Artificial Intelligence and User-generated Data are Transforming how Firms come to Understand Customer Needs

By John R. Hauser (MIT), Zelin Li (MIT), Chengfeng Mao (MIT)[1]

1. Introduction

The rapid growth of social media and online platforms provides users and customers abundant opportunities to express their opinions, thoughts, and feelings about companies and products. Digital footprints provide extremely valuable data with which to understand and manage customers' needs. While the explosion of such unstructured data being generated by customers can be daunting for traditional methods, the advancement in artificial intelligence (AI) and machine learning (ML) algorithms enables firms to process such data rapidly, efficiently, and effectively. In this chapter, we provide an overview of how artificial intelligence is transforming the ways that firms identify customer needs, structure customer needs for insight, and prioritize customer needs. This practice has come to be called the voice of customers (VOC). We aim to contribute to the marketing literature in the following ways. First, we summarize how VOC helps firms to gain insights on using user-generated data. Second, we discuss the types of user-generated data and how the challenges associated with analyzing each type of data. Third, we describe common methods, matched to the firms' goals and the structure of the data, that are used to analyze VOC. Fourth, and most importantly, we map the methods used to relevant applications, providing guidance on how to select the appropriate method of analysis to address the desired research questions.

2. Artificial Intelligence & the Voice of the Customer

2.1. Importance of the Voice of the Customer (VOC)

Understanding customer needs are essential to designing profitable products and selecting the right marketing strategies. The VOC provides a deep understanding of how customers and

[1] Names are listed alphabetically. All authors contributed equally.

potential customers will respond to changes in a product, changes in price, or changes in the way in which the firm communicates to customers. There are many proven examples,

- latent unmet customer needs identify valuable new directions for firms to explore when designing products
- “must-have” customer needs help a firm make sure that its products are considered by customers
- knowing customers’ implicit or explicit beliefs about the image of a brand improves brand management, especially if that image can be tracked over time or in response to marketing actions
- customers follow a purchase journey which might include becoming aware of a product and its message, searching for the product, buying the product, experiencing the product, and communicating to others, say via social media, about the product

Other applications include understanding a firm’s relative positioning in the market, identifying better customer segments, creating better branding based on impressions and perceptions and either reinforcing brand image or adjusting the brand image, developing improved content, learning consideration and thoughts throughout the purchase journey, targeting better, and boosting returns from advertising and other forms of persuasion.

2.2. VOC Practice before Artificial Intelligence (AI) and Big Data

Customer needs are deep and complex (often textual and sometimes visual). They are more than concrete features of a product. For example, a customer may want a smartphone screen that is easy to read while viewing text messages. Physical features such as the number of pixels, brightness, refresh rate, and screen dimensions affect the customers’ perception of “easy to read while viewing text messages,” but are not the same. Easy to read may be more nuanced, dependent upon use (text alone, text with emojis, text with pictures), dependent upon context (dark room, bright sunny day, at a glance or reading, the proportion of scene used), There may be other ways to fulfill the customer need such as better algorithms, colors, layout, or something not yet invented. Because firms need to understand nuance, content, use, and other aspects of customer needs, the traditional practice of identifying customer needs has focused on rich observation such as focus groups, one-on-one interviews, ethnography, or metaphor elicitation (Burchill and Brodie 1997; Calder 1977; Griffin and Hauser 1993; Gupta 2020; Vilà and Camps 2018; Zaltman 1997)

Customer needs can be numerous. It is not uncommon for customer needs for complex products to number in the hundreds. Industry practice is to structure customer needs into hierarchies and focus on the top level (primary) or the next level (secondary) of the hierarchy. The hierarchy is often obtained by clustering customer needs using a form of similarity as the distance measure. For example, customers might be asked to state similarities or to sort the customer's needs into piles of similar meanings. When such data are not available, customers or experts are asked to agree on a hierarchy, although using real customers is the preferred practice..

Priorities are measured with a variety of methods. Conjoint analysis is the most widely used, although often for physical features (Rao 2014), for more-complex customer needs stated importance measures have also proven effective (Griffin and Hauser 1993; Vonderembse and Van Fossen 1998). Firms can even let customers express their needs directly by inviting users to come up with their own new designs (Kaulio 1998).

2.3. The Promise of User-Generated Content (UGC)

Transaction data enable a firm to track sales, manage its customer communications, and evaluate the effect of marketing actions. But transaction data rarely provides a rich understanding of why the customer is acting as observed. On the other hand, primary data such as questionnaires, focus groups, one-on-one interviews, metaphor elicitation, and ethnographic observation provide rich understandings, but are expensive and time-consuming, and, by necessity, limited in scope.

UGC provides many benefits. It is often readily available by scraping various websites and feeds (with permissions of course). It is often high volume. For example, as of February 2022, there are 95 million images posted on Instagram every day, 500 million tweets on Twitter each day, and 250 million reviews now posted on Amazon. UGC is updated constantly and, hence, can be tracked over time. And, while any given product review may be relatively short, the set of all product reviews can be quite broad and contain rich complex information. (The same is true for images, tweets, and other forms of UGC.)

2.4. Challenges the Use of UGC

While UGC data have tremendous potential for understanding the VOC, there are many challenges. The sheer volume of customer-generated data makes it infeasible for humans to process all but a small sample. The data are unstructured and often too complex and high-dimensional to be analyzed with traditional methods such as regression. For example, a single image may contain a million pixels. Video is even higher-dimensional with 24 frames per second over many minutes and relationships among the frames. Even textual data is complex.

By its very nature, UGC is selected by the customers themselves. There is no guarantee that the corpus, or even a sample, is a representative set of customers. The selection process itself can create a bias as it is often hypothesized that only customers who feel strongly will post. (Fortunately, Timoshenko and Hauser (2019) provide data that self-selection does not seem to bias UGC, although the lack of a self-selection effect might be unique to their data or at least their goals.)

In summary, AI augments the analysis to examine *a priori* defined theoretical constructs or identifies hidden structures that can be interpreted post hoc. When the data are continuously updated by users, AI enables firms to automate the tracking of these data and the constructs and insights obtained from these data. Firms can timely monitor the new trends in customer needs and the competitive landscape so that they can promptly react to changes.

2.5. The Promise of AI and Machine Learning (ML)

In recent years, AI and ML have made many advances and marketing science has been on the cutting edge of using these methods to analyze UGC. We review a sampling of these applications in this summary. These applications use AI and ML to (1) convert large amounts of unstructured data to low dimensional constructs and (2) model the often complex interactions among the reduced constructs.

Among the advances we review are:

- the same customer needs can be identified in product reviews more quickly and with lower cost than from traditional methods

- AI augments humans to understand better the topics in which customers are interested and how their interests affect their actions in complex social networks
- AI provides a means to track brand perceptions from images posted by customers
- clusters of engineering characteristics in customer reviews provide insight on deeper customer needs,
- brand sentiment can be identified automatically,
- and many more.

3. Applications of AI-Powered VOC

3.1. Identifying and Organizing Customer Needs

Effective product development designs products that customers desire. To seek such profitable designs among the seemingly infinite variations of a given product or product line, one powerful approach is to break down the product into attributes and assess customers' preferences of different combinations of attributes. The high-dimensional design space can be reduced to the lower dimensions of attributes to facilitate an easier search for a design that matches customer desirability. This approach is at the core of conjoint analysis, where customers are presented with a set of products with variations in the same set of attributes so that the part-worth of each attribute can be estimated (Green, Krieger, and Wind 2001). Similarly, in Quality Function Deployment (QFD), the development of a new product starts with understanding customers' needs and mapping them to product attributes (Vonderembse and Van Fossen 1998). This attribute-based approach plays a central role when AI is applied to product development. The procedures and theoretical constructs remain the same, while AI enables the approach to achieve higher optimality and efficiency than that which was infeasible with traditional methods.

