis, e.g., brand names for brand buzz monitoring (Klostermann, Plumeyer, Böger, & Decker, 2019) or for social listening (Liu et al., 2019; Netzer et al., 2012). A related task in NLP is commonly known as named entity extraction (NER), which can be used to identify organizations, people, or locations. Among the most popular methods of NER is the Stanford Named Entity recognizer (Finkel, Grenager, & Manning, 2005) and the more advanced InferNER approach (Shahzad, Amin, Esteves, & Ngomo, 2021).

Moving from words to topics, unsupervised topic models such as LDA (Blei, Ng, & Jordan, 2003) have been employed across a wide variety of contexts for topic extraction. Among the first applications of LDA in marketing is by Tirunillai and Tellis (2014), who employ LDA to enable strategic brand analysis from user-generated content. LDA has also been used to extract brand-relevant information from social tags (Nam, Joshi, & Kannan, 2017), content of loan application requests (Netzer et al., 2019), and the practical relevance of marketing articles (Jedidi, Schmitt, Ben Sliman, & Li, 2021).

Several extensions to the traditional LDA have been proposed. Toubia, Iyengar, Bunnell, and Lemaire (2019) use a seeded LDA to extract psychological themes from entertainment products. Büschken and Allenby (2016) extend the traditional LDA approach to construct a sentence-based LDA that avoids the

Table 1. Established and Novel Text Analysis Tools in Marketing.

	Established Tools	Marketing Examples	Novel Tools	Marketing Examples
Established tasks				
Concept and topic extraction	• Traditional machine learning methods (e.g., naïve Bayes, support vector machines) • Stanford NER • Latent Semantic Analysis • Latent Dirichlet Allocation • Poisson Factorization	• Netzer et al. (2019) • Tirunillai and Tellis (2014) • Toubia et al. (2019)	• Deep contextual language models (e.g., RoBERTa) • InferNER • BERTopic • Embedded Topic Model (ETM) • Supervised Deep Topic Modeling (sDTM) • Supervised Hierarchical Dirichlet Process (SHDP)	• Puranam et al. (2021) • Hartmann et al. (2021) • Boughanmi and Ansari (2021) • Yang, Zhang, and Fan (2022)
Relationship extraction	• Word co-occurences • SimLex-999 • Handwritten rules • WordNet • Stanford Parser	• Netzer et al. (2012)	• Static embeddings (e.g., word2vec, GloVe) • Contextual embeddings (e.g., BERT-based, SentenceBERT)	• Timoshenko and Hauser (2019) • Toubia et al. (2021)
Sentiment and writing style extraction	• Dictionaries (e.g., Evaluative Lexicon 2.0) • General-purpose (e.g., LIWC) and specialized sentiment dictionaries (e.g., VADER) • Traditional machine learning methods	• Berger and Milkman (2012) • Villarroel Ordenes et al. (2019)	• Dictionaries (automated; e.g., Wordify) • Aspect-based sentiment analysis (ABSA) • Domain-adapted language models (e.g., SiEBERT) • BART-NLI and SetFit for few and zero-shot learning	• Hovy, Melumad, and Inman (2021) • Chakraborty et al. (2022) • Hartmann et al. (2022)
Novel tasks				
Text generation	–	–	• ChatGPT, GPT-3 • Plug & Play LM (PPLM) • LSTM	• Reisenbichler et al. (2022) • Carlson et al. (2022)
Text summarization	–	–	• Sequence-to-sequence models (e.g., BART, T5)	• –
Multimodal representation learning	–	–	• CLIP • Custom multimodal network architectures	• Dew, Ansari, and Toubia (2022)

bag-of-word approach assumption, in which the word order of the original textual unit is not maintained. Liu and Toubia (2018) extend the traditional LDA to a hierarchical LDA that combines the short text in the search query with the longer text of the search page results.

An alternative approach to topic modeling is based on Poisson factorization (Gopalan, Hofman, & Blei, 2013). Poisson factorization has several advantages relative to the LDA approach. First, Poisson factorization is better suited for sparse and short textual responses (Canny, 2004). Second, unlike LDA, Poisson factorization does not assume that the distribution of topics in a document sums up to 1. Hence, some documents can include more topics than others. Several recent marketing papers have used Poisson factorization for topic modeling (e.g., Chung et al., 2022; Liu, Toubia, & Hill, 2021; Toubia et al., 2021).

3.2 Relationship Extraction

Relationship extraction seeks to extract and identify relationships among words and entities. At the most basic level, relationship extraction can be captured by the mere co-occurrence of words. For example, to create market maps leveraging co-occurrences of brands (Netzer et al., 2012). Toubia and Netzer (2017) assess the creativity of ideas using semantic subnetworks of the words or concepts in the idea. Boghrati and Berger (2019) analyze misogyny in song lyrics by assessing how different traits (e.g., competence) are related to men and women. To extract more complex relationships like customer needs (Timoshenko & Hauser, 2019) or social media adverse drug reactions (Feldman, Netzer, Peretz, & Rosenfeld, 2015), one needs to go beyond mere co-occurrence and bag-of-words approaches.

While the opportunities for relationship extraction in marketing are vast, their application to date has been limited, possibly due to the relative complexity involved in capturing such relationships and assessing semantic similarity. This is likely to change with the rise and increasing accessibility of embedding-based methods, which we discuss later in this chapter.

3.3 Sentiment and Writing Style Extraction

Lexicons such as the Linguistic Inquiry and Word Count (LIWC; Pennebaker, Boyd, Jordan, & Blackburn, 2015), VADER (Hutto & Gilbert, 2014), and the Evaluative Lexicon 2.0 (Rocklage, Rucker, & Nordgren, 2018) have been a prevalent choice for marketing researchers for sentiment and writing style extraction (Hartmann, Huppertz, Schamp, & Heitmann, 2019). TextAnalyzer (http://textanalyzer.org/) summarizes multiple lexicons (Berger, Sherman, & Ungar, 2020). To quantify the information conveyed by a document, the document's words are compared to the word lists contained in the lexicon. LIWC has been used in numerous marketing articles (e.g., Berger & Milkman, 2012; Hartmann et al., 2021; Netzer et al., 2019) to capture the effect of writing styles on different behavioral outcomes. Simple word lists can allow users to measure specific text dimensions like the number of personal pronouns (Packard & Berger, 2020) or more nuanced dimensions such as text concreteness (Brysbaert, Warriner, & Kuperman, 2014) or authenticity (Pennebaker et al., 2015).

Lexicons are often defined in a top-down approach allowing theory to inform which words to include in a lexicon (Humphreys, 2019). This contributes to the interpretability of lexicons. Due to their relative ease of use for researchers who are novice at NLP research, these tools are popular in the marketing and social science fields, and novel lexicon-based tools continue to be developed, e.g., Wordify (Hovy et al., 2021).

However, there are several limitations to lexicons. For complex constructs, creating a well-performing lexicon can be difficult and require intense human labor (Chapman, 2020). Moreover, lexicons tend to be inferior even to traditional machine learning methods regarding classification accuracy (Hartmann, Heitmann, Siebert, & Schamp, 2022), which may lead to erroneous conclusions (Jaidka et al., 2020). One of the reasons for the limited performance is the context-dependent performance of lexicons (Berger, Humphreys, et al., 2020). The same lexicon may not be appropriate to extract writing style in both social media and news corpora. It is important to note that when developing a lexicon, special care should be taken to ensure internal, external, and construct validity.

If the desired lexicon does not exist, researchers can train their own task-specific writing style classifier using machine learning methods such as a random forest, naïve Bayes, or neural networks. For example, Villarroel Ordenes et al. (2019) train a support vector machine to classify assertive, expressive, and directive text.

Extracting the sentiment of a text is among the most popular NLP tasks in marketing and has been employed across many different applications. Sentiment analysis is the "identification of positive or negative orientation of textual language" (Hirschberg & Manning, 2015, p. 265). Kübler, Colicev, and Pauwels (2020) investigate how different sentiment extraction tools are suitable to predict social media's impact on consumer mindset metrics. Numerous studies have investigated the effect of sentiment in user-generated content on aspects like sales (e.g., Tang, Fang, & Wang, 2014) or stock market performance (e.g., Tirunillai & Tellis, 2012). Chakraborty, Kim, and Sudhir (2022) perform aspect-based sentiment analysis to create a more nuanced understanding of user-generated communication about particular elements of a product or service. Closely related to sentiment analysis is the detection of different emotions (e.g., anger, joy, disgust). For example, Rocklage and Fazio (2020) analyze contexts in which emotional content can backfire.

While many applications of sentiment analysis have relied on lexicon-based tools (Hartmann et al., 2019), state-of-the-art machine learning models often outperform more traditional approaches for sentiment analysis (Hartmann et al., 2022). In the next section, we discuss how frontier NLP approaches have both improved our ability to extract concepts and topics, relationships, and sentiment and writing style, as well as open opportunities for additional novel applications (see also Table 1).

4. THE NEXT METHODOLOGICAL FRONTIER: EMBEDDINGS, LANGUAGE MODELS, TRANSFER LEARNING

NLP applications in marketing continue to move from extracting single words, to combining words into topics, to relationship extraction. This development is likely to be accelerated by novel technologies introduced in the NLP field. A defining element of the next methodological frontier in marketing are embedding-based methods, including static word embeddings such as word2vec (Mikolov, Chen, Corrado, & Dean, 2013), GloVe (Pennington, Socher, & Manning, 2014), and fastText (Bojanowski, Grave, Joulin, & Mikolov, 2017) as well as pretrained deep contextual language models such as BERT (Devlin, Chang, Lee, & Toutanova, 2019), RoBERTa (Liu et al., 2019), and GPT-3 (Brown et al., 2020).

Deep contextual language models allow researchers to study fine-grained relationships between words and concepts (e.g., drugs appearing with certain side effects or problems with certain product features). In addition, they have the appealing property that they can benefit from the use of *transfer learning*. Transfer learning in NLP applications typically consists of two steps. First, a large language model is *pretrained* on massive datasets such as English Wikipedia or the PILE, an 800GB dataset containing diverse text for language model training (Gao et al., 2020). This approach benefits from work by large organizations like Google, Amazon, and Meta that have access to large amounts of data and advanced computing. Second, the researcher can *fine-tune* the pretrained, publicly available language model to complete a task in a particular application context. Transfer learning substantially reduces the amount of annotated training data needed for the focal research objective and has contributed to the remarkable performance levels that these models have achieved (Hartmann et al., 2022). We discuss static embeddings and vector semantics next, followed by a discussion of transfer learning and different types of language models.

4.1 Embeddings and Vector Semantics

The idea behind word embedding methods is to represent a word by the words that tend to appear in its vicinity. The problem is that the space of words that appear next to each word can be very large and very sparse. To resolve the sparsity problem, the word is represented by a dense vector of lower dimensionality that is learned using the semantic information from local, i.e., neighboring, words (word2vec) or from global word co-occurrences (GloVe). Once word embeddings are constructed, one can look for relationships between words by calculating similarities between words such that words that appear in similar contexts tend to carry similar meanings. This concept already originated in the 1950s and is also known as the *distributional hypothesis* (Harris, 1954). Consider, for example, the near-synonymous words "couch" and "sofa." These words are likely to occur in comparable sentence environments with similar neighboring words such as seating or television (Jurafsky & Martin, 2021). Consequently, their embedding representations would be closely aligned in multidimensional semantic space (Le & Mikolov, 2014).

Among the most popular methods to represent words as vectors is word2vec (Mikolov, Chen et al., 2013). Word2vec embeddings are considered static because the same word used in different contexts will have the same fixed embedding.

Consequently, static word embeddings cannot reflect polysemy, i.e., the same word carrying multiple related meanings such as *bank* (as a noun versus as a verb or as the land next to a river versus a financial institution). Despite this limitation, word2vec has proven powerful in various downstream applications to study the semantic relationship of words (e.g., Boghrati & Berger, 2019; Charlesworth et al., 2022), explore relational meanings and analogies (e.g., Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), and historic change in word meaning, so-called diachronic word embeddings (Hamilton, Leskovec, & Jurafsky, 2016).

In contrast to labor-intensive feature engineering, embeddings make use of self-supervised representation learning, in which the vector representations can be learned automatically from text without supervision (Mikolov, Chen et al., 2013) and transferred to different applications (Kim, 2014). Another advantage of embedding techniques, compared to traditional bag-of-words approaches, is that they often can deal with out-of-vocabulary words, words that the algorithm has not seen before, as long as the words in the proximity of unknown words are familiar. fastText, for example, works on character-level n-grams instead of on word-level. Consequently, it can return embeddings also for unseen or misspelled words if they contain substrings of characters that are similar to words it has been trained on. In some situations, embeddings are needed at the sentence level, for example, to identify the similarity between product reviews or between different social media users based on their self-descriptions. A simple approach is to average the word-level embeddings which belong to the sentence. Alternatively, methods such as doc2vec (Le & Mikolov, 2014) and SentenceBERT (Reimers & Gurevych, 2019) provide specialized architectures that are optimized to return semantically meaningful sentence embeddings.