Machine learning algorithms can be seamlessly integrated into conjoint analysis to optimize the configurations of products presented to the respondents to achieve a more accurate assessment of customer preferences for complex products such as digital cameras (Huang and Luo 2016) whereas conventional conjoint analysis would require too many questions for the respondents to cover a large number of different product configurations. UGC, combined with AI, provides valuable data with which to identify customers. AI enables researchers to analyze the enormous qualities of UGC to comprehensively and accurately understand customer needs. Timoshenko and Hauser (2019) illustrate one approach. The most informative sentences are filtered from the UGC dataset with natural-language process methods (NLP) and evaluated by human analysts. To identify new user innovations, NLP focuses experts on the relatively small set of posts that contain user innovations (von Hippel and Kaulartz 2021).

AI further assists in organizing customer needs. In traditional VOC, either experts or customers manually aggregate customer demands. When customers are used, the surveys (customer need sorts) are expensive and time-consuming. Clustering algorithms use words and sentence embeddings to automatically group similar UGC content into clusters, each representing a unique facet of customer needs. Human analysts investigate a much smaller sample by choosing from the clusters. The end result is a broad-based sampling from the entire space of potential customer needs (Timoshenko and Hauser 2019). In other papers, hierarchical clustering automates semantically similar technical attributes. Careful analysis organizes further to form clusters representing more abstract customer needs (Wang et al. 2021).

3.2. Understanding and Forecasting Demand

Understanding customer needs and preferences is important to managerial implications in areas such as predicting sales (Wang et al. 2021), business success (Chakraborty, Kim, and Sudhir 2021), and managing social media engagement (Ma, Sun, and Zhang 2019). The challenge of these tasks with UGC is to identify the exact attributes of the product or service from the vast amount of content. In survey-based methods, attributes are based on VOC studies, product development teams, or domain experts. The free-form language in UGC requires more effort to identify the attributes from a large corpus. Wang et al. (2021) focus on a set of most frequent and informative noun phrases selected by heuristics and manually identify the product attributes. When the set of attributes is small and mostly known *a priori*, supervised machine learning has been applied to automatically identify the attributes. For instance, Chakraborty, Kim, and Sudhir (2021) focus on five attributes of restaurants such as food and service and use a supervised machine learning model trained on manually annotated samples to identify the customer reviews containing these attributes. Zhang and Luo (2021) take a similar approach to extracting restaurant attributes such as food or interiors from photos posted by customers. *A posteriori* human interpretation is useful when the attributes or needs are not known exactly or a more granular and open-ended analysis is preferred. Ma, Sun, and Zhang (2019) cluster similar images on social media and characterize their attributes with word clouds of user descriptions. Manual interpretation becomes feasible with such succinct representations of the small number of image clusters.

After customer needs or product attributes are identified successfully, the next step is to assess customers' perception of the attributes. Sentiment analysis is a common approach. Supervised machine learning models output binary labels of +1/-1 to indicate whether customers have a positive or negative experience with an attribute, or numerical values which indicate the strength (Chakraborty, Kim, and Sudhir 2021; Zhang and Luo 2021). Vermeer et al. (2019) suggest that sentiment is not the only relevant aspect. The context and content type of the UGC are also important, such as rejection, complaint, and suggestions.

It is often desired to further quantify the impact of the customers' perception of the attributes on business outcomes. One direct measurement is the importance of attribute perception in the prediction models complementing econometrics models used to explore the causal effects (Zhang and Luo 2021).

In the highly connected environment of the Internet, product features are not the only factor affecting customers' demands. Reviews posted by other customers have a significant impact (Chakraborty, Kim, and Sudhir 2021; Liu, Lee, and Srinivasan 2019). To quantify the impact of other customers' UGC requires identifying different theoretical constructs from those related to product attributes.

We summarize many of the methods in Table 1.

Table 1. Example AI Methods to identify and use Customer Needs and Attributes

	Data	Filtering	Product / service Attributes	Primary Needs	Perception of attributes	Importance of attributes
Traditional methods	Survey	N/A	manual identification		Respondents' answers	
Timoshenko and Hauser (2019)	UGC text	embeddings, supervised ML, clustering	manual identification		N/A	

von Hippel and Kaulartz (2021)	UGC text	embeddings, keyword matching	manual identification		N/A	
Wang et al. (2021)	UGC text	N/A	heuristics, manual identification	embedding. hierarchical clustering, post-hoc manual interpretation	sentiment analysis	N/A but can be obtained from the sales forecasting model
Chakraborty et al. (2021)	UGC text	N/A	manual identification		sentiment analysis	structural model
Zhang and Luo (2021)	UGC text + images	N/A	manual identification		sentiment analysis for image captions	feature importance + causal model
Ma (2019)	UGC text + images + social network	N/A	embedding, clustering, post-hoc manual interpretation		N/A	
Vermeer et al. (2019)	UGC text	N/A			context+ content type	N/A

3.3. Reflection of Branding Through Users

3.3.1. Brand Perception

Extracting brand perception is an important application. Although UGC is heterogeneous and varies across and within platforms and scenarios, UGC provides a corpus of user posts on social media and reviews in which users share their feedback, evaluation, and thoughts about brands. UGC and AI help brand managers to understand what perceptions, images, and personalities their brands have among users. These summaries help brands better manage their relationships with users, perform self-congruency tests on whether users' brand perceptions match the firm's intended image, and adjust branding strategies to best deliver the right images and products to the right users.

Traditionally, brand perceptions are measured with surveys, focus groups, or manual screening of users feedback. These traditional techniques are time consuming and labor-intensive, often resulting in limitations on sample size and the number of pre-specified attributes that are collected. Responses and content generated this way are not guaranteed to reflect the feelings users had at the moments of purchase and consumption. With the development of the Internet AI algorithms, these limitations are being addressed. Posts and contents are being generated in real-time, and contents in various kinds of formats are analyzed efficiently using AI to extract brand perceptions.

The literature augments the traditional techniques and extracts new types of information in diverse directions. For example, the utilization of machine learning algorithms in Dzyabura and Peres (2021) transform a well-known manual marketing research tool, metaphor elicitation technique (ZMET, Zaltman 1997), into a procedure that is automated and scaled easily. Liu, Dzyabura, and Mizik (2020) and Pamuksuz, Yun, and Humphreys (2021) use social media images and posts, respectively, to extract and evaluate the presence of brand-related attributes (glamorous, rugged, healthy, fun in Liu, Dzyabura, and Mizik; sincerity, excitement, competence, sophistication, and ruggedness in Pamuskuz, Yun, and Humphreys), Their methods investigate whether the brand perceptions are conveyed through UGC and match brands' intended images and personalities. Other methods enable brands to capture brand perceptions and images with a fuller set of contexts and users' emotions. Klostermann et al. (2018) take into account the heterogeneous online posting environment to understand brand perceptions in diverse domains and contexts. They use their algorithm to understand the strength and weaknesses of the products. Chakraborty, Kim, and Sudhir (2021) augment traditional sentiment analysis to capture the degree of positivity and negativity in users' perceptions of brands. These methods help firms understand the emotions associated with the diverse images and attributes of a brand. Wang et al. (2021) bridge the gap between consumer perceived attributes and engineering attributes so that brands might match perceptions and reactions to the right product design.

3.3.2. Brand Positioning

Positioning is a branding strategy by which to identify and deliver differentiated values to target customers. Positioning refers to the target perceptions of a brand relative to competitive brands in a particular segment of the market. For example, BMW is positioned as fun to drive and luxury in the premium market for sporty sedans. Different brands of related products coexist, and different segments of heterogeneous customers have distinctive product needs. Each brand needs to communicate its unique benefits to the target customer segment(s) with the right branding message(s).

Brand positioning identifies the market structure. Knowing a brand position helps the firm identify its perceptual location on a market map. The brand learns from its relationship to other brands and from positionings in adjacent markets. For example, Wang et al. (2021) develop an algorithm that allows the comparison of brand positions using seven meta-attributes as extracted from a large UGC corpus. Positioning helps brands to be perceived as different by the target segment, to be remembered and occupy a unique space in their consumers' perceptions, and to be purchased, hopefully repeatedly. As described, brand positioning is a complex, multi-task process that involves market segmentation, targeting, perception mapping, value identification (conceptualization), and strategy implementation.

Typically, survey-based perceptual maps, often augmented with eye-tracking, place brands in a perceptual space of relevant perceptual attributes, e.g., quality, personal service, value, and convenience for HMOs. Leveraging widely available UGC enriches the ability of the firm to understand the market structure construction that best represents the competitive market. The perceptions can be tracked over time, by segment, and how they evolve in the "wild." The continuous flow of UGC enables a firm to validate the implemented positioning strategy and gain insight into the causes of any change.

Ma, Sun, and Zhang (2019) use user clusters and image/post clusters learned through an embedded network. Tied to brand labels they find the average distance between each user cluster and brand. The closer a user cluster is to a brand, the more likely those users will pin images of that brand. Using this network of clusters enables brands to match with the right consumer segments and direct their efforts on this segment. Yang, Zhang, and Kannan (2021) build a visualization algorithm from users' engagement with brands' public fan pages that allows for real-time monitoring of the highly-fluid market structure. The tool enables brands to identify opportunities (e.g., collaboration with proximal brands in different industries) and threats (e.g., competitor proximal brands or outside brands that show potential threats) both inside and outside the defined industry boundaries. Brands can adjust their strategies to accommodate the market structure. AI is important to enable the methods to track the frequent user posts and monitor how fast the market is changing.

3.3.3. User-Brand Interaction

UGC is used to understand engagement between users and brands. Engagement can mean many different things under various contexts. Brands use UGC to learn how to drive user engagement or to predict future user engagement. The proliferation of the new, unstructured data types (e.g., images and videos) presents new opportunities to investigate how UGC drives user actions. Li and Xie (2020) find that the mere presence of images increases sharing. Characteristics of images (e.g., color variation, presence of a human face, professionalism of the image) have different effects on user engagement depending on contexts. Ma, Sun, and Zhang (2019) build a heterogeneous user-brand network using Pinterest data and use user interactions and curation actions to predict future engagement. Hartmann et al. (2021) propose a new concept called “brand selfies” where users post a selfie holding a branded product without actually showing their own face. They find that images from brand selfies are more effective for promoting brand engagement and result in higher likelihoods of likes and comments.

UGC does not have to be one-way from the customer to the brand; brands can engage with UGC to interact with users and maintain brand-user relationships. Vermeer et al. (2019) break the traditional stereotypes of comments with negative sentiments being the only worthy response and use AI algorithms to identify response-worthy reviews. With such systems, brands can better manage engagement with the users.

Other research focuses on brand profits testing whether UGC can predict well, has an effect on purchase intention, and, perhaps, an effect on sales. Besides the Hartmann et al. (2021) paper described above, Wang et al. (2021) find that the mean sentiments on the seven meta-attributes improve the prediction of sales rank. Liu, Lee, and Srinivasan (2019) constructed purchase journeys and found causal impacts of read-review content on sales. They found that UGC plays a larger role in final sales when products are more expensive, when the number of reviews is more modest, when products are in a more competitive market, and when the products are new. Branded products are, in some papers, less affected by reviews. In a different setting, Zhang and Luo (2021) find that user-generated restaurant photos in Yelp are strong predictors of whether a restaurant would survive. Photos are more predictive for the survival of independent restaurants compared to chain restaurants, for young and mid-aged restaurants compared to established restaurants, for medium-priced than for low-priced restaurants. The proportion of food photos has the largest positive association with restaurant survival, followed by proportions of outside and interior photos. With these insights, brands create campaigns that ask users to post food pictures on Yelp. Such strategies should boost sales.

3.4. Advertising, Persuasion, and Communication

The use of AI and UGC to address advertising, persuasion, and communication is a research area of great potential. Zhang and Luo (2021) hint at the ability of firms to use UGC actively. Interventions can be tracked, evaluated, and optimized with AI if suitable theories are developed.

4. Data Available for AI and VOC

4.1. Data Sources

4.1.1. Customer Reviews

Customer reviews are the most direct source of customer opinions on products and businesses. A wide range of constructs about customers can be extracted from the rich information in customer reviews, such as customers' levels of satisfaction about various attributes of a product or service (Chakraborty, Kim, and Sudhir 2021; Liu, Lee, and Srinivasan 2019; Zhang and Luo 2021), customers' needs (Timoshenko and Hauser 2019; Wang et al. 2021), or customer innovations (von Hippel and Kaulartz 2021).

Customer reviews also impact customer behaviors and business outcomes. The constructs extracted from customer reviews can aid further studies about such impact, including restaurant survival (Chakraborty, Kim, and Sudhir 2021; Zhang and Luo 2021), sales (Liu, Lee, and Srinivasan 2019), and many more which await exploration.

4.1.2. Social Media

Social media provides the perspective of customers sharing opinions and information with one another. Unlike customer reviews, most posts on social media are not directly related to a direct purchase experience. AI must filter relevant content about specific products. As a result, social media opinions are best for studying brand level perceptions (Klostermann et al. 2018; Liu, Dzyabura, and Mizik 2020; Ma, Sun, and Zhang 2019).

Brands themselves generate social media content. In turn, customers' posts contain responses to the brands' social media marketing content. Social media provides insight into brands' marketing strategies on social media, which can be compared to customer posts and the content of other brands. Researchers have studied many constructs in social media activities to understand and evaluate success. These constructs include customer satisfaction (Vermeer et al. 2019), user engagement (Li and Xie 2020; Ma, Sun, and Zhang 2019), congruence between brand image and customer perception (Liu, Dzyabura, and Mizik 2020; Pamuksuz, Yun, and Humphreys 2021) and purchase intent (Hartmann et al. 2021). Other constructs include brand image composition and qualities (Hartmann et al. 2021; Li and Xie 2020) or responses to customers' comments (Vermeer et al. 2019).

Added data such as likes, following, and commenting provide measures of brand similarity, which provide a view of social network structures indicating strong and weak ties, similarity among customers, and customer affinity groups. These measures are linked to managerial impact. For instance, Yang, Zhang, and Kannan (2021) identify the market structure and competition from brand-user interactions on social media. Ma, Sun, and Zhang (2019) extract brand positioning from similar types of information.

4.1.3. Direct Queries to Customers

AI has opened up new possibilities with direct customer queries that augment traditional surveys. AI-based queries can be used at scale. Dzyabura and Peres (2021) ask participants to create a collage of images to represent their perception of each given brand. Prior applications of such metaphor analysis were limited to in-person queries to a small sample of customers. AI further enables queries about more complex subjects, such as eliciting respondents' preferences about complex products with too many attributes for traditional survey research methods (Huang and Luo 2016).

Direct queries are most valuable when the researcher has a specific question to address or when UGC is not available, for example, in low-volume products and some B2B products. Direct queries overcome self-selection biases when they exist and can explore topics and constructs on which customers do not post. On the other hand, researchers incur costs and delays relative to UGC and panel studies when direct queries are difficult to manage. Direct queries to large samples often do not enable the in-depth and complex reactive queries possible with qualitative studies.