Returning to the earlier embeddings example of the words "couch" and "sofa", these two words are indeed most similar when looking up the words with the highest similarity using a pretrained word2vec model, having a similarity of 0.83 (see Fig. 1). The underlying model ("word2vec-google-news-300") was trained on approximately 100 billion words and is publicly available.[2]

4.2 Deep Learning Architectures for NLP

"Language is a sequence that unfolds in time" (Jurafsky & Martin, 2021, Chapter 9, p. 1). Several deep learning architectures exist to model human language and some can reflect its temporal nature. Recurrent neural networks (RNNs, Elman, 1990) are among these architectures. RNNs process a sentence, i.e., a sequence of words, sequentially, one word at a time. In contrast to classical feedforward neural networks, which are sequence agnostic, an RNN's hidden layer includes a recurrent connection. At time t, the layer processes a hidden state h_{t-1}, which captures information from words that have already been processed at preceding points in time, in addition to the input x_t. There is no limit in terms of the length of the prior context to be included in this "fluid representation" of a text sequence's meaning (Chollet, 2021, p. 293). RNNs can be used for different NLP tasks. Intuitively, they can be used for text generation, with the objective to predict the next word in a sequence of text from the current word w_t and the

```
[ ] import gensim.downloader
```

▾ word2vec

```
[ ] w2v = gensim.downloader.load('word2vec-google-news-300')

    [==================================================] 100.0% 1662.8/1662.8MB downloaded
```

```
[ ] w2v.most_similar('sofa')

    [('couch', 0.8309178352355957),
     ('settee', 0.7764685750007629),
     ('sofas', 0.7543261051177979),
     ('loveseat', 0.7152645587921143),
     ('recliner', 0.7101271152496338),
     ('futon', 0.6624690294265747),
     ('leather_sofa', 0.6620596647262573),
     ('plush_sofa', 0.6556485295295715),
     ('ottoman', 0.6525834798812866),
     ('couches', 0.6501914262771606)]
```

```
[ ] w2v.most_similar('couch')

    [('sofa', 0.8309179544448853),
     ('recliner', 0.7366936802864075),
     ('couches', 0.701655209064836),
     ('comfy_couch', 0.6747691035270691),
     ('futon', 0.6523739695549011),
     ('al_Jabouri_slept', 0.6240309476852417),
     ('loveseat', 0.617920994758606),
     ('beanbag_chair', 0.616889476776123),
     ('recliner_chair', 0.6121512055397034),
     ('settee', 0.6086535453796387)]
```

```
[ ] w2v.similarity('couch', 'sofa')

    0.8309179
```

Fig. 1. Code Example for Similarity Comparison Between the Words "Couch" and "Sofa." *Note:* All code examples from this book chapter are available on GitHub: https://github.com/j-hartmann/nlp-in-marketing.

previous hidden state h_{t-1} (Mikolov, Karafiát, Burget, Cernocký, & Khudanpur, 2010).[3] However, they can also be employed for text classification tasks, where the final hidden state is used as the input for a classifier such as an emotion classifier given the words up to word w_t.

A more complex derivative of RNNs are Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997). LSTMs address the limitation of RNNs that the hidden states are biased to contain more information at the end of the sequence. Moreover, in training RNNs, due to the chain of computations, one may be faced with the *vanishing gradients problem* (Hochreiter, Bengio, Frasconi, & Schmidhuber, 2001), a problem that occurs when the noise that is induced in the sequential transmission of error information during back-propagation overwhelms the gradient information (Chollet, 2021). For this purpose, LSTMs contain gates to control the flow of information along the network. These gates help LSTMs to selectively forget and remember information. Recent LSTM applications in marketing include sentiment analysis (Chakraborty et al., 2022; Li, Liao, & Xie, 2021) as well as sales conversion prediction (Liu et al., 2019).

Although commonly used for image classification tasks, convolutional neural networks (CNNs) trained with backpropagation (LeCun et al., 1989) have also been used for textual analysis in marketing applications (e.g., Chakraborty et al., 2022; Liu et al., 2019; Puranam et al., 2021; Timoshenko & Hauser, 2019). In contrast to RNNs, CNNs do not model the sequential nature of language and have no temporal memory (Chollet, 2021). However, CNNs can account for adjacent word groups or phrases and perform well in combination with LSTMs (e.g., Chakraborty et al., 2022).

4.3 Transfer Learning and Transformer Models

The advent of the transformer architecture (Vaswani et al., 2017) has pushed language models' performance and possibilities to unprecedented levels, quickly overtaking LSTMs as the state-of-the-art architectures. Transformers do not rely on convolutional or recurrent connections and can process a text sequence in parallel. Using so-called self-attention layers, when processing a word, the network can selectively attend to the other words in the word's context, allowing it to model linguistic relationships across words (more precisely, *tokens*). Thereby, in contrast to static word embeddings such as word2vec, transformer models can represent words using contextualized embeddings, which helps for word sense disambiguation. Instead of being assigned a fixed position in geometric feature space, contextual embeddings reflect the context-dependent meaning of words in the specific context the word appeared in. For static word embedding methods, the word "apple" will have the same vector representation whether it appears in a technology context and refers to the technology powerhouse, or in Snow White referring to the poisoned fruit. Transformer models help solve this ambiguity. Moreover, the parallel processing of a text addresses the vanishing gradients problem of RNNs mentioned earlier (Jurafsky & Martin, 2021). Combining the word embeddings with positional encodings is crucial for the transformer to capture word order information. In addition, by processing the same text multiple times in parallel, transformers can model different types of relationships between words in the same sentence (e.g., between a noun and its pronoun, or other longer-range and shorter-range word dependencies). This design is known as multihead attention (Vaswani et al., 2017).

One of the most popular transformer models is BERT (Bidirectional Encoder Representations from Transformers; Devlin et al., 2019). BERT is trained through next sentence prediction and masked language modeling, where words in the training data are randomly masked, and the training objective is to recover these words correctly. This approach helps the models to capture the linguistic structure of language without explicit supervision signals (Manning, Clark, Hewitt, Khandelwal, & Levy, 2020). Due to their ability to process text in parallel, transformer models scale well to large amounts of data. BERT, for example, was trained on the BooksCorpus dataset (800 million words) and Wikipedia (2.5 billion words; Devlin et al., 2019). This contributes to transformer models' appealing property of being suited well for *transfer learning* for various downstream tasks, e.g., sequence classification, or part-of-speech tagging.