4.2. Data Types

4.2.1. Text Data

Prior to the advent of AI and UGC, text data were best analyzed with human readers and human judges. Because these processes were labor-intensive, and often involved experts, the analysis of text data was limited to small focused samples. In the past few years, many researchers have developed methods to make sense of large text corpora enabling analysis to be automated and done at scale. The most common form of text data are user product reviews (Chakraborty, Kim, and Sudhir 2021; von Hippel and Kaulartz 2021; Liu, Lee, and Srinivasan 2019; Timoshenko and Hauser 2019; Wang et al. 2021; Zhang and Luo 2021), user-created posts and content including but not limited to post content, post caption, and blogs (Hartmann et al. 2021; von Hippel and Kaulartz 2021; Klostermann et al. 2018; Li and Xie 2020; Ma, Sun, and Zhang 2019; Pamuksuz, Yun, and Humphreys 2021), user direct comments and feedback to brands (Vermeer et al. 2019), and responses elicited from users (Huang and Luo 2016). Researchers augment text data with tags (Klostermann et al. 2018), links, emojis, and other quantifying information.

Many relatively new methods have proven effective in the analysis of text data. Timoshenko and Hauser (2019) and von Hippel and Kaulartz (2021) extract customer needs and innovations through product reviews. Vermeer et al. (2019) use users' comments to brands to determine which comments are worthy of responding based on comments' content and sentiments. Besides these applications, attributes and other information are extracted from text data with the facilitation of AI algorithms. Pamuksuz, Yun, and Humphreys (2021) and Klostermann et al. (2018) extract brand information and personalities; Chakraborty, Kim, and Sudhir (2021)extract attributes on the degree of emotions; and Wang et al. (2021) automatically extract meta-attributes with unsupervised learning that requires no a priori assumptions.

4.2.2. Images

The utilization of visuals and images has always been essential to marketing and customer feedback. Internet and social media platforms enable and encourage users to create and share images about themselves, about products, and about opinions and metaphors. Like text data prior to AI, the analysis of images depended upon low sample tagging. Analyses were heavily dependent upon human judgment. Features of images were largely limited to predefined basic attributes such as color, shape, and texture. AI algorithms open a gateway to new ways of incorporating and understanding image data in marketing research. Images can be conveniently classified into different categories based on researchers' needs and features can be automatically extracted in a more expedited way.

With these tools at hand, researchers have been interested in analyzing how user-generated images impact firms. One direction of interest is how images affect user and brand engagement. Li and Xie (2020) investigate how different aspects of images (e.g., mere presence, image characteristics, and level of match between image and its caption) in tweets mentioning brands increase engagement metrics such as likes, comments, and retweets. Hartmann et al. (2021) study how brand selfies affect brand engagement. Other researchers investigate how user-generated images convey users' perceptions of the brands, either through naturally-created posts (Klostermann et al. 2018; Liu, Dzyabura, and Mizik 2020) or through direct elicitation, e.g., brand-perception collages elicited from users (Dzyabura and Peres 2021). Zhang and Luo (2021) use attributes of user-review photos in Yelp to predict the survival rate of restaurants. Dzyabura et al. (2022) use fashion-item images to understand the drivers of product returns. In most of these applications, researchers combine image data and text data to achieve their research objectives.

4.2.3. User Engagement

User engagement (e.g. likes, comments, shares) are essential metrics in measuring social media performance; researchers use user-engagement metrics to predict outcomes or as target variables (Hartmann et al. 2021). Engagement data often occur in complex networks that connect brands and users and provide insights on users' actions and behaviors. The challenge is in dealing with the large volume of user-engagement data. New methods combine AI and network analysis to better decipher and visualize networks generated by user-engagement data. Yang, Zhang, and Kannan (2021) use Facebook user-engagement data to create a brand-user network and use dimensionality reduction to visualize the brand positioning and generate market structure maps. Ma, Sun, and Zhang (2019) investigate Pinterest users' curation actions (e.g. ping / save a post or image) to identify user and brand clusters that can best facilitate brands to better understand the prospective audiences for firms' products.

5. Tools and Methods to Understand UGC

5.1. Data Preprocessing

Given the highly unstructured nature of the data from UGC, data preprocessing is necessary to format data in a way that can be used by the algorithms. UGC data are typically noisy without predefined features and labels and can be biased due to self-selection. Preliminary cleaning and filtering are essential. For example, Yang, Zhang, and Kannan (2021) use Facebook engagement data (e.g. likes, comments, shares) to create a brand-user network. However, given the existence of potential bots on Facebook, they filter their data by removing fake users using cues such as highly unusual interaction activities with huge numbers of brands. They further exclude users who posted duplicative comments with URL links under multiple posts. Li and Xie (2020) eliminate tweets that include hyperlinks to the original website. Such links artificially increase engagement rates. Wang et al. (2021) source their data from Amazon; they exclude reviews of Amazon Fire tablets to correct for Amazon's tendency to use algorithms that make Amazon products look more favorable. Such data screening requires judgment and should be explicit in any analysis so that the reader can properly evaluate potential biases in the underlying data or biases introduced by screening.

For user-generated text (e.g. reviews, comments, captions), standard text preprocessing procedures should be used. These procedures include but are not limited to excluding stop words, numbers, and punctuations(Chakraborty, Kim, and Sudhir 2021; Timoshenko and Hauser 2019; Wang et al. 2021), dropping meaningless phrases and infrequent words, transforming emojis into standard emoticons, and retaining only text from a target language, say English (Klostermann et al. 2018). Although these procedures have become accepted practice, improved precision is not guaranteed, Vermeer et al. (2019) show that original text, without any processing, can generate better performance in some scenarios.

Researchers often seek to generate features from text content. Sentiment classification is used frequently as a preprocessing step so that researchers can represent the sentiments of sentences as a feature for their algorithms (Chakraborty, Kim, and Sudhir 2021; Klostermann et al. 2018; Li and Xie 2020; Vermeer et al. 2019; Zhang and Luo 2021). Other examples include extracting predefined attributes and topics and detecting the presence of linguistic features by relying on external dictionaries such as the Linguistic Inquiry and Word Count dictionary (Chakraborty, Kim, and Sudhir 2021; Li and Xie 2020).

When preprocessing images, image tagging is used frequently. Researchers seek to label images in their data with categories that best capture features in the images. This can be done either with a pre-existing set of labels (Hartmann et al. 2021; Ma, Sun, and Zhang 2019; Zhang and Luo 2021) or through an off-the-shelf service that automatically generates labels (Dzyabura and Peres 2021; Klostermann et al. 2018; Li and Xie 2020; Zhang and Luo 2021). Researchers might be interested in using cinema graphic features in pictures, such as pixels, length, width, brightness, color, and composition. These features become part of the feature collection set in the data preprocessing step (Li and Xie 2020; Ma, Sun, and Zhang 2019; Zhang and Luo 2021).

Researchers may find it useful to transform data to make it more usable than it would be in its natural format. For example, Hartmann et al. (2021) augment their data with expanded training data – mirroring images horizontally and selectively over-sampling comments that express purchase intentions. Their goal is to address class imbalance issues. Liu, Lee, and Srinivasan (2019) transform review and purchase data into purchase journeys to better capture users' previous purchase experiences.

5.2. Unsupervised Learning

Unsupervised learning is a popular family of techniques that extract innate patterns from data without the need for labeled data. Unsupervised methods are quite useful as a means to project high-dimensional unstructured data onto a much smaller vector space or to summarize a large amount of data with a much smaller set of patterns.

5.2.1. Embeddings

Embedding is a foundational technique for converting unstructured data into numerical vectors of fixed size to facilitate more efficient computation. Language embedding is now a standard initial step for applying deep learning models in NLP tasks. Word2Vec (Mikolov et al. 2013) is the most popular language embedding algorithm in the papers we reviewed. Word2Vec embeds words into a vector space where words of similar semantic meaning are also proximate to each other in vector-space distance. This representation is simple yet powerful because it captures semantic similarity and contextual information of the words. Von Hippel and Kaulartz (2021) find posts containing user innovation ideas with Word2Vec which otherwise could not be found with exact keyword matching. Word2Vec combined with deep learning methods such as long short-term memory (LSTM) can often outperform lexicon methods (Chakraborty, Kim, and Sudhir 2021). Word embeddings from Word2Vec are widely used in NLP tasks, such as classification with CNN (Timoshenko and Hauser, 2019) or with convolutional-LSTM (Chakraborty, Kim, and Sudhir 2021). More recent Transformer-based models such as BERT (Devlin et al. 2018) and RoBERTa (Liu et al. 2019) are equipped with further improved subword embeddings, which capture more nuances such as changes in word meaning in different contexts or order. The improved representations contribute to the improved performance of RoBERTa in the text classification tasks of Hartmann et al. (2021)and Pamuksuz, Yun, and Humphreys (2021).

Embedding is not limited to words. Pamuksuz, Yun, and Humphreys (2021) embed whole documents as vectors with Doc2Vec (Le and Mikolov 2014), Timoshenko and Hauser (2019) embed complete sentences, Ma, Sun, and Zhang (2019) embed users, images, and annotations in a social network into the same vector space with metapath2vec++ (Dong, Chawla, and Swami 2017) and Burnap, Hauser, and Timoshenko (2021) embed images into lower-dimensional spaces. Because of the versatility, efficiency, and rich information preserved by embedding methods, these methods have been adopted for learning the representations in various supervised or unsupervised learning models.

5.2.2. Clustering

Clustering methods group data into distinct clusters while maximizing the similarity within the clusters. The resulting number of clusters is smaller than the original dataset. Clustering has been used in marketing for decades. For example, it is common to cluster demographics, psychographics, or other consumer descriptors to identify consumer segments. See Green et al. (1973) among many others. One enduring application of clustering is the use of Ward’s algorithm to cluster customer needs based on co-occurrence within sorted piles of customer needs (Griffin and Hauser 1993).

More recently, Timoshenko and Hauser (2019) summarize customer reviews, and Ma, Sun, and Zhang (2019) identify user interest groups and image types with clustering. Clusters are often grouped in a tree-like form with lower-level clusters grouped to form higher-level clusters via hierarchical clustering. Wang et al. (2021) use this method to group product attributes into more abstract customer needs. Because similarity in clustering is measured by a distance metric, the representation of the data is crucial to the effectiveness of clustering. The data must be mapped into vector space that preserves the semantic and structural similarity. Therefore, clustering is commonly applied to the vector space of the aforementioned embedding methods. Word2Vec is a popular choice (Chakraborty, Kim, and Sudhir 2021; Klostermann et al. 2018; Timoshenko and Hauser 2019; Wang et al. 2021; Yang, Zhang, and Kannan 2021). The network embedding applied by Ma, Sun, and Zhang (2019) maps users with similar interests in proximity with each other to identify user interest groups.

5.2.3. Topic Modeling

Topic modeling seeks to find a set of topics that can best represent all the documents in the data. Each topic can be interpreted by the words most likely to appear in that topic, and the topic composition of each document is estimated from the likelihood of its words belonging to each topic. Latent Dirichlet Allocation (LDA, Blei, Ng, and Jordan 2003) is a popular topic modeling algorithm. It is often used for summarizing texts as a mixture of topics (Dzyabura and Peres 2021; Zhang and Luo 2021). While clustering is often used for summarizing texts, it only allows one data point to belong to one cluster. In contrast, topic modeling represents a document as a mixture of topics, which captures the complexity of language better. LDA is normally applied with a bag-of-word representation of texts, but it can also be combined with Word2Vec as in LDA2Vec (Das, Zaheer, and Dyer 2015) to achieve better performance. There are a few challenges with LDA. One challenge is that common words such as “the,” “and,” and others end up in the bags of words. To address such a challenge, Hsu and Glass (2006) extends LDA to capture first-order Markov relationships among the topics. With HMM LDA (hidden Markov models LDA), words such as “the” cluster into the first topic, and can be ignored. The other challenge with LDA and similar topic models is that the bags of words are often hard to interpret and often contain many constructs. See for example, Büschken and Allenby (2016).

5.3. Prediction-Based Algorithms (Supervised Learning)

Prediction-based machine-learning algorithms are one of the most common types of applications in marketing research. Traditional statistical-learning methods, such as support vector machines (SVM) and XGBoost, are used actively and improved by researchers for predictive tasks. Huang and Luo (2016) use a fuzzy SVM, a weighted variant of a soft-margin SVM, to improve conjoint analysis by adaptively selecting subsequent questions based on previous answers from both the respondent and the other prior participants. Vermeer et al. (2019) use an SVM to decide which electronic Word-of-Mouth (eWOM) is relevant and/or requires an urgent response by brands managing customer relationships. Li and Xie (2020) use an SVM to detect behavioral drivers (self-enhancement, information provision), sentiments, and topics that are present in user-generated texts. Wang et al. (2021) use two SVMs to conduct sentiment analysis by extracting the presence of positive and negative sentiments separately. Zhang and Luo (2021) use XGBoost as their predictive algorithm to perform one-year-ahead predictions on restaurant survival based on features generated from photos posted on Yelp. These traditional statistical-learning methods are easy to implement and often provide transparency on the contributions of different factors within the prediction process.

Traditional machine-learning methods often lack the ability to use unstructured data directly. Fortunately, recent advances in machine learning algorithms such as embedding enable researchers to train models directly based on unstructured data such as user-generated text and images. The flexibility, adaptive ability, and the capability of modeling complex, non-linear relationships among features make deep-learning algorithms ideally suited to learn relationships within the data and deliver better predictive performance. Deep-learning algorithms often do better on accuracy and other metrics such as precision and recall. For example, convolutional neural networks (CNN) are best used when analyzing images because the convolutional filters enable CNN to capture spatial structures of the images; recurrent neural networks (RNN), on the other hand, work well with text data. RNN algorithms, such as LSTM and gated recurrent unit methods (GRU), have the ability to capture sequential data.

Most recently, the emergence of Transformer-based deep learning models has become an exciting new development in NLP. Transformer-based models gain an additional level of flexibility on top of RNNs by allowing the algorithm to attend to any order it finds most useful in accomplishing the task. The Transformer-based models do not have to process an input text sequence from start to finish. Instead, the models focus on searching for contextual information that delivers meaning to each word in the sequence. This ability enables efficient parallelization of the training process to greatly reduce training time.

Although off-the-shelf algorithms are often quite powerful and well-suited to the marketing tasks, researchers have improved performance by modifying existing machine-learning / deep-learning algorithms to best accomplish the prediction tasks. Timoshenko and Hauser (2019) build a CNN to identify informative content from online text reviews. Liu, Lee, and Srinivasan (2019), Liu, Dzyabura, and Mizik (2020), Zhang and Luo (2021) build their own versions of multi-task CNN to predict the level of presence of predefined attributes within the user-generated images. Hartmann et al. (2021) use a VGG-16 algorithm (pretrained on 1.2 million images on ImageNet) and add layers to the CNN to enable reliable classification of image types. Similar adaptations have also been used in the NLP context. Chakraborty, Kim, and Sudhir (2021) combine convolutional models with recurrent neural networks (convolutional-LSTM) to account for language structure and predict sentiment and the presence of restaurant attributes.