The three main types of transformer models are *autoregressive, autoencoding, and sequence-to-sequence* models (Hugging Face, 2022). Autoregressive or causal language models such as GPT-2 (Radford et al., 2019) and GPT-3 (Brown et al., 2020) are suited well for text generation (e.g., Reisenbichler et al., 2022). In contrast to BERT, which is trained to recover masked words based on left (past) and right (future) contexts (Devlin et al., 2019), the training objective of GPT-2 is to create a generative tool that generates text given a textual prompt, i.e., a starting sequence of words that the model will complete, e.g., "The sun is shining. Let's go…." Hence, it only considers the left (past) context. Autoregressive models estimate the probability $P(y|x)$, where y represents a sequence of predicted words $y = y_1, y_2, \ldots, y_n$, and x represents the prompt $x = x_1, x_2, \ldots x_n$ of existing words (Tunstall, von Werra, & Wolf, 2022). Once trained, an autoregressive model such as GPT-3 can create a text of arbitrary length by starting with a prompt and iteratively appending the most likely next word to that sequence. The example prompt could be completed to: "The sun is shining. Let's go *outside and have some ice cream*."

Autoencoding models such as BERT and RoBERTa use both the left and right context when encoding a sequence of words. Clearly, this is beneficial when trying to infer complex relationships between words for tasks like sentiment analysis that often requires a nuanced understanding of the dependencies across words in a sentence and would hence suffer if restricted to only the left context. They are also suited well for tasks such as sentence classification, NER, and extractive question answering (Hugging Face, 2022). In marketing, autoencoding models have been used to infer purchase intentions from user-generated text (Hartmann et al., 2021) and to study customer perceptions of service (Puranam et al., 2021).

Lastly, there are sequence-to-sequence (or encoder–decoder) models such as BART (Lewis et al., 2019) and T5 (Raffel et al., 2019). As their name suggests, these models take as input a sequence and return as output a sequence. The encoder transforms the input sequence into an intermediate representation, while the decoder predicts the next word w_t in a target sequence, considering both the encoded input as well as the previous words w_1 to w_{t-1}. NLP tasks that require this kind of encoder–decoder architecture include text summarization, language translation, and generative question answering (e.g., from "How is the weather in NYC?" to "You can expect a sunny day.").

It is important to note that the three main transformer types are not mutually exclusive. For example, recent work demonstrates that pretrained BERT, GPT-2, and RoBERTa checkpoints can be used to initialize a model for various sequence-to-sequence tasks (Rothe et al., 2020). Similarly, ERNIE 3.0 (Sun et al., 2021) fuses an autoregressive and autoencoding network, enabling efficient application for natural language understanding and generation tasks. Despite this overlap, the taxonomy described above may facilitate effective method choice from an applied marketing perspective (e.g., using established autoregressive models such as GPT-2 for text generation; see Reisenbichler et al., 2022).

5. MARKETING APPLICATIONS OF LANGUAGE MODELS

As Table 1 showcases, novel embedding-based tools can be used not only to obtain better results in established tasks (i.e., concept and topic extraction, relationship extraction, sentiment and writing style extraction) but also to tackle novel tasks (i.e., text generation, text summarization, multi-modal content analysis). A limited number of recent marketing applications demonstrate this potential, for example, for multimodal representation learning (Dew et al., 2021) or automated text generation (Carlson et al., 2022; Reisenbichler et al., 2022).

5.1 Novel Approaches for Established Tasks

The main advantage of applying novel transformer tools for established tasks is their performance across various downstream tasks. For example, state-of-the-art language models can outperform lexicons by more than 20 percentage points in accuracy for sentiment analysis (Hartmann et al., 2022). This increased performance allows these methods to better capture linguistic nuances relative to established tools, which do not consider word order or the contextual meaning of words. In addition to the accuracy advantage, transfer learning methods often require smaller annotated training data than traditional machine learning methods (Hartmann et al., 2022).

For concept and topic extraction, the Embedded Topic Model (ETM) addresses the problem of working with large and heavy-tailed vocabularies, blending word embeddings with traditional topic models such as LDA (Dieng, Ruiz, & Blei, 2020). ETM has been shown to outperform LDA in terms of both topic quality and predictive performance (Dieng et al., 2020). For relationship extraction, language models can help grasp more fine-grained relationships across words and entities compared to co-occurrence analyses (e.g., BERTopic; Grootendorst, 2022). These relationships can be analyzed at the word level using static embeddings (e.g., word2vec or GloVe) or on sentence level using contextual embeddings (e.g., SentenceBERT).

5.2 Novel Approaches for Novel Tasks

However, beyond improving performance on existing tasks, advances in NLP open the window to address challenges that were impossible to address before. The large-scale pretraining allows language models to learn a fine-grained linguistic understanding that is helpful for various downstream tasks. Among the most impressive novel tasks is automated text generation at nearly human-like standards.[4] The recently released ChatGPT by OpenAI shows a great promise in that respect. Firms communicate with their customers, employees, investors, and the public across a large variety of channels (Berger, Humphreys, et al., 2020).

Content generation can be employed to serve users via chatbots or digital voice assistants that can automatically respond to users' inputs. This highlights the potential economic impact that this novel NLP task may generate. Similarly, automatic text summarization allows advertisers to extract rich knowledge beyond word counts from large-scale text corpora. E-commerce websites and review platforms can summarize the content of a product or a service, and elaborate consumer reviews can be aggregated into concise summaries increasing value to other users.

Recent marketing examples have employed text generation in the context of search engine optimization (SEO) (Reisenbichler et al., 2022) and the creation of wine reviews (Carlson et al., 2022). For example, Reisenbichler et al. (2022) build on GPT-2 and show how natural language generation can support content marketing by proposing a semiautomated "human in the loop" methodology that can create human-like SEO content. Carlson et al. (2022) even formulate a "kind of 'Turing test'" to test if human judges can tell which reviews were generated by humans and which ones were automatically created by a transformer model.[5]

For traditional machine learning methods, lack of training data have often been a limiting factor (Berger, Humphreys, et al., 2020). Due to their large-scale pretraining, language models (such as BART) can also be used for new tasks for which no training exist, often referred to as zero-shot classification (Yin, Hay, & Roth, 2019). One approach for zero-shot classification is to formulate a sentence–class pair (e.g., "What a great product.", "positive") as a natural language inference (NLI) task, where the sentence represents the premise and the candidate label is formulated as a hypothesis (i.e., premise: "What a great product", hypothesis: "This example is *positive*."). Label probabilities are obtained by converting the probabilities for entailment and contradiction from the NLI task. Relatedly, recent few-shot methods such as SetFit have produced remarkable accuracy levels with as little as a dozen of training examples (Tunstall, Reimers et al., 2022).

To facilitate the application of state-of-the-art language models for novel marketing applications, Table 2 provides a list of pretrained language models available for the marketing community. All models are open-access and can be applied with fairly limited coding knowledge, e.g., for off-the-shelf sentiment analysis (SiEBERT; Hartmann et al., 2022), multimodal representation learning (CLIP; Radford et al., 2021), text generation (GPT-J 6B; Wang & Komatsuzaki, 2021), and text-to-image generation (Stable Diffusion; Rombach et al., 2022). Note that commercial alternatives exist, e.g., DALL·E 2 for text-to-image generation (Ramesh, Dhariwal, Nichol, Chu, & Chen, 2022) or ChatGPT for conversational text generation (Ouyang et al., 2022).

6. DISCUSSION

For more than a decade, NLP has enabled various marketing applications across diverse contexts. In this chapter, we discussed applications of textual analysis in marketing along the dimensions of the dual role of language, its use as a

Table 2. Pretrained Models for Marketing Applications (Selection).

Models	Description	Training Data	References
Established tasks			
Concept and topic extraction			
FLERT (or BERT-base-NER)	Extracts named entities	CoNLL-03	Schweter and Akbik (2020)
BART-large-NLI	Returns probabilities for label candidates	MultiNLI (MLNI)	Lewis et al. (2019)
Relationship extraction			
multi-qa-MiniLM-L6-cos-v1	Computes sentence similarity	17 datasets (215M question–answer pairs)	Reimers and Gurevych (2019)
Sentiment and writing style extraction			
SiEBERT	Predicts binary sentiment	15 datasets	Hartmann et al. (2022)
Sentiment (English)	Predicts three-class sentiment	5,304 social media comments	Hartmann et al. (2021)
Emotion (English)	Predicts Ekman's basic emotions	6 datasets (19,677 texts)	Hartmann, Zhang, and Netzer (2022)
Novel tasks			
Text generation			
GPT-J 6B	Generates text based on prompt	The PILE (Gao et al., 2020)	Wang and Komatsuzaki (2021)
DistilGPT2	Generates text based on prompt	OpenWeb-TextCorpus	Sanh, Debut, Chaumond, and Wolf (2019)
Text summarization			
DistilBERT-base-cased-distilled-squad	Returns answers to questions	SQuAD v1.1	Sanh et al. (2019)
Multimodal representation learning			
CLIP	Different tasks (including image classification)	400M image-text pairs	Radford et al. (2021)
Whisper	Transcribes speech to text	680,000 hours of audio	Radford et al. (2022)
Stable Diffusion	Generates images based on text prompts	LAION dataset (Schuhmann et al. 2022)	Rombach, Blattmann, Lorenz, Esser, and Ommer (2022)

Note: The supplementary GitHub repository contains examples for each task: https://github.com/j-hartmann/nlp-in-marketing.

dependent and independent variable, as well as the opportunity to study causal effects using text data. Moreover, we discussed how NLP methods have been employed for established tasks in marketing, i.e., *concept and topic extraction, relationship extraction, sentiment and writing style extraction*, and delineated novel tasks enabled by recent technological advancements, i.e., *text generation, text summarization, and multimodal representation learning*.

6.1 Roadmap and Future Trends

While the body of work leveraging textual analysis in marketing already seems vast, we believe there are many opportunities ahead, leading to a rapid proliferation of language models in applied marketing research. These opportunities will stem from new business applications that involve textual data (e.g., sharing economy, streaming, and metaverse) as well as advances in the development of NLP tools. As outlined before, the novel methods, including state-of-the-art transformer models, can be used to address existing tasks with higher performance levels or to generate novel applications (e.g., natural language generation). Next, we emphasize three trends that are likely to shape the future of NLP applications in marketing.

First, there is an immense opportunity in leveraging *transfer learning* in the marketing community. While the NLP and computer science communities do well in freely distributing their models, data, and code, (e.g., BLOOM by Scao et al. (2022), trained on 59 languages, or Meta's recent release of the open pre-trained transformer with 175 billion parameters, called OPT-175B; Zhang et al., 2022), the marketing community has room to grow in that respect. Rather than each paper starting from scratch training its own model on its own data, if we share our data and trained models, we can transfer learning from one application to another. The Open Science movement is a positive step in that direction. Consider, for example, the dozens of papers analyzing consumer reviews. If these papers would share their data and trained models, subsequent papers can build on these to generate new insights with higher accuracy levels and richer language content. In that sense, the community will be able to build on each other's work not only substantively but also methodologically. Combining multiple datasets for more extensive and more diverse training data (e.g., SiEBERT and the PILE) can lead to better outcomes. Moreover, the community is likely to benefit from benchmarks that cover a broader variety of NLP tasks (e.g., XTREME; Hu et al., 2020), whose leaderboards can serve as an effective starting point for researchers to explore method options (e.g., GLUE, SQuAD).

Second, capturing the *true* relationship across words, sentences, and concepts is a promising area of research. While embedding-based methods such as word2vec and related applications have enabled many fascinating applications, there is still a long way to go to automatically disentangle related constructs such as similarity, coherence, and relatedness using automated methods. Despite the impressive performance leap introduced by transformer models, they still struggle with many language patterns that humans can easily interpret, like simple negations (Hossain, Chinnappa, & Blanco, 2022) or sarcasm. Further work is needed to identify and address these shortcomings to enable reliable applications in marketing and social sciences.

Third, we see a promising opportunity in multimodal representation learning using methods such as CLIP (Radford et al., 2021) and others as well as blending multiple sources of multimedia data such as text, image, and video using custom multimodal network architectures (see Grewal et al., 2021 for a recent discussion). Until now, applied examples in marketing remain rare (e.g., Boughanmi & Ansari, 2021; Dew et al., 2022).

6.2 Challenges, Biases, and Potential Harms

Human language is riddled with ambiguity and in constant flux. Several challenges remain that can be addressed by the marketing community. Lack of interpretability is often called out as a limitation of black-box deep learning models (Rai, 2020). Better understanding the errors and sensitivity of a transformer model will be helpful in building trust in the model's prediction. Moreover, adding such a layer of interpretability helps to understand the variables of interest and their nuances better. Methods such as Local Interpretable Model-agnostic Explanations (LIME; Ribeiro et al., 2016) can help shed light on "why" certain predictions are made. For example, Hartmann et al. (2021) combine LIME with RoBERTa to highlight which words are associated with expressed purchase intentions. Alternatively, one can use variable selection tools such as regularized regression to reduce the number of predictors (e.g., Netzer et al., 2019). Dimensionality reduction tools such as topic modeling, or the use of categorized dictionaries, often allow the researcher to reduce the dimensionality of the problem such that the textual variables can enter into traditional econometric inference approaches in an intuitive and explainable manner.