5.4 Hybrid of Unsupervised and Supervised Learning

Supervised learning can make accurate predictions but its training requires costly labeled data. The hybrid models use unsupervised learning to identify innate patterns from more readily available unlabeled data to help supervised learning achieve better performance. A common approach of the hybrid model is to pre-train the representation learning part with unsupervised learning and then fine-tune the whole model on the prediction task with supervised learning. Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018) is a Transformer-based language model that follows this approach. During pre-training, the algorithm learns a representation that can best predict each token based on both the left and the right context, and also predict the next sentence. The representation is then used for prediction and further optimized during the fine-tuning phase. Many models have then been developed on top of BERT, such as distilBERT (a light version of BERT, Sanh et al. 2019) and RoBERTa (a robust pretrained algorithm that can be used off-the-shelf, Liu et al. 2019). Pamuksuz, Yun, and Humphreys (2021) and Hartmann et al. (2021) both fine-tune a pre-trained RoBERTa model to predict predefined attributes (brand personalities and purchase intention, respectively). These models are widely available and pretrained on corpora with billions of words. Researchers can use them to achieve their objectives with relatively high performance to facilitate research in the field of marketing. Many large technology companies have provided APIs and services with state-of-the-art deep learning models built in. These off-the-shelf APIs and models have effectively extracted interpretable features (Dzyabura and Peres 2021) and label detection (Klostermann et al. 2018; Li and Xie 2020).

Supervised and unsupervised learning components in a single model can be jointly trained. Burnap, Hauser, and Timoshenko (2021) apply variational autoencoders (VAEs) to encode the automotive images as embeddings and generate images from the embeddings. VAEs are trained in an unsupervised manner to recreate the original images and generate new images from the embeddings. In order to generate realistic images, adversarial learning is applied to train the encoder to distinguish the generated images from the original images and the generative model to generate images indistinguishable from the original ones. The embeddings in the VAE with adversarial training are jointly optimized using supervised and unsupervised learning to predict the aesthetic appeal of the automobile images and to generate new images that customers evaluate as having high aesthetic appeal.

5.5. Model interpretation

Applying machine-learning algorithms to data is just the start of AI-powered VOC. Managerial impact and/or theoretical contribution requires researchers to accurately communicate their findings from the algorithms in succinct language. In this section, we discuss the roles and purposes of model interpretation.

5.5.1 Model Interpretation with Manual Inspection and Data Transformation

Often, unsupervised learning methods enhance interpretability, especially with textual data. The low-dimensional output from these models allows manual inspection and intuitive interpretation. Topics discovered by topic modeling can be represented by a list of words most likely associated with the topics (Dzyabura and Peres 2021; Zhang and Luo 2021). Clusters can be represented by examples from each cluster (Timoshenko and Hauser 2019) or post-hoc manual analysis (Wang et al. 2021). The texts from clusters can be further summarized into word clouds to facilitate a more intuitive interpretation (Ma, Sun, and Zhang 2019). The caveat is that some of these interpretations are like a Rorschach inkblot test; the human mind can often see patterns in topics or clusters that may not exist. At a minimum, researchers should test reliability, that is, do they get the same interpretations from random splits of the data?

For more complex data formats, data transformation is a common strategy. Images can be converted to texts and more easily summarized. Dzyabura and Peres (2021) and Zhang and Luo (2021) obtain textual descriptions about the content of the images and then further summarize the texts as topics with LDA. Ma, Sun, and Zhang (2019) characterize the images and users in a social network with texts associated with the images. More complex data can be visualized as images. Klostermann et al. (2018) visualize the image clusters with network graphs where nodes are image tags and edges are scaled co-occurrence frequencies. Yang, Zhang, and Kannan (2021) project the high-dimensional representations of brands in a two-dimensional space with t-SNE (Van der Maaten and Hinton 2008) to visualize the market structure.

5.5.2. Post-hoc Model Explanation

The black-box nature of deep-learning models makes inspection of internal mechanisms challenging. In response, explainable AI methods help researchers better understand the prediction mechanism of complex machine learning models (Gilpin et al. 2018; Guidotti et al. 2018). For instance, Hartmann et al. (2021) apply gradient-weighted class activation mapping (Grad-CAMs; Selvaraju et al. 2017) to understand how their CNN model classifies images. Saliency masks from Grad-CAMs highlight the most important areas in the images for the prediction. For example, Burnap, Hauser, and Timoshenko (2022) use masks to encode the "automotiveness" of the images. Hartmann et al. (2021) apply Local Interpretable Model-agnostic Explanations (LIME; Ribeiro, Singh, and Guestrin 2016) to highlight the words that are most predictive in classifying user comments with the RoBERTa model.

Recently, researchers have begun to use methods that represent the marginal effect of features in deep-learning models for model explanation. SHAP values (Shapley Additive exPlanations, Lundberg et al. 2020) draw on Shapley values from game theory to compute the marginal effect of a feature averaged over all possible values of other features. (This is often approximated by the set of observed items in the data.) For example, Zhang and Luo (2021) use SHAP values to quantify the importance of photos and reviews in predicting restaurant survival in their XGBoost model. Dzyabura et al. (2022) use SHAP values to quantify the marginal impact of their interpretable features. They combine SHAP values with Pareto charts from quality management to highlight the features with the greatest leverage. Because SHAP values average over the feature sets in the data, Dzyabura et al. (2022) use bee-swarm charts to indicate the leverage and values of each individual data point. The bee-swarm charts enable those who use the model to zero in on particular instances to understand the impact of a feature. SHAP values are correlational rather than causal, and best used to form hypotheses and to interpret the impact of machine-learned features.

5.6. Evaluation

Evaluation of AI methods is important. Researchers need to understand whether their proposed models address their research questions in a generalizable manner that is not tied to the estimation/calibration data.

In the case of prediction machine learning algorithms, researchers often split their labeled dataset into training and test datasets (Chakraborty, Kim, and Sudhir 2021; Hartmann et al. 2021; Liu, Dzyabura, and Mizik 2020; Timoshenko and Hauser 2019) so that algorithms do not overfit the original dataset. Usually, performance on the training dataset is validated using cross-validation. The ideal scenario is when researchers have ground truth and thus can justify their proposed models' capabilities using widely accepted metrics. Examples of frequently mentioned metrics are accuracy, F1-score, recall, and precision, used by researchers to evaluate their predictive models against an externally observed outcome.

With a single model, it is difficult to determine how good is good. For example, consider the classical example from transportation demand modeling. If we know that 90% of the respondents drive and 10% use public transit, then a model predicting everyone will drive will have a hit rate of 90%. Randomly predicting in proportion to market shares with give an 82% hit rate. A complex model with a 75% hit rate sounds good, but not compared to these naïve models.

It is critical that performance be compared to both naïve models and the best-available baseline models. Such baseline models can be traditional statistical learning models (e.g., linear regression, SVM, LDA), lexicon-based methods in the context of the natural language processing (Chakraborty, Kim, and Sudhir 2021), qualitative methods (Dzyabura and Peres 2021; Huang and Luo 2016; Wang et al. 2021), or alternative deep-learning models (Timoshenko and Hauser 2019). Researchers must be careful that the statistics for their proposed model are inflated relative to viable alternatives because researchers know "their" model best. It is imperative that researchers devote sufficient tuning to alternative models so that the comparison is fair. For example, Burnap, Hauser, and Timoshenko (2021) spend substantial time tuning the best-available pre-trained deep-learning model so that any increment from their custom deep-learning model can be fairly evaluated.

6. Mapping Methods to Research Questions

AI has been applied to a wide range of research questions within the general topic of the voice of customer research. Research has proven valuable to identify constructs and test hypotheses. It has also proven valuable to uncover customer needs, structure those customer needs, tie those customer needs to branding, and predict outcomes.

6.1. *A Posteriori* Identified Phenomena and Constructs

AI has been successful at identifying phenomena and constructs in a data-driven manner. The simple maxim, “we don’t know what we don’t know”, often applies. Breakthroughs often come when we identify new directions, new phenomena, or new constructs. AI is valuable, particularly unstructured methods, to find regularities in data and identify the constructs that might be mediators or moderators of causal relationships (Shadish et al. 2002). AI methods help distill interesting constructs and phenomena from a large amount of UGC. For example, von Hippel and Kaulartz (2021) find twenty-six high-quality user innovations from more than 200,000 online posts. With further analysis, these user innovations might be categorized to help understand the basis of innovation. Wang et al. (2021) filter out sentences containing engineering attributes from over 700,000 review sentences. These engineering attributes might be categorized further to understand how such engineering attributes map to customer needs.