Transfer learning dramatically reduces the need for annotated training data in the fine-tuning stage. At the same time, large-scale pretraining can introduce biases into the model. Language models pretrained on unfiltered massive internet text data can replicate toxic language, amplify implicit biases (Hartmann, Schwenzow, & Witte, 2023), perpetuate stereotypes of religion (Abid, Farooqi, & Zou, 2021), gender (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016), or sexual orientation (Sheng, Chang, Natarajan, & Peng, 2019), and may pose privacy threats when trained on sensitive data (Carlini et al., 2021). Debiasing these models has received recent attention in the literature (e.g., Schramowski, Turan, Andersen, Rothkopf, & Kersting, 2022; Zhao, Wang, Yatskar, Ordonez, & Chang, 2017), but remains an open research question. Human biases can also creep into the models in the data annotation stage. Careful coding guidelines and quality checks can help address these risks. As marketing researchers, we should be aware of the different sources of biases and develop methods to address them.

NLP methods rely on large amounts of textual training data for pretraining and relevant textual data for any focal application. Sparse model architectures (e.g., ST-MoE; Zoph et al., 2022) can help reduce the appetite of data-hungry deep learning models and improve their training efficiency. Relatedly, researchers should pay close attention to legal and ethical privacy concerns when working with large-scale textual data. For example, many platforms restrict data scraping for academic research purposes. In such cases, using an Application Programming Interface (API) may be a more appropriate route (Boegershausen, Datta, Borah, & Stephen, 2022). Researchers should also ensure to remove any identifiable information when dealing with consumer-level textual data like call center conversations or chats. NER techniques can help identify such personal data.

When extracting constructs from textual data, validation is a core concern. How do researchers know that they have extracted what they think they did? Assessing a method's accuracy is important not only for top-down lexicons that are derived from theoretical considerations (Berger, Humphreys, et al., 2020; Humphreys, 2019) but also for bottom-up methods such as machine learning classifiers. For the latter, evaluation tends to be more common as the training process requires annotated training data. Using off-the-shelf lexicons without assessing their performance for a given task may be risky. The same applies to commercial black-box solutions, which seldom release their source code. For example, Borah and Tellis (2016) employ commercial software for text classification, but used research assistants for validation, to study the effect of product recalls on competitor brands. In addition, nontextual data can and, if possible, should be used to validate the text mining algorithm. For example, Netzer et al. (2012) compared the market structure maps derived from text mining social media posts with those derived for actual car switching, obtaining high degree of convergence. Similarly, Schweidel and Moe (2014) compared brand health measures from social media with those obtained from traditional surveys.

Lastly, it is important to note that even for machine learning models with equivalent training domain performance, the interim representations of such models may still look and perform very differently due to a challenge known as "underspecification" (D'Amour et al., 2020). Once deployed in practice, model performance can deteriorate unexpectedly, which requires special attention for research projects that bridge academia and practice.

7. CONCLUSION

To conclude, NLP in marketing is here to stay. The advent of pretrained transformer models has opened new and exciting avenues for marketing scholars. These include text generation, text summarization, and multimodal content representation and can be used for a myriad of applications, including chatbots, voice assistants, and fine-grained semantical relationships across words, sentences, and concepts. We hope this chapter and the supplementary code examples help interested scholars in exploring these rich opportunities related to NLP in marketing.

ACKNOWLEDGMENTS

The authors thank Matthias Aßenmacher for feedback on an earlier version of this chapter. Jochen Hartmann is grateful for the grant "Challenging the Boundaries of Natural Language Processing," received from the Claussen-Simon Foundation.

NOTES

1. https://github.com/j-hartmann/nlp-in-marketing
2. For details on the underlying data, see Gensim-data: https://github.com/RaRe-Technologies/gensim-data.

3. To increase the learning capacity of recurrent neural networks, multiple layers can be stacked. In addition, bidirectional recurrent neural networks that process a sequence both chronologically and anti-chronologically and concatenate the resulting representations, may further improve performance (Chollet, 2021; see Wang, Qin, Luo, & Kou, 2022 for a recent marketing application).

4. When GPT-2 was introduced in February 2019 the public release of the full model was held back for six months, supposedly due to the fear that GPT-2 would allow ill-willed users to produce fake news and deceptive content at scale (https://openai.com/blog/gpt-2-6-month-follow-up/).

5. Early attempts to automatic text generation and text summarization have existed since at least the 1990s (Mitra, Singhal, & Buckley, 1997; Rambow & Korelsky, 1992). However, their adoption in marketing and related fields was slow until their performance increased with the development of transformer models.

REFERENCES

Abid, A., Farooqi, M., & Zou, J. (2021). Large language models associate Muslims with violence. *Nature Machine Intelligence*, *3*(6), 461–463.

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, *46*(4), 557–590.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1–25.

Berger, J., & Milkman, K. L. (2012). What makes online content viral?. *Journal of Marketing Research*, *49*(2), 192–205.

Berger, J., Sherman, G., & Ungar, L. (2020). TextAnalyzer. Retrieved from http://textanalyzer.org

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Boegershausen, J., Datta, H., Borah, A., & Stephen, A. T. (2022). Fields of gold: Scraping web data for marketing insights. *Journal of Marketing*, *86*(5), 1–20.

Boghrati, R., & Berger, J. (2019). *Quantifying 50 years of misogyny in music*. Working Paper.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, *29*.

Borah, A., & Tellis, G. J. (2016). Halo (spillover) effects in social media: do product recalls of one brand hurt or help rival brands?. *Journal of Marketing Research*, *53*(2), 143–160.

Boughanmi, K., & Ansari, A. (2021). Dynamics of musical success: A machine learning approach for multimedia data fusion. *Journal of Marketing Research*, *58*(6), 1034–1057.

Bovet, A., Morone, F., & Makse, H. A. (2018). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *Scientific Reports*, *8*(1), 1–16.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, *35*(6), 953–975.

Canny, J. (2004). GaP: A factor model for discrete data. In *Proceedings of the 27th annual international ACM SIGIR conference on research and development in information retrieval*, New York, NY, USA (pp. 122–129). ACM.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., . . . Raffel, C. (2021). Extracting training data from large language models. In *30th USENIX security symposium* (USENIX Security 21) (pp. 2633–2650).

Carlson, K., Kopalle, P. K., Riddell, A., Rockmore, D., & Vana, P. (2022). Complementing human effort in online reviews: A deep learning approach to automatic content generation and review synthesis. *International Journal of Research in Marketing*. In Press. doi:10.1016/j.ijresmar.2022.02.004

Chakraborty, I., Kim, M., & Sudhir, K. (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research*, *59*(3), 600–622.

Chapman, C. (2020). Commentary: Mind your text in marketing practice. *Journal of Marketing*, *84*(1), 26–31.

Charlesworth, T. E. S., Caliskan, A., & Banaji, M. R. (2022). Historical representations of social groups across 200 years of word embeddings from Google Books. *Proceedings of the National Academy of Sciences*, *119*(28), 1–17.

Chollet, F. (2021). *Deep learning with Python* (2nd Ed.). Shelter Island, NY: Manning.

Chung, J., Johar, G. V., Li, Y., Netzer, O., & Pearson, M. (2022). Mining consumer minds: Downstream consequences of host motivations for home-sharing platforms. *Journal of Consumer Research*, *48*(5), 817–838.

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., … Sculley, D. (2020). Underspecification presents challenges for credibility in modern machine learning. arXiv preprint arXiv:2011.03395.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805.

Dew, R., Ansari, A., & Toubia, O. (2022). Letting logos speak: Leveraging multiview representation learning for data-driven branding and logo design. *Marketing Science*, *41*(2), 401–425.

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439–453.

Egami, N., Fong, C. J., Grimmer, J., Roberts, M. E., & Stewart, B. M. (2018). How to make causal inferences using texts. arXiv preprint arXiv:1802.02163.

Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2007). From story line to box office: A new approach for green-lighting movie scripts. *Management Science*, *53*(6), 881–893.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Feldman, R., Netzer, O., Peretz, A., & Rosenfeld, B. (2015). Utilizing text mining on online medical forums to predict label change due to adverse drug reactions. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1779–1788).

Finkel, J. R., Grenager, T., & Manning, C. D. (2005, June). Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)* (pp. 363–370).

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., … Leahy, C. (2020). The PILE: An 800GB dataset of diverse text for language modeling. arXiv preprint arXiv:2101.00027.

Goldfarb, A., Tucker, C., & Wang, Y. (2022). Conducting research in marketing with quasi-experiments. *Journal of Marketing*, *86*(3), 1–20.

Gopalan, P., Hofman, J. M., & Blei, D. M. (2013). *Scalable recommendation with poisson factorization*. arXiv preprint arXiv:1311.1704.

Grewal, R., Gupta, S., & Hamilton, R. (2021). Marketing insights from multimedia data: Text, image, audio, and video. *Journal of Marketing Research*, *58*(6), 1025–1033.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. arXiv preprint arXiv:1605.09096.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2–3), 146–162.

Hartmann, J., Heitmann, M., Schamp, C., & Netzer, O. (2021). The power of brand selfies. *Journal of Marketing Research*, *58*(6), 1159–1177.

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2022). More than a feeling: Accuracy and application of sentiment analysis. *International Journal of Research in Marketing*. In Press. doi: 10.1016/j.ijresmar.2022.05.005

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, *36*(1), 20–38.

Hartmann, J., Schwenzow, J., & Witte, M. (2023). The political ideology of conversational AI: Converging evidence on ChatGPT's pro-environmental, left-libertarian orientation. SSRN preprint: https://ssrn.com/abstract=4316084.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266.

Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (2001). Gradient flow in recurrent nets: The difficulty of learning long-term dependencies.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hossain, M. M., Chinnappa, D., & Blanco, E. (2022). An analysis of negation in natural language understanding corpora. arXiv preprint arXiv:2203.08929.

Hovy, D., Melumad, S., & Inman, J. J. (2021). Wordify: A tool for discovering and differentiating consumer vocabularies. *Journal of Consumer Research*, *48*(3), 394–414.

Hu, J., Ruder, S., Siddhant, A., Neubig, G., Firat, O., & Johnson, M. (2020). Xtreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation. In *International conference on machine learning* (pp. 4411–4421).

Hugging Face. (2022). Hugging face course. Retrieved from https://huggingface.co/course/

Humphreys, A. (2019, May). Automated text analysis. In *Handbook of market research* (pp. 1–32).

Humphreys, A., & Wang, R. J. H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, *44*(6), 1274–1306.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8(1), pp. 216–225).

Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*, *117*(19), 10165–10171.

Jedidi, K., Schmitt, B. H., Ben Sliman, M., & Li, Y. (2021). R2M Index 1.0: Assessing the practical relevance of academic marketing articles. *Journal of Marketing*, *85*(5), 22–41.

Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing* (Vol. 3). Prentice-Hall.

Keith, K. A., Jensen, D., & O'Connor, B. (2020). Text and causal inference: A review of using text to remove confounding from causal estimates. arXiv preprint arXiv:2005.00649.

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv: 1408. 5882.

Klostermann, J., Plumeyer, A., Böger, D., & Decker, R. (2019). Extracting brand information from social networks: Integrating image, text, and social tagging data. *International Journal of Research in Marketing*, *35*(4), 538–556.

Kübler, R. V., Colicev, A., & Pauwels, K. H. (2020). Social media's impact on the consumer mindset: When to use which sentiment extraction tool? *Journal of Interactive Marketing*, *50*, 136–155.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188–1196). PMLR.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, *48*(5), 881–894.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., . . . Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.

Li, X., Liao, C., & Xie, Y. (2021). Digital piracy, creative productivity, and customer care effort: Evidence from the digital publishing industry. *Marketing Science*, *40*(4), 685–707.

Liu, X., Lee, D., & Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research*, *56*(6), 918–943.

Liu, X., Shi, S. W., Teixeira, T., & Wedel, M. (2018). Video content marketing: The making of clips. *Journal of Marketing*, *82*(4), 86–101.

Liu, J., & Toubia, O. (2018). A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science*, *37*(6), 930–952.

Liu, J., Toubia, O., & Hill, S. (2021). Content-based model of web search behavior: An application to tv show search. *Management Science*, *67*(10), 6378–6398.

Manning, C. D., Clark, K., Hewitt, J., Khandelwal, U., & Levy, O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences*, *117*(48), 30046–30054.

Melumad, S., Inman, J. J., & Pham, M. T. (2019). Selectively emotional: How smartphone use changes user-generated content. *Journal of Marketing Research*, *56*(2), 259–275.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In *Interspeech* (Vol. 2(3), pp. 1045–1048).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*.

Mitra, M., Singhal, A., & Buckley, C. (1997). Automatic text summarization by paragraph extraction. In *Intelligent scalable text summarization*.

Nam, H., Joshi, Y. V., & Kannan, P. K. (2017). Harvesting brand information from social tags. *Journal of Marketing*, *81*(4), 88–108.

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*(3), 521–543.

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, *56*(6), 960–980.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... Lowe, R. (2022). Training language models to follow instructions with human feedback. arXiv preprint arXiv: 2203.02155.

Packard, G., & Berger, J. (2020). Thinking of you: How second-person pronouns shape cultural success. *Psychological Science*, *31*(4), 397–407.

Packard, G., & Berger, J. (2021). How concrete language shapes customer satisfaction. *Journal of Consumer Research*, *47*(5), 787–806.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).

Puranam, D., Kadiyali, V., & Narayan, V. (2021). The impact of increase in minimum wages on consumer perceptions of service: A transformer model of online restaurant reviews. *Marketing Science*, *40*(5), 985–1004.

Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors. *Marketing Science*, *36*(5), 726–746.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., … Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748–8763).

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision*. OpenAI Blog. Retrieved from https://cdn.openai.com/papers/whisper.pdf

Radford, A., Wu, J., Rewon, C., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., . . . Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:1910. 10683.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141.

Rambow, O., & Korelsky, T. (1992, March). Applied text generation. In *Third conference on applied natural language processing* (pp. 40–47).

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. arXiv preprint arXiv:2204.06125.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese Bert-networks. arXiv preprint arXiv:1908.10084.

Reisenbichler, M., Reutterer, T., Schweidel, D., & Dan, D. (2022). Frontiers: Supporting content marketing with natural language generation. *Marketing Science*, *41*(3), 441–452.

Rocklage, M. D., & Fazio, R. H. (2020). The enhancing versus backfiring effects of positive emotion in consumer reviews. *Journal of Marketing Research*, *57*(2), 332–352.

Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The Evaluative Lexicon 2.0: The measurement of emotionality, extremity, and valence in language. *Behavior Research Methods*, *50*(4), 1327–1344.

Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2021). Mass-scale emotionality reveals human behaviour and marketplace success. *Nature Human Behaviour*, *5*(10), 1323–1329.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684–10695).

Rothe, S., Narayan, S., & Severyn, A. (2020). Leveraging pre-trained checkpoints for sequence generation tasks. *Transactions of the Association for Computational Linguistics*, *8*, 264–280.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., ... Manica, M. (2022). BLOOM: A 176B-parameter open-access multilingual language model. arXiv preprint arXiv:2211.05100.

Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A., & Kersting, K. (2022). Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nature Machine Intelligence*, *4*(3), 258–268.

Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., ... Jitsev, J. (2022). Laion-5b: An open large-scale dataset for training next generation image-text models. arXiv preprint arXiv:2210.08402.

Schweidel, D. A., & Moe, W. W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research*, *51*(4), 387–402.

Schweter, S., & Akbik, A. (2020). Flert: Document-level features for named entity recognition. arXiv preprint arXiv:2011.06993.

Sejnowski, T. J. (2020). The unreasonable effectiveness of deep learning in artificial intelligence. *Proceedings of the National Academy of Sciences*, *117*(48), 30033–30038.

Shahzad, M., Amin, A., Esteves, D., & Ngomo, A. C. N. (2021, April). InferNER: An attentive model leveraging the sentence-level information for named entity recognition in microblogs. In *The international FLAIRS conference proceedings* (Vol. 34).

Sheng, E., Chang, K. W., Natarajan, P., & Peng, N. (2019). The woman worked as a babysitter: On biases in language generation. arXiv preprint arXiv:1909.01326.

Simonov, A., & Rao, J. (2022). Demand for online news under government control: Evidence from Russia. *Journal of Political Economy*, *130*(2), 259–309.

Stone, P. J. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, MA: MIT Press.

Sun, Y., Wang, S., Feng, S., Ding, S., Pang, C., Shang, J., . . . Wang, H. (2021). Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. arXiv preprint arXiv:2107.02137.

Tang, T., Fang, E., & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing*, *78*(4), 41–58.

Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, *38*(1), 1–20.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, *31*(2), 198–215.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research*, *51*(4), 463–479.

Toubia, O., Berger, J., & Eliashberg, J. (2021). How quantifying the shape of stories predicts their success. *Proceedings of the National Academy of Sciences*, *118*(26).

Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2019). Extracting features of entertainment products: A guided latent Dirichlet allocation approach informed by the psychology of media consumption. *Journal of Marketing Research*, *56*(1), 18–36.

Toubia, O., & Netzer, O. (2017). Idea generation, creativity, and prototypicality. *Marketing Science*, *36*(1), 1–20.

Tunstall, L., Reimers, N., Jo, U. E. S., Bates, L., Korat, D., Wasserblat, M., & Pereg, O. (2022). Efficient few-shot learning without prompts. arXiv preprint arXiv:2209.11055.

Tunstall, L., von Werra, L., & Wolf, T. (2022). *Natural language processing with transformers*. Sebastopol, CA: O'Reilly Media, Inc.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Veitch, V., Sridhar, D., & Blei, D. (2020, August). Adapting text embeddings for causal inference. In *Conference on uncertainty in artificial intelligence* (pp. 919–928). PMLR.

Villarroel Ordenes, F., Grewal, D., Ludwig, S., Ruyter, K. D., Mahr, D., & Wetzels, M. (2019). Cutting through content clutter: How speech and image acts drive consumer sharing of social media brand messages. *Journal of Consumer Research*, *45*(5), 988–1012.

Wang, B., & Komatsuzaki, A. (2021). GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model. Retrieved from https://github.com/kingoflolz/mesh-transformer-jax

Wang, Y., Qin, M. S., Luo, X., & Kou, Y. (2022). How support for Black Lives Matter impacts consumer responses on social media. *Marketing Science*. In Press. doi:10.1287/mksc.2022.1372

Woolley, K., & Sharif, M. A. (2021). Incentives increase relative positivity of review content and enjoyment of review writing. *Journal of Marketing Research*, *58*(3), 539–558.

Yang, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences*, *112*(47), 14473–14478.

Yang, Y., Zhang, K., & Fan, Y. (2022). sDTM: A supervised Bayesian deep topic model for text analytics. *Information Systems Research*. In Press. doi:10.1287/isre.2022.1124

Yin, W., Hay, J., & Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. arXiv preprint arXiv:1909.00161.

Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., … Zettlemoyer, L. (2022). OPT: Open pre-trained transformer language models. arXiv preprint arXiv:2205.01068.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. W. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. arXiv preprint arXiv:1707.09457.

Zoph, B., Bello, I., Kumar, S., Du, N., Huang, Y., Dean, J., … Fedus, W. (2022). Designing effective sparse expert models. arXiv preprint arXiv:2202.08906.