AI enables efficient discoveries of regularities from large amounts of unstructured data. Brand perception is an example where organically-discovered patterns of customer opinions update researchers' *a priori* assumptions. Traditional methods rely on human interpretation to interpret the free-form responses from customer surveys, but deep learning enables the automatic extraction of abstract patterns from customers' responses and UGC. Unsupervised learning organizes the patterns in an easy to interpret manner (Dzyabura and Peres 2021; Klostermann et al. 2018; Liu, Dzyabura, and Mizik 2020). Similar strategies are also used for understanding customer needs (Timoshenko and Hauser 2019; Wang et al. 2021).

AI can operationalize construct identification from the complex data structures, and produce succinct construct representation that captures essential information in the high volume raw data. These constructs would otherwise be difficult to define by language or mathematical formula. For instance, Ma, Sun, and Zhang (2019) represent a large heterogeneous social network consisting of users, images, and image posting activities with network embeddings, which capture user similarities and image similarities based on their connections in the social network. This representation is the foundation for defining the construct of user interest groups. The ease of operationalization of construct identification in this process facilitates the discovery of empirical generalizations (Bass 1995) and thus prepares researchers for developing theories with better generalizability.

6.2. *A Priori* Defined Constructs

Researchers can often propose hypotheses inspired by extant literatures or manual inspection of the data. To validate the hypotheses, they first need to identify the constructs defined a priori in the theories. AI can automate the extraction of these constructs from unstructured data that would otherwise be cumbersome and costly to identify. The simplest example could be sentiment analysis, where the goal of the algorithm is to detect the positive or negative sentiment with its degree in customer's sentences. In image processing scenarios, this could mean using convolutional neural networks (CNN) to label the pictures so that researchers can use these labels for their research questions. After attributes and constructs are identified and extracted, AI facilitates the development of predictive models which can be used to validate hypotheses originated from the marketing literature, business practice, or researchers' own observations. Liu, Dzyabura, and Mizik (2020) identified brand personalities on social media according to the definitions in existing theories and examined the portrayal and perception of the brand personalities. Pamuksuz, Yun, and Humphreys (2021) observed the taxonomy of three brand image types and categorized the images with CNN accordingly. They further examined the impact of brand image types on purchase intention. Other examples include predicting fashion-item returns, the aesthetics of an image, and sales.

There is a growing shift in how predictive machine learning models are used in marketing. Initially, predictive ML algorithms were used primarily to predict an outcome, say sales, even though the model was hard to interpret. Such models were evaluated by their ability to forecast an outcome based on the data. Huang and Luo (2016) proposed using fuzzy SVM to select questions for conjoint analysis. Liu, Dzyabura, and Mizik (2020) use CNNs as the main part of the proposed model to classify the level of presence of predefined attributes (rugged, fun, glamorous, healthy) in images and average across all images of a brand to extract brand perception. More recently, predictive models have become part of a larger goal in which the research seeks to understand and interpret the data to create new products, strategies, or communications. These applications include predefined constructs or constructs uncovered from the data with unsupervised methods. For example, researchers might label the data with independently defined features of interest and manually label training data using platforms such as Amazon MTurk. Prediction algorithms (e.g. RNN / Transformer-based model for language tasks and CNN for image processing tasks) identify the best transformations of the data (interactions, non-linearities) to provide strong predictive performance, while being careful not to exploit noise in the data. Methods such as SHAP values then identify the marginal impact of features in predicting the outcome of interest.

Researchers also use fully labeled data as their primary dataset to validate hypotheses of interest. AI algorithms are evolving. New methods are appearing to enable better predictions, better structuring of data, stronger hypothesis testing, and improved generality. The new methods, combined with the proven applications of existing methods means that AI will continue to grow as a means to understand and use the voice of the customer.

6.3. Validation

One challenge to the use of AI is construct validity: does AI actually find the constructs that the researchers intend to identify? Human judgment is still a gold standard in verifying construct validity, such as expert reviews or comparison with prior expert opinions (Timoshenko and Hauser, 2019; von Hippel and Kaulartz, 2021; Wang et al., 2021). Model interpretation can also be used as robustness checks for construct validity. For instance, Hartmann et al. (2021) classify brand images and texts with deep-learning models. Because the brand image categories and purchase intention in text are the central constructs in their work, they also apply post-hoc interpretation methods to highlight the features in the images and texts that are most important for identifying their target constructs. The visualization clearly confirms that the deep learning models correctly identify the features relevant to the predefined constructs.

An equally important question is external validity: how effective are the constructs in explaining and predicting phenomena of interest and do the constructs and the nomological network generalize to other domains? Prediction tasks are often used to assess external validity, but prediction can be unique to a data set or, without careful modeling, exploit random variation. It is incumbent upon researchers that the answer applies to the data, and especially to new situations not tied to the data that were used to develop the model. The ideal comparison is ground truth. For example, can predictions of purchase behavior be compared to actual purchase behavior. Without ground truth, AI methods are in danger of over-interpretation. Similarly, it is tempting to try many alternative methods on the same data set in a tournament to determine the best method for the data. While useful, such tournament research risks exploiting random variation in the data to find a model which is best for the data at hand, but may not be best for new data. Occam's razor, complexity control, regularization, hold-out testing, cross-data testing, and many other methods assure that the model is a true representation of the data.

Researchers are also often interested in measuring the causal validity or the nomological validity of their proposed constructs. Some researchers use post-hoc causal inference models to identify the causal effects while others use lab controlled experiments to make results more ascertained. Hartmann et al. (2021) use both econometric modeling and lab experiments to validate whether there is indeed a causal impact of brand selfies on brand engagements and purchase intention. Li and Xie (2020) measure the causal impact of text features and posting features on consumers' attention to posts. Liu, Lee, and Srinivasan (2019) construct consumer purchase journeys and estimate the causal impact of review reading behaviors. Zhang and Luo (2021) train a cluster-robust causal forest (Athey and Wager 2019) to obtain consistent estimates of the causal effect of Yelp review photos on restaurant survival.

AI is extremely powerful for research into the voice of the customer, but researchers must be careful so that it is not misused.

7. Future Directions

AI for VOC has been valuable, and the future is rich.

Researchers have used text and image UGC effectively, but video and audio data are substantially higher-dimensional and contain interrelations worth exploration. For example, a video file is not a succession of independent frames but rather tells a story through both audio and visual means with story arcs, characterization, and metaphors. AI is beginning to be applied to such data (Toubia, Berger, and Eliashberg 2021).

New methods are being developed and older methods improved. We have mentioned CNNs, SVMs, LDA, HMM LDA, LSTM, boosted trees, VAEs, transformer-based models, adversarial training, embeddings, clustering, SHAP values, Pareto charts, bee-swarm charts, and other methods. These just scratch the surface of what is now available, not to mention that which will be discovered in the next few years. Problems and issues that now seem difficult to solve are likely solvable in the near future.

Finally, most AI VOC applications have focused on extracting customer needs, engineering characteristics, user-discovered innovations, and brand perceptions. Such methods are likely to be applied to advertising, persuasion, and communication as they relate to VOC customer needs. And, of course, AI is likely to improve applications that we have yet to identify.

References

Athey, Susan and Stefan Wager (2019), “Estimating Treatment Effects with Causal Forests: An Application,” *Observational Studies*, 5 (2), 37–51.

Bass, Frank M. (1995), “Empirical Generalizations and Marketing Science: A Personal View,” *Marketing Science*, 14 (3), G6–19.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), “Latent Dirichlet allocation,” *Journal of machine learning research: JMLR*, 3 (null), 993–1022.

Burchill, Gary and Christina Hepner Brodie (1997), *Voices Into Choices: Acting on the Voice of the Customer*, Oriel Incorporated.

Burnap, Alex, John R. Hauser, and Artem Timoshenko (2021), “Design and Evaluation of Product Aesthetics: A Human-Machine Hybrid Approach.”

Büschken, Joachim and Greg M. Allenby (2016), “Sentence-Based Text Analysis for Customer Reviews,” *Marketing Science*, 35 (6), 953–75.

Calder, Bobby J. (1977), “Focus Groups and the Nature of Qualitative Marketing Research,” *JMR, Journal of marketing research*, 14 (3), 353–64.

Chakraborty, Ishita, Minkyung Kim, and K. Sudhir (2021), “Attribute Sentiment Scoring with Online Text Reviews: Accounting for Language Structure and Missing Attributes,” *JMR, Journal of marketing research*, 00222437211052500.

Das, Rajarshi, Manzil Zaheer, and Chris Dyer (2015), “Gaussian LDA for topic models with word embeddings,” in *Proceedings of the 53rd Annual Meeting of the Association for*

Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), aclanthology.org, 795–804.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018), "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv [cs.CL]*.

Dong, Yuxiao, Nitesh V. Chawla, and Ananthram Swami (2017), "metapath2vec: Scalable Representation Learning for Heterogeneous Networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, New York, NY, USA: Association for Computing Machinery, 135–44.

Dzyabura, Daria, Siham El Kihal, John R. Hauser, and Marat Ibragimov (2022), "Leveraging the Power of Images in Managing Product Return Rates."

Dzyabura, Daria and Renana Peres (2021), "Visual Elicitation of Brand Perception," *Journal of marketing*, 85 (4), 44–66.

Gilpin, Leilani H., David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal (2018), "Explaining Explanations: An Overview of Interpretability of Machine Learning," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89.

Green, Paul E., Wind, and Yoram (1973), *Multiattribute decisions in marketing: a measurement approach*, Hinsdale, Ill.: Dryden Press.

Green, Paul E., Abba M. Krieger, and Yoram (jerry) Wind (2001), "Thirty Years of Conjoint Analysis: Reflections and Prospects," *Interfaces*, 31 (3), S56–73.

Griffin, Abbie and John R. Hauser (1993), "The Voice of the Customer," *Marketing Science*, 12 (1), 1–27.

Guidotti, Riccardo, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi (2018), "A Survey of Methods for Explaining Black Box Models," *ACM Comput. Surv.*, 51 (5), 1–42.

Gupta, Sunil (2020), "Are You Really Innovating Around Your Customers' Needs?," *Harvard Business Review*.

Hartmann, Jochen, Mark Heitmann, Christina Schamp, and Oded Netzer (2021), "The Power of Brand Selfies," *JMR, Journal of marketing research*, 00222437211037258.

von Hippel, Eric and Sandro Kaulartz (2021), "Next-generation consumer innovation search: Identifying early-stage need-solution pairs on the web," *Research policy*, Treading New Ground in Household Sector Innovation Research: Scope, Emergence, Business Implications, and Diffusion, 50 (8), 104056.

Hsu, Bo-June (Paul) and James Glass (2006), “Style & topic language model adaptation using HMM-LDA,” in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP ’06, USA: Association for Computational Linguistics, 373–81.

Huang, Dongling and Lan Luo (2016), “Consumer Preference Elicitation of Complex Products Using Fuzzy Support Vector Machine Active Learning,” *Marketing Science*, 35 (3), 445–64.

Kaulio, M. A. (1998), “Customer, consumer and user involvement in product development: A framework and a review of selected methods,” *Total Quality Management & Business Excellence*, 9 (1), 141–49.

Klostermann, Jan, Anja Plumeyer, Daniel Böger, and Reinhold Decker (2018), “Extracting brand information from social networks: Integrating image, text, and social tagging data,” *International Journal of Research in Marketing*, 35 (4), 538–56.

Le, Quoc and Tomas Mikolov (22--24 Jun 2014), “Distributed Representations of Sentences and Documents,” in *Proceedings of the 31st International Conference on Machine Learning*, Proceedings of Machine Learning Research, E. P. Xing and T. Jebara, eds., Beijing, China: PMLR, 1188–96.

Li, Yiyi and Ying Xie (2020), “Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement,” *JMR, Journal of marketing research*, 57 (1), 1–19.

Liu, Liu, Daria Dzyabura, and Natalie Mizik (2020), “Visual Listening In: Extracting Brand Image Portrayed on Social Media,” *Marketing Science*, 39 (4), 669–86.

Liu, Xiao, Dokyun Lee, and Kannan Srinivasan (2019), “Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning,” *JMR, Journal of marketing research*, 56 (6), 918–43.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (2019), “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” *arXiv [cs.CL]*.

Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee (2020), “From Local Explanations to Global Understanding with Explainable AI for Trees,” *Nature machine intelligence*, 2 (1), 56–67.

Ma, Liye, Baohong Sun, and Kunpeng Zhang (2019), “Image network and interest group – A heterogeneous network embedding approach to analyze social curation on Pinterest,” (accessed November 4, 2021), [available at https://www-2.rotman.utoronto.ca/userfiles/seminars/marketing/files/PinterestNetworkEmbedding_MaSunZhang_2019.pdf].

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013), "Efficient Estimation of Word Representations in Vector Space," *arXiv [cs.CL]*.

Pamuksuz, Utku, Joseph T. Yun, and Ashlee Humphreys (2021), "A Brand-New Look at You: Predicting Brand Personality in Social Media Networks with Machine Learning," *Journal of Interactive Marketing*, 56, 55–69.

Rao, Vithala R. (2014), *Applied Conjoint Analysis*, Springer, Berlin, Heidelberg.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016), "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, New York, NY, USA: Association for Computing Machinery, 1135–44.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf (2019), "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv [cs.CL]*.

Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra (2017), "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 618–26.

Shadish, W. S., William R. Shadish, Thomas D. Cook, and Donald Thomas Campbell (2002), *Experimental and Quasi-experimental Designs for Generalized Causal Inference*, Houghton Mifflin.

Timoshenko, Artem and John R. Hauser (2019), "Identifying Customer Needs from User-Generated Content," *Marketing Science*, 38 (1), 1–20.

Toubia, Olivier, Jonah Berger, and Jehoshua Eliashberg (2021), "How quantifying the shape of stories predicts their success," *Proceedings of the National Academy of Sciences of the United States of America*, 118 (26).

Van der Maaten, L. and G. Hinton (2008), "Visualizing data using t-SNE," *Journal of machine learning research: JMLR*.

Vermeer, Susan A. M., Theo Araujo, Stefan F. Bernritter, and Guda van Noort (2019), "Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media," *International Journal of Research in Marketing*, 36 (3), 492–508.

Vilà, Joaquim and Xavier Camps (2018), "From Identifying Needs to Generating Opportunities," *IESE Insight Reviews*, (37), 15–22.

Vonderembse, Mark A. and Tom Van Fossen (1998), "Quality function deployment," in *Handbook of Total Quality Management*, C. N. Madu, ed., Boston, MA: Springer US, 460–75.

Wang, Xin (Shane), Jiaxiu He, David J. Curry, and Jun Hyun (Joseph) Ryoo (2021), "Attribute Embedding: Learning Hierarchical Representations of Product Attributes from Consumer Reviews," *Journal of marketing*, 00222429211047822.

Yang, Yi, Kunpeng Zhang, and P. K. Kannan (2021), "Identifying Market Structure: A Deep Network Representation Learning of Social Engagement," *Journal of marketing*, 00222429211033585.

Zaltman, Gerald (1997), "Rethinking Market Research: Putting People Back In," *JMR, Journal of marketing research*, 34 (4), 424–37.

Zhang, Mengxia and Lan Luo (2021), "Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp."