frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 15 November 2023
DOI 10.3389/fpsyg.2023.1192020

Check for updates

OPEN ACCESS

EDITED BY
Francesco Walker,
Leiden University, Netherlands

REVIEWED BY
Leslie M. Blaha,
Air Force Research Laboratory, United States
August Capiola,
Wright-Patterson Air Force Base, United States
Sarah Jessup,
Universities of the Washington Metropolitan Area, United States,
in collaboration with reviewer AC

*CORRESPONDENCE
Areen Alsaid
✉ alsaid@umich.edu

RECEIVED 22 March 2023
ACCEPTED 27 October 2023
PUBLISHED 15 November 2023

CITATION
Alsaid A, Li M, Chiou EK and Lee JD (2023) Measuring trust: a text analysis approach to compare, contrast, and select trust questionnaires.
Front. Psychol. 14:1192020.
doi: 10.3389/fpsyg.2023.1192020

COPYRIGHT
© 2023 Alsaid, Li, Chiou and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Measuring trust: a text analysis approach to compare, contrast, and select trust questionnaires

Areen Alsaid[1]*, Mengyao Li[2], Erin K. Chiou[3] and John D. Lee[2]

[1]Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, MI, United States, [2]Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, United States, [3]Department of Human Systems Engineering, Arizona State University, Mesa, AZ, United States

Introduction: Trust has emerged as a prevalent construct to describe relationships between people and between people and technology in myriad domains. Across disciplines, researchers have relied on many different questionnaires to measure trust. The degree to which these questionnaires differ has not been systematically explored. In this paper, we use a word-embedding text analysis technique to identify the differences and common themes across the most used trust questionnaires and provide guidelines for questionnaire selection.

Methods: A review was conducted to identify the existing trust questionnaires. In total, we included 46 trust questionnaires from three main domains (i.e., Automation, Humans, and E-commerce) with a total of 626 items measuring different trust layers (i.e., Dispositional, Learned, and Situational). Next, we encoded the words within each questionnaire using GloVe word embeddings and computed the embedding for each questionnaire item, and for each questionnaire. We reduced the dimensionality of the resulting dataset using UMAP to visualize these embeddings in scatterplots and implemented the visualization in a web app for interactive exploration of the questionnaires (https:// areen.shinyapps.io/Trust_explorer/).

Results: At the word level, the semantic space serves to produce a lexicon of trust-related words. At the item and questionnaire level, the analysis provided recommendation on questionnaire selection based on the dispersion of questionnaires' items and at the domain and layer composition of each questionnaire. Along with the web app, the results help explore the semantic space of trust questionnaires and guide the questionnaire selection process.

Discussion: The results provide a novel means to compare and select trust questionnaires and to glean insights about trust from spoken dialog or written comments.

KEYWORDS

trust, trust assessment, trust measurement, questionnaires, text analysis, trust layers

1. Introduction

Trust has been studied in myriad contexts, from the internet to consumer products, healthcare, the military, and transportation. One challenge for advancing trust research is being able to measure trust precisely, and in a way that can generalize across contexts.

The study of trust in diverse contexts has resulted in multiple definitions: as a belief, an attitude, an intention, and a behavior. While these definitions are conceptually distinct, they are also interrelated; beliefs are derived from an individual's past experiences, affective processing of

beliefs governs attitudes, attitudes modulate intentions, and intentions are turned into behaviors (Ajzen and Fishbein, 1980). Nonetheless, trust is fundamentally an attitude. In the study of trust for systems design, trust is considered as a mediator between beliefs and behaviors, and hence, trust is defined as "the *attitude* that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee and See, 2004). Because of the variation in how trust has been conceptualized and subsequently operationalized, there exists a multitude of trust measures (Kohn et al., 2021).

Observable behaviors have been used as proxy measures of trust because they are often seen as more objective and less obtrusive than self-report measures. Behaviors are also often the main outcomes of interest when it comes to the study of trust, such as understanding what affects people's decisions to rely on the advice of a virtual real estate agent (Cassell and Bickmore, 2000), with trust being one important factor. Some behavioral indicators used to study trust have included: compliance, reliance, eye gaze, voice, and facial expression, and even pedal presses in automated vehicles (Meyer and Lee, 2013; Price et al., 2017; Lee et al., 2019).

Behavioral measures of trust are useful for understanding trust as a socio-cognitive construct between agents interacting in real-time (Takayama, 2009; Schilbach et al., 2013) and can serve as inputs to models used to dynamically predict human behavior in specific task contexts (Domeyer et al., 2018; Yang et al., 2021). However, behavioral measures are often tied to a specific context or experimental setup and are considered indirect measures of trust because it is possible to engage in trust-related behaviors without actually involving trust (Chiou and Lee, 2021). Therefore, when generalizing from laboratory studies to real-world scenarios, there is a substantial risk of misinterpreting or misapplying behavioral measures of trust. This risk is often associated with the potential for compromised construct validity, because of the lack of ground truth, which hinders the progress in establishing a robust theoretical foundation of trust (Campbell and Fiske, 1959; Lee and See, 2004).

In contrast to behavioral measures, questionnaires (i.e., self-report) are a more direct measures of trust, because trust is fundamentally an *attitude* and not a behavior. Therefore, asking a person about their attitude and closely associated factors, such as their beliefs and expectations, is important for understanding that person's trust. Although a person's reflective responses are also imperfect measures and not without limitations, questionnaires have the added advantage of being straightforward to administer, are typically rigorously developed based on trust theory, have established methods for validating empirically, and can more easily generalize across task contexts. Indeed, questionnaires have been widely used to measure trust. Yet, the literature indicates that several trust questionnaires have been developed for more specific task environments, perhaps to increase the sensitivity of the instrument. This has led to many trust questionnaires spanning multiple fields and contexts (Kohn et al., 2021).

The large pool of existing questionnaires presents a challenge to researchers in selecting the appropriate questionnaire. Questionnaire selection depends on several factors such as the application domain, the context, and the trust layer of interest. Questionnaire items can characterize different layers of trust such as dispositional, learned, and situational trust (Hoff and Bashir, 2014). A comparison between the questionnaires and their constituent items and words can guide the selection process. However, these relationships have not been systematically explored.

A recent paper described how nine questionnaires, measuring trust in automation specifically, related to one other based on a semantic network analysis of their constituent words (Jeong et al., 2018). Using Latent Semantic Analysis (LSA) in combination with network analysis, the paper identified 14 highly central words that could be used to create an integrated scale. While promising, this paper focused on the similarity between keywords. Focusing on the words only might overlook the contextual information contained by a questionnaire item or a questionnaire as a whole. In this study, we investigate the similarity across the words, the questionnaire items, and the questionnaires.

Text analysis could be used to reveal connections between the many different trust questionnaires. These connections can be condensed and visualized in two-dimensional semantic spaces. The manifestation of these connections in the semantic space at different levels of analysis (i.e., words, items, and questionnaires) allows researchers to compare and select the questionnaires that best support their research needs. Hence, text analysis provides one lens for considering the differences and similarities between various trust questionnaires.

Accordingly, the present analysis is not aimed at developing a new scale, nor finding a single ideal one, as there is no single ideal questionnaire that works for all experiment and contexts (Kohn et al., 2021). However, it gathers the most commonly used questionnaires in three most common research area: humans, e-commerce, automation. It also provides high-level comparison and guidance to researchers to choose the best-suited questionnaire for their research question.

1.1. A primer on text analysis

In text analysis, words are often represented as embeddings. Embeddings are vectors of numbers that describe the location of a word in a high-dimensional semantic space relative to other words (Pennington et al., 2014). For example, words like "cat" and "dog" would be closer to each other than "cat" and "mailbox." Words with similar meanings have similar vector representations and will thus be close to each other in the semantic space. The vector representation of words allows for mathematical operations that quantify the similarities of words and hence allows for advances in natural language processing applications like sentiment analysis and text autocompletion.

Methods that learn the vector representation of words are categorized into (i) global matrix factorization methods and (ii) local context window methods (Pennington et al., 2014). The first method exploits statistical information contained by the words, such as Latent Semantic Analysis (LSA). The intuition is to extract relationships between the words in the corpus, assuming that words similar in meaning will appear in similar contexts (Landauer and Dumais, 2008). LSA relies on the frequency of word occurrence and ignores the context in which the words appear. It represents the text data in a corpus matrix that consists of word frequencies in each document. Each word occurrence in each document is counted, and the entire matrix is reduced using Singular Value Decomposition (SVD). As a result, documents that share more words are considered similar, even if the similar words were used in a different context [e.g., the "bank" in "river bank" and "bank ATM" is considered equal (Hu et al., 2016)]. LSA produces semantic spaces that are high-level abstractions that are useful but lack context information.

The second method uses skip-gram models to capture the local context in which the word occurs. In skip-gram models, a constant

length window is moved along the corpus, and a neural network is trained to capture the co-occurrence of words in that entire window, and to predict context based on the central word (Altszyler et al., 2016). One example is a technique called word2Vec (Mikolov et al., 2013), which preserves the local context and provides a more precise description of the relationships between words compared to LSA and SVD. In word2Vec, embeddings are estimated by predicting words based on the words in the predefined window which enables the embeddings to capture relationships between words such that vector operations on the embeddings can complete word analogies in a meaningful fashion. In line with this paper's goals, we chose Global Vectors for Word Representation (GloVe) because it is a suitable approach for text analysis tasks that require considering the context within the data and the broader context of spoken language (Pennington et al., 2014). GloVe combines the benefits of global factorization and local context methods: it uses the statistical information contained by the words while also accounting for context by considering the co-occurrence statistics of words within a corpus. GloVe is trained on the non-zero elements of aggregated global word-to-word co-occurrence probability matrix and shows improved interpretability and accuracy compared to Word2Vec.

The vector representation of the words defines the position of each word in a high dimensional space, typically 100–500 dimensions. However, high-dimensional data is hard to visualize making it hard to identify what words similar to each other (Patel, 2016). Dimensionality reduction techniques reveal the underlying structure of the data. Principal component analysis (PCA) is a common dimensionality reduction technique that finds the linear combinations of the variables that capture the most variance in a dataset (Hubert et al., 2005). t-Distributed Stochastic Neighbor Embedding (t-SNE) is another technique that accommodates non-linear relationships between the variables and more precisely captures the micro-structure of the data: it maps similar instances to nearby points and dissimilar instances to distant point in the lower-dimensional space (van der Maaten and Hinton, 2008).

In this paper, we use a non-linear dimensionality reduction technique, called Uniform Manifold Approximation and Projection (UMAP), which captures non-linear relationships, like t-SNE, but in a more reproducible and computationally efficient manner (McInnes et al., 2018). UMAP creates a low-dimensional representation where similar items are near each other, and it preserves the micro and macro structure of the data, which is appropriate for this study since it is beneficial to understand both the inter-cluster and intra-cluster relationships for the trust questionnaires, their constituent items, and words (i.e., highlighting how they relate or diverge). For visualization purposes, we used a two-dimensional space. However, the results of dimensionality reduction might not be directly comprehensible since it is highly non-linear (McInnes et al., 2018), nonetheless, they can reveal important relationships between the variables (Alsaid et al., 2018; Alsaid and Lee, 2022). For more details on text analysis and dimensionality reduction techniques, see Appendix A.

1.2. Research objective

In this paper, we use word-embedding text analysis to understand different aspects of trust questionnaires and selecting the appropriate ones. First, we conduct a scoping literature review to gather existing questionnaires. Second, we apply text analysis techniques to quantify the relationships between the words used in the questionnaires, the questionnaire items within the questionnaires, and the overall questionnaires. These relationships were quantified using GloVe vector representations of the words. Third, we develop charts that quantify the composition of each questionnaire (i.e., application domain and trust layer composition) to guide researchers to select a questionnaire suited for the research task at hand. Finally, we generate a lexicon of the trust-related words that could be used to develop trust questionnaires and trust-focused sentiment analysis. The results are implemented in a web application that can help the researchers compare and contrast the different trust questionnaires and select the best fit for their research needs.

2. Method

2.1. Compiling and labeling the corpus

A scoping literature review (Grant and Booth, 2009) using google scholar was conducted using the keywords: "trust in automation, trust in humans, trust in e-commerce, trust, assessment, scales," their variants (e.g., "technology," "robots," "interpersonal trust," "surveys," "questionnaires") and their combinations.

Titles and abstracts were read to select those that developed or used rating-based trust measures. Then all articles were read in detail, and only *unique* developed questionnaires were included in the final selection and multiple questionnaires had overlapping items. A total of 80 articles were downloaded that met these inclusion criteria. Of these 80 articles, 46 questionnaires were extracted, with a total of 626 questionnaire items. After assessing the final selection of questionnaires, the questionnaires were categorized and labeled based on the domain for which they were developed (Chita-Tegmark et al., 2021):

1. Automation: questionnaires developed for assessing trust in automation, including robots and technology more generally.
2. E-Commerce: questionnaires developed to assess consumers' trust in brands, trust in retailers' websites, and online shopping in general.
3. Human: questionnaires developed to assess interpersonal trust.

Because not all questionnaire items assessed the same layer of trust, the items within each questionnaire were also categorized and labeled according to the layer of trust that they measured, based on Hoff and Bashir's model of trust layers (2014), for its comprehensiveness:

1. Dispositional: measures a person's general tendency to trust, independent of context or a specific system. Dispositional trust arises from long-term biological and environmental influences.
2. Learned: measures a person's trust based on previous experiences with a specific automated system.
3. Situational: measures trust in a specific context or situation including both the external environment and the internal, context-dependent characteristics of the operator.

Earlier trust scales primarily examined interpersonal trust and dispositional trust as measures of individual differences and personal

characteristics. However, as technology advanced, there emerged a growing interest in studying trust within specific domains such as e-commerce and automation. Furthermore, the scales appeared to span a spectrum ranging from general assessments, such as Rotter (1967) scale, which assessed overall trusting tendencies, to more specific questionnaires like Körber (2018) which focuses on particular aspects of automation.

2.2. Data cleaning

Once the corpus was compiled and labeled, the first step of our text analysis was to pre-process the data. We converted all words to lowercase, removed one-letter words, and punctuation. We also converted plural words to their singular form, such as 'decisions' to 'decision. To focus the analysis on words relevant to trust assessment, we excluded stop words. Stop words refer to unimportant, uninformative, frequently used words such as pronouns, prepositions, and auxiliary verbs. Here, we used a list of stop words from the tidytext package, specifically, the Onix stop word lexicon. The Onix stop word list was moderately aggressive in removing words compared to other stop word lists. The Onix list includes words such as "become," "know," "fully," "great" and "interesting." In addition, we removed words that referred to either the trustor or trustee like "product," "system," "user," "technology," or "consumer" because of their high frequency and limited relevance to making conceptual distinctions regarding trust. A complete list of removed word can be found in Appendix C.

2.3. Data analysis

Using the Wikipedia 2014 + Gigaword 5 pre-trained word vectors dataset provided on the GloVe website (Pennington et al., 2014), we calculated embeddings for each word, questionnaire item, and questionnaire. At the word level, we matched the words' vector embeddings with those in the pre-trained data. At the item level, we calculated the log odds ratios weighted by an uninformative Dirichlet prior for the words in each item. Using the log odds increases the weights for words that are common in a specific item, and relatively uncommon among all other items. This method gives greater weight to distinguishing words (Monroe et al., 2008). The log odds ratios were then used to create a weighted mean of the embeddings of the words that comprise each item. We used this same process to calculate an embedding for each questionnaire. For more details on the log odds ratio calculations see Appendix A.

To develop the trust lexicon, we calculated the log odds ratio of the words in the trust questionnaires, given the 5,000 most common English words list from the wordfrequency website (Davies and Gardner, 2010), and extracted the 20 most unique trust words. We calculated the cosine similarity distance between each of the 20 words and the GloVe word embeddings, similar to the approach of Fast et al. (2017), and for each of those word, we extracted the five closest words.

2.4. Software tools

We used R statistical software (R Development Core Team, 2016) to create plots with the 'ggplot2' package (Wickham and Winston, 2019); for data cleaning, we used the 'tidyverse' (Wickham, 2016) and 'tidytext' (Silge and Robinson, 2016) packages, and for dimensionality reduction, we used the 'umap' package (Konopka, 2019).

3. Results

Table 1 shows the questionnaires included in the text analysis from our mapping review, the number of items in each questionnaire, the number of citations per article, and the labeled domain category.

3.1. Word-level analysis

Figure 1 shows the relevant trust words gleaned from the questionnaires' 626 items. The size of the word reflects its frequency; the bigger the word, the more often it occurs. The words are arrayed based on the UMAP dimensionality reduction of the word embeddings. The dimension of the GloVe word embeddings was 100 and using UMAP it was reduced to two. The choice of two dimensions was for visualization purposes. This approach has been used in several application and showed high performing results (Wang et al., 2021). The words near each other in this space are expected to have similar or complementary meanings. Some of the words create themes that directly map to different trust dimensions.

For example, in the middle of the figure, there is a cluster that includes "believe," "expect," "advice," and "decision." This area generally seems to be about truth reasoning and anticipating behavior and predictability. This contrasts with the bottom right cluster that focuses on fairness, and security and includes words like "fair," "welfare," "trust," and "secure." This is much more about how the behavior is valued. Also, the words in the upper right cluster include "experience," "skills," and "knowledge" which seem to characterize competence and ability. Finally, the upper left cluster includes words such as "dependable," "reliable," "competent," "honest" which seem to characterize performance, and measure the integrity and reliability dimensions of trust. The upper left cluster also includes words like "cheat," "honest," and "sincere" which seem to characterize integrity. Whether or not these provide a comprehensive account of trust is the topic for other papers (Lee and See, 2004; Chiou and Lee, 2021; Malle and Ullman, 2021), but it certainly gives us an idea about the current and most common state of how researchers are measuring trust perceptions. Commonly, these dimensions of trust characterize trustworthiness in different objects, and this might explain why they are showing in certain clusters. For example, although not uniquely, integrity typically characterizes trust in humans whereas reliability characterizes trust in automation (Malle and Ullman, 2021).

3.2. Item-level analysis

Figure 2 shows the UMAP representation of the items. The items from the most cited questionnaire in each domain are encircled and color-coded. Note that the figure shows *all* items of *all* questionnaires. The (Rotter, 1967) questionnaire was developed to measure human-human trust, the (Gefen et al., 2003) questionnaire was developed to measure trust in online shopping, and the (Jian et al., 2000) questionnaire was designed to measure trust in automation. Their

TABLE 1 Published articles with trust questionnaires ordered by the number of citations in each category.

Paper	Items	Citations	Category
Jian et al. (2000)	12	1,139	Automation
Muir and Moray (1996)	9	1,014	Automation
Dzindolet et al. (2003)	1	967	Automation
Mcknight et al. (2011)	23	564	Automation
Madsen and Gregor (2000)	25	338	Automation
Singh et al. (1993)	12	278	Automation
Yagoda and Gillan (2012)	36	155	Automation
Merritt (2011)	6	145	Automation
Schaefer (2013)	40	140	Automation
Charalambous et al. (2016)	10	86	Automation
Chancey et al. (2017)	15	61	Automation
Körber et al. (2018)	1	58	Automation
Körber (2018)	19	54	Automation
Montague (2010)	28	52	Automation
Merritt et al. (2015)	10	51	Automation
Wiczorek and Manzey (2014)	1	50	Automation
Chien et al. (2014)	21	39	Automation
Goillau et al. (2003)	8	21	Automation
Albert et al. (2009)	1	20	Automation
Walliser et al. (2016)	1	18	Automation
Khalid et al. (2016)	14	15	Automation
Malle and Ullman (2021)	20	12	Automation
Moeckli et al. (2015)	12	11	Automation
Holthausen et al. (2020)	6	8	Automation
Salcedo et al. (2011)	20	5	Automation
Schneider et al. (2017)	6	5	Automation
Byrne and Marín (2018)	5	1	Automation
Gefen et al. (2003)	7	9,222	E-Commerce
McKnight et al. (2002)	3	5,981	E-Commerce
Bhattacherjee (2002)	7	1,835	E-Commerce
Delgado-Ballester et al. (2003)	8	1,202	E-Commerce
Sohaib and Kang (2015)	32	35	E-Commerce
Rotter (1967)	25	5,666	Human-Human
Rempel et al. (1985)	26	4,947	Human-Human
Mayer and Davis (1999)	38	2,863	Human-Human
Yamagishi and Yamagishi (1994)	24	2,782	Human-Human
Gefen and Straub (2004)	6	2,296	Human-Human
Larzelere and Huston (1980)	8	1,989	Human-Human
Yamagishi (1986)	5	1,750	Human-Human
Johnson-George and Swap (1982)	25	1,630	Human-Human
Wrightsman (1964)	13	526	Human-Human
Miller and Mitamura (2003)	2	380	Human-Human
Evans and Revelle (2008)	21	332	Human-Human
Frazier et al. (2013)	4	86	Human-Human
Goto (1996)	3	74	Human-Human
Allen et al. (2004)	7	33	Human-Human

Number of citations as of October 2021.

different purposes are reflected in their placement in the semantic place.

Although very frequently cited, the questionnaire by Gefen et al. (2003) spans a small area, mainly because the items were similar and were developed to assess trust in an online vendor (i.e., "e-Commerce") based on past experiences. The questionnaire items were framed as, "Based on my experience with the online vendor in the past, I know it is…," with items assessing factors such as predictability and trustworthiness. On the other hand, the (Rotter, 1967) questionnaire has items that are widely spread across the semantic space. This is because it assesses dispositional trust through a large set of questions related to views of the future, attitudes toward society, and hypothetical ethical scenarios. Jian et al. (2000) trust in automation questionnaire has some items close in the UMAP space to those of Rotter's interpersonal trust questionnaire; this is because it includes items that assess the reliability and dependability of automation and people.

Figure 2 showed that questionnaires vary in how their items spread across the semantic space. The questionnaire's spread is an indicator of the breadth of the questionnaire, and the different dimensions of trust it covers. Table 2 shows the spread value of each questionnaire. Spread was calculated as the average Euclidian distance between the questionnaire's items and the questionnaire centroid in the semantic space. As such, single-item questionnaires have a spread of zero.

3.3. Questionnaire-level analysis

Figure 3 shows the UMAP semantic space representation of the questionnaires, which shows how the questionnaires relate to each other. The upper region of the semantic space is dominated by trust in automation questionnaires, whereas the bottom left region mostly consists of human-human trust questionnaires. The e-commerce questionnaires are spread across the entire space.

Questionnaires close to each other share terms or similar terms that make them close in the UMAP space. The bottom left cluster consisted of questionnaires about trust in humans and had many questions related to peoples' behaviors such as "honesty" and "cheating" and how the person viewed "relationships" with and "personalities" of others. The questionnaires in the upper cluster mostly assessed efficiency, dependability, reliability, and safety in the specific domain that the questionnaire was developed. Finally, the questionnaires in the bottom right cluster commonly asked about the general tendency to trust (e.g., "I usually trust machines until there is a reason not to.)" The similarities and differences between these questionnaires can be further explored ls in the web app.

By calculating the log odds ratio of words in a specific category given all words used in all questionnaires, we identified the words that are most unique and distinguishing of trust across domains. Figure 4 shows the 10 highest frequencies of words in each domain (along the horizontal axis) and how unique these words are to each of the specific domains (along the vertical axis). Some words had tied frequencies, and thus the figure shows more than 10 words in e-commerce and automation domains. The figure shows that some words like "reliable" and "perform" were *more* unique to Automation, whereases words like "experience, or "transactions" were *more* unique to e-commerce, and words like "honest" or "cheat" were more unique to human-human. The log odds ratio emphasizes words that are *common in a specific category* and relatively *uncommon in others*, making them more frequent in their respective domains, though not exclusive to them.

This is different than the Term Frequency-Inverse Document Frequency (TF-IDF) metric. TF-IDF is used to assess the rarity or importance of a term in the document collection, while log odds ratio measures the association or relevance of terms to specific categories or classes. IDF focuses on the overall collection of documents, whereas log odds ratio emphasizes term relevance within specific categories or classes. For more details on the log odds ratio calculations see Appendix A.

3.4. Questionnaire composition

To help guide questionnaire selection, we assessed the questionnaire composition of domain-related words and layers of trust. The domain categories refer to trust in automation, e-commerce, or humans; the trust layers refer to dispositional, situational, or learned trust.

To calculate the proportion of domain-related words in each questionnaire (e.g., the questionnaire can have 20% automation-related terms, 30% e-commerce-related terms, and 50% human-human-related terms), we used the log odds ratio results shown in Figure 4. The log odds ration compares the frequency of words in a certain domain to all words form all questionnaires. The questionnaires' domain composition results are illustrated in Figure 5, where the questionnaires are ordered by the proportion of automation, e-commerce, and human-human content. For instance, Figure 5 shows that Goto's (Goto, 1996) questionnaire is composed of 100% human-related words while the Delgado-Ballester et al. (2003) words involve a combination of each human, e-commerce, and automation-related words. In addition, we looked at words' usage over the years and noticed that the language and terminology used in questionnaires underwent changes, but these changes were a result of the evolving nature of the domains rather than mere temporal shifts.

Similarly, we calculated the proportion of each of the trust layers for each item. The items' layers were coded by three researchers. One researcher led the labeling, while the other two provided reviewed the labels. The finalization of the labels occurred when all three researchers reached a consensus and agreed upon them. All items were categorized as either dispositional, learned or situational. The result is shown in Figure 6. Evans and Revelle (2008) is a questionnaire for measuring trust in humans and 100% of the questionnaire's items are learned; all items assess trust based on previous experiences. The Rotter (1967) questionnaire also measures trust in humans, but its items are 100% dispositional, meaning all items assess a person's innate tendency to trust. This is a starting point for selecting the questionnaire that best suits the research objective.

3.5. Trust lexicon

Figure 4 compared trust words in a specific domain category (i.e., automation, e-commerce, human-human) to trust words across all questionnaires. However, to identify trust-related words across domains, we compared them to the common words in the English language. We used a list of 5,000 frequent words from the Corpus of Contemporary American English (COCA) (Davies and Gardner,

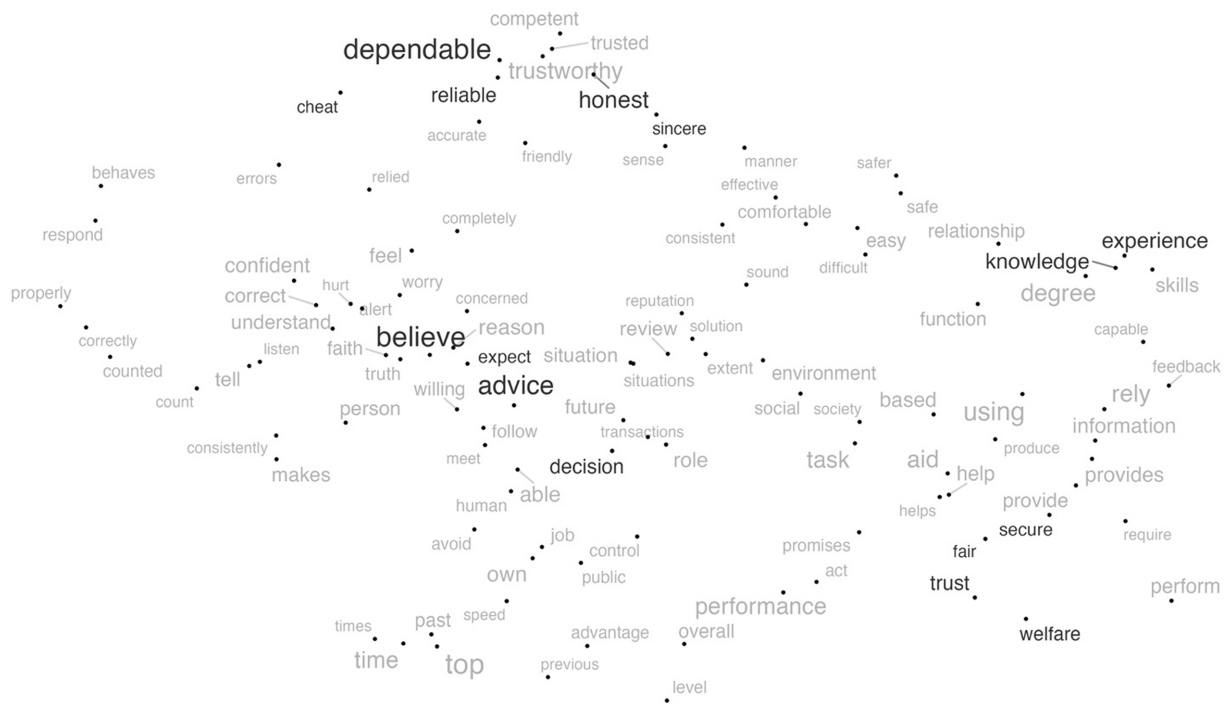

FIGURE 1
The UMAP two-dimensional representation of the questionnaire words. The words discussed in the manuscript as examples of specific themes are highlighted in black.

2010). The corpus includes words from different genres; spoken, fiction, magazines, newspapers, and academic texts as shown in Figure 7. Words above the dashed line ($y = 0$), indicate words that are more common in the trust questionnaires than in common English usage, while words below the dashed line are more prevalent in common English than in the trust questionnaires. Not surprisingly "trust" and "people" occur frequently in the questionnaires and occur much more frequently in the questionnaires than they do in English usage.

Based on the 20 most unique trust words revealed in Figure 7, we created a trust lexicon shown in Figure 8. For each of the 20 words from Figure 7, we extracted the five closest neighboring words in the high-dimensional GloVe embeddings space using the cosine distance. Before extracting the five closest words, we conducted the same data cleaning steps described in the methods on the GloVe words, to avoid words like "the" and "from." The 20 most unique trust words from the questionnaires are shown in light gray, while the other words in dark gray are their closest neighbors.

4. Discussion

The results of the word-embedding text analysis identified the most common words used to assess trust, revealed semantic similarities and differences across the trust questionnaires, and provided a detailed comparison of the questionnaires' composition based on domain and trust layers. The results were implemented in an interactive web application that allows further exploration of the analysis. Particularly, the word-level and item-level results can support further questionnaire development and the questionnaire-level results can aid questionnaire selection given the research objective for measuring trust, and consideration of the trusting context.

4.1. Word-level analysis: common words and trust lexicon

The word-level analysis allowed us to identify the frequently used words to measure trust. Common words included "dependable," "experience," and "advice." The neighboring words in the semantic space also had similar or complementary meanings such as "reliable," "knowledge" and "helpful," respectively. Word clusters shown in Figure 1 can be mapped to trust dimensions identified by Malle and Ullman (2021). "Dependable" and "reliable" are closely situated next to each other, which can be mapped to the performance dimension of trust, whereas "secure" and "fair" show the moral dimension of trust. The themes revealed by the most frequent words were consistent with literature on trust dimensions (i.e., integrity and competence) (Malle and Ullman, 2021). When categorized into trust in automation, humans, and e-commerce, more specific themes emerged (e.g., reliability and dependability to describe automation. Honesty and sincerity to describe humans, and prior experiences and services to describe e-commerce).

Understanding the words associated with trust questionnaires can be helpful in different ways. The identified trust-related words can be used as a single-word trust assessment tool, e.g., by asking subjects to rate how well these words describe the system of interest. The trust-related words can be descriptors used in card sorting tasks to measure trust, similar to the microsoft desirability toolkit (Benedek and Miner, 2002). In addition, these words can be used as

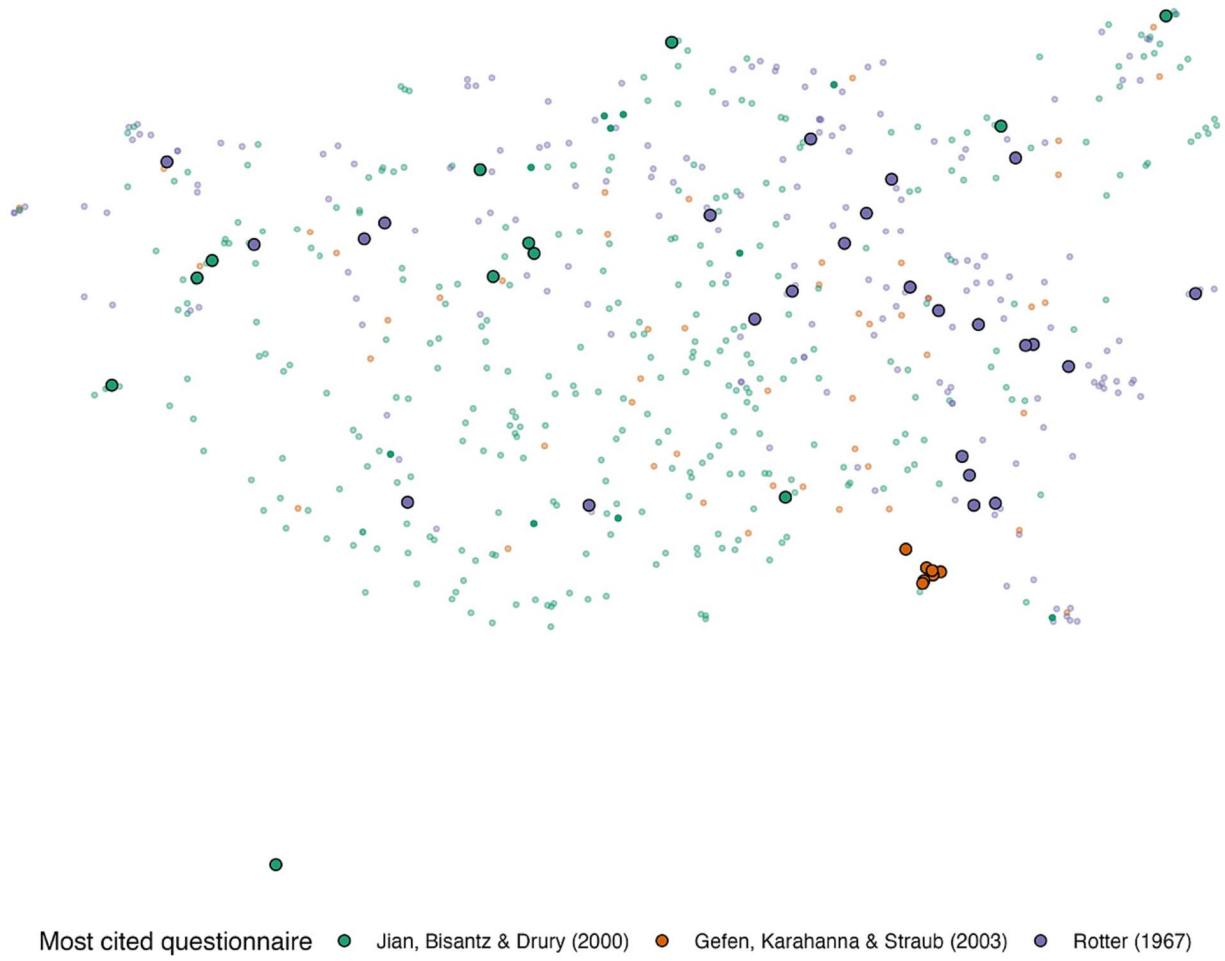

FIGURE 2
The UMAP two-dimensional representation of the embeddings of the questionnaire items. The most cited questionnaire in each category (Gefen et al., 2003 for e-commerce; Muir and Moray, 1996 for automation; Rotter, 1967 for human-human). Green, orange, and purple represent items in automation, e-commerce, and human-human trust questionnaires, respectively.

a basis robust or more precise instruments for measuring trust. For example, the Jian et al. (2000) scale development process consisted of three key steps: collecting comprehensive list of trust-related words, assessing word similarity through a questionnaire, and rating the similarity of word pairs through a paired comparison study. The resulting words were then clustered and used to construct the trust scale. This process can be intensive and tedious. Leveraging the identified words can help us study their similarity without subjective bias or intensive manual effort, which could make future instrument development more robust and precise.

By examining the most common words in each domain, we found similarities and differences in what questionnaires typically use to describe trust in automation compared to trust in humans. This can be because some forms of automation are more human-like, such as anthropomorphic agents or virtual humans, hence questionnaires involving more human-like qualities in these research contexts may be more appropriate (Lankton et al., 2015). Furthermore, there are certain trust dimensions that are relevant to both the human and the automation such as integrity and competence, which can explain the similarities.

The words' semantic space revealed by this analysis was used to create a trust lexicon. Word sentiment lexicons can estimate people's emotional state (Tausczik and Pennebaker, 2010) and attitudes on social media (Pang and Lee, 2008) or in conversations (Li et al., 2020). Word sentiment lexicons are typically created through tedious manual labeling of each word in the dictionary, which produces a sentiment rating for each word. But when word embeddings are combined with dimensionality reduction techniques, they reveal correlations between words and how they might relate which can expedite the development of lexicons. The lexicon resulting from our analysis can build on and improve similar lexicons of trust-related words (Mohammad and Turney, 2013). However, our approach relied solely on the machine learning algorithms to identify similarities and difference between words used in trust assessment questionnaires. While machine-learning-led approaches are helpful and efficient, they have inherent limitations. For instance, in Figure 8 we can see the proximity of words like "degree," which is commonly used in trust questionnaires, to words like "postgraduate" which is irrelevant to trust. Future work might explore a mixed-initiative approach to lexicon development leveraging machine learning efficiency and human expertise (Alsaid

TABLE 2 Questionnaires' spread measured as the mean Euclidian distance between questionnaires items in the semantic space.

Paper	Category	Spread
Albert et al. (2009)	Automation	0
Dzindolet et al. (2003)	Automation	0
Körber et al. (2018)	Automation	0
Walliser et al. (2016)	Automation	0
Wiczorek and Manzey (2014)	Automation	0
Chancey et al. (2017)	Automation	3.2787649
Goillau et al. (2003)	Automation	9.8178918
Merritt (2011)	Automation	11.0535503
Singh et al. (1993)	Automation	13.3077262
Holthausen et al. (2020)	Automation	19.8532091
Merritt et al. (2015)	Automation	24.075077
Moeckli et al. (2015)	Automation	26.7285164
Yagoda and Gillan (2012)	Automation	26.8700501
Khalid et al. (2016)	Automation	26.9118744
Charalambous et al. (2016)	Automation	27.3718082
Byrne and Marín (2018)	Automation	32.1872312
Salcedo et al. (2011)	Automation	35.4811105
Muir and Moray (1996)	Automation	40.9128248
Schneider et al. (2017)	Automation	44.2913665
Madsen and Gregor (2000)	Automation	45.7479433
Jian et al. (2000)	Automation	52.9878988
Chien et al. (2014)	Automation	58.3499561
Malle and Ullman (2021)	Automation	63.6366837
Mcknight et al. (2011)	Automation	64.6944447
Körber (2018)	Automation	72.247291
Montague (2010)	Automation	117.608309
Schaefer (2013)	Automation	141.482416
McKnight et al. (2002)	E-Commerce	3.7473223
Gefen et al. (2003)	E-Commerce	4.047058
Bhattacherjee (2002)	E-Commerce	14.6113045
Delgado-Ballester et al. (2003)	E-Commerce	29.7256945
Sohaib and Kang (2015)	E-Commerce	73.1215945
Goto (1996)	Human-Human	0.1458184
Miller and Mitamura (2003)	Human-Human	1.1019821
Frazier et al. (2013)	Human-Human	7.5817419
Yamagishi (1986)	Human-Human	19.049358
Gefen and Straub (2004)	Human-Human	21.7186439
Wrightsman (1964)	Human-Human	23.8513333
Evans and Revelle (2008)	Human-Human	30.5084731
Larzelere and Huston (1980)	Human-Human	31.4059937
Rempel et al. (1985)	Human-Human	43.247883
Johnson-George and Swap (1982)	Human-Human	43.3757231
Allen et al. (2004)	Human-Human	44.0729489
Rotter (1967)	Human-Human	51.903995
Yamagishi and Yamagishi (1994)	Human-Human	85.8964228
Mayer and Davis (1999)	Human-Human	122.881738

FIGURE 3
The UMAP two-dimensional representation of the questionnaire embeddings with domain category color-coded. Green, orange, and purple represent the automation, e-commerce, and human-human categories, respectively.

and Lee, 2022; Alsaid et al., 2023). This synergy can provide more robust and accurate results in various natural language processing tasks.

4.2. Item-level analysis: items' spread and trust characteristics

The item-level analysis showed that some questionnaires had items close to each other while items in other questionnaires were more dispersed. The distribution of the items was linked to the variation of the trust characteristics being evaluated by different items in each questionnaire (e.g., reliability, performance, cheating, prior experience, etc.); the greater the spread of a questionnaire's items, the more layers and characteristics are captured.

In general, the human-human questionnaires were broadly distributed. Because, in line with the questionnaire-level results, items in this domain tended to assess a wide range of human characteristics in different hypothetical scenarios. This implies that these questionnaires used words from varying contexts, which would explain their spread in the semantic space. On the other hand, some questionnaires' items were contained in a very small area in the semantic space. One illustrative example is the most cited e-commerce questionnaire (Gefen et al., 2003); the questionnaire's items assess similar characteristics (and thus used closely related words) of online vendors such as "reliability," "honesty" and "trustworthiness."

In summary, the distribution of the questionnaire items in the semantic space can reflect the variety of trust dimensions being measured – the more dimensions the trust questionnaire captures, the more spread the items are. Therefore, one important consideration when selecting a trust questionnaire is the spread of its items. If the research question requires evaluating a specific quality that is associated with trust (e.g., reliability), then researchers could pick questionnaires with items that are closer together in the semantic space, whereas if the research questions require evaluating multiple qualities associated with trust (e.g., prior experience, performance, deception), a more spread questionnaire is likely more appropriate. Careful examination of the questionnaire and its constituent items is necessary.

4.3. Questionnaire-level analysis: domain- and layer- based selection

At the questionnaire level, the results revealed three main clusters, one consisting of mainly human-human trust questionnaires, and two

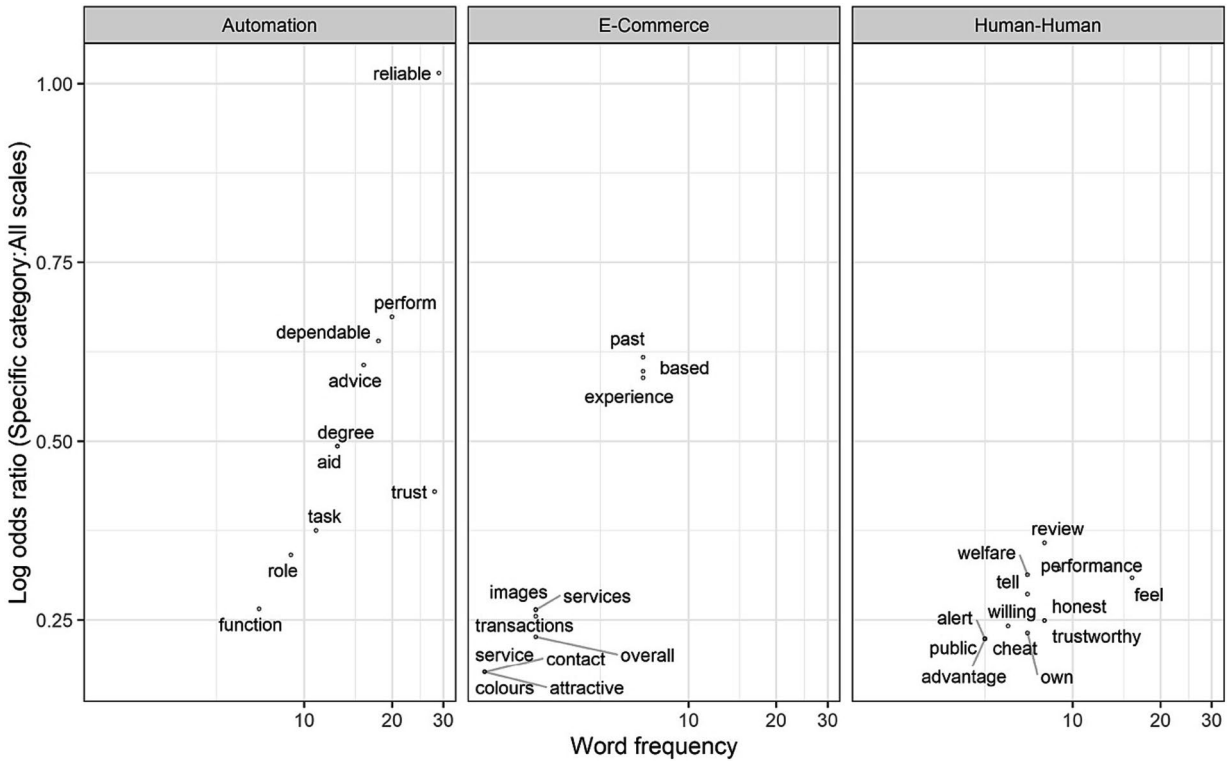

FIGURE 4
The log odds ratio of the 10 most common words in a specific category given all words in the questionnaires. The x-axis represents words frequency and is displayed on a logarithmic scale (base 10) to allow for the visualization of a wide range of values where the spacing between points increases exponentially as you move further right. The y-axis represents the log odds ratio of each word: the higher the word on the y-axis. The log odds ratio emphasizes words that are common in a specific category and relatively uncommon in others.

containing a mix of trust in automation and trust in e-commerce questionnaires as shown in Figure 3. Questionnaires assessing a person's trust in other people were typically broad and contained diverse items assessing learned, dispositional, and situational trust through hypothetical scenarios, general views about the world, and the overall tendency to trust others. Questionnaires assessing trust in automation had two common themes with those assessing trust in e-commerce. Because after removing domain-specific words (e.g., "website" and "vendor" for e-commerce and "robot" and "automation" for automation), items of both domains were largely similar. One theme focused on assessments of reliability, accuracy, and trustworthiness of a system, while the other theme focused on the general tendency of people to trust or not trust new technologies. This explains the proximity in the semantic space, nonetheless, questionnaires developed for trust in automation might not be appropriate to assess trust in e-commerce and vice versa. Depending on the context and the research question, one theme or a combination of themes might be more appropriate, and researchers should carefully consider the aspects of trust being evaluated by each questionnaire (Kohn et al., 2021).

The questionnaire composition analysis provided a descriptive map for questionnaire selection based on layer and domain. In the layer composition analysis, questionnaires were summarized based on constituent items' layers whereas in the domain composition analysis, questionnaires were summarized based on their constituent words uniqueness to certain domains. This provides an overview of each questionnaire's composition and enables their comparison, however, aggregating and summarizing data in this way may result in overlooking certain details, like the overlaps highlighted in Figure 2. For instance, while the Rotter (1967) questionnaire exhibited large overlap in Figure 2 and high spread in Table 2, it appeared to be mostly composed of human-human words in Figure 5. This similar to reducing experimental data to mean values: while it might not capture every detail, it provides a useful overall summary.

4.4. Practical implications: questionnaire selection guidelines

The questionnaire composition analysis provided further clarity on how the myriad trust questionnaires compare to one another and can thus serve as initial guidance for selecting a questionnaire. Here, we outline general guidelines and considerations for the trust questionnaire selection process: identifying the domain and layer, and considering items' dispersion, and evaluating the tradeoff between number of items and sampling frequency.

After carefully defining the research questions and the underlying hypotheses, the researcher needs to identify *the domain* in which trust is being measured. This is important because trust questionnaires are typically developed to measure trust in a specific context and the way trust is characterized varies from one domain to another (Lee and See, 2004; Lewis and Weigert, 2012). This was evident in the word-level analysis in Figure 4: the words used to describe trust differed across domains. For example, Lee and Moray (1992) conceptualization of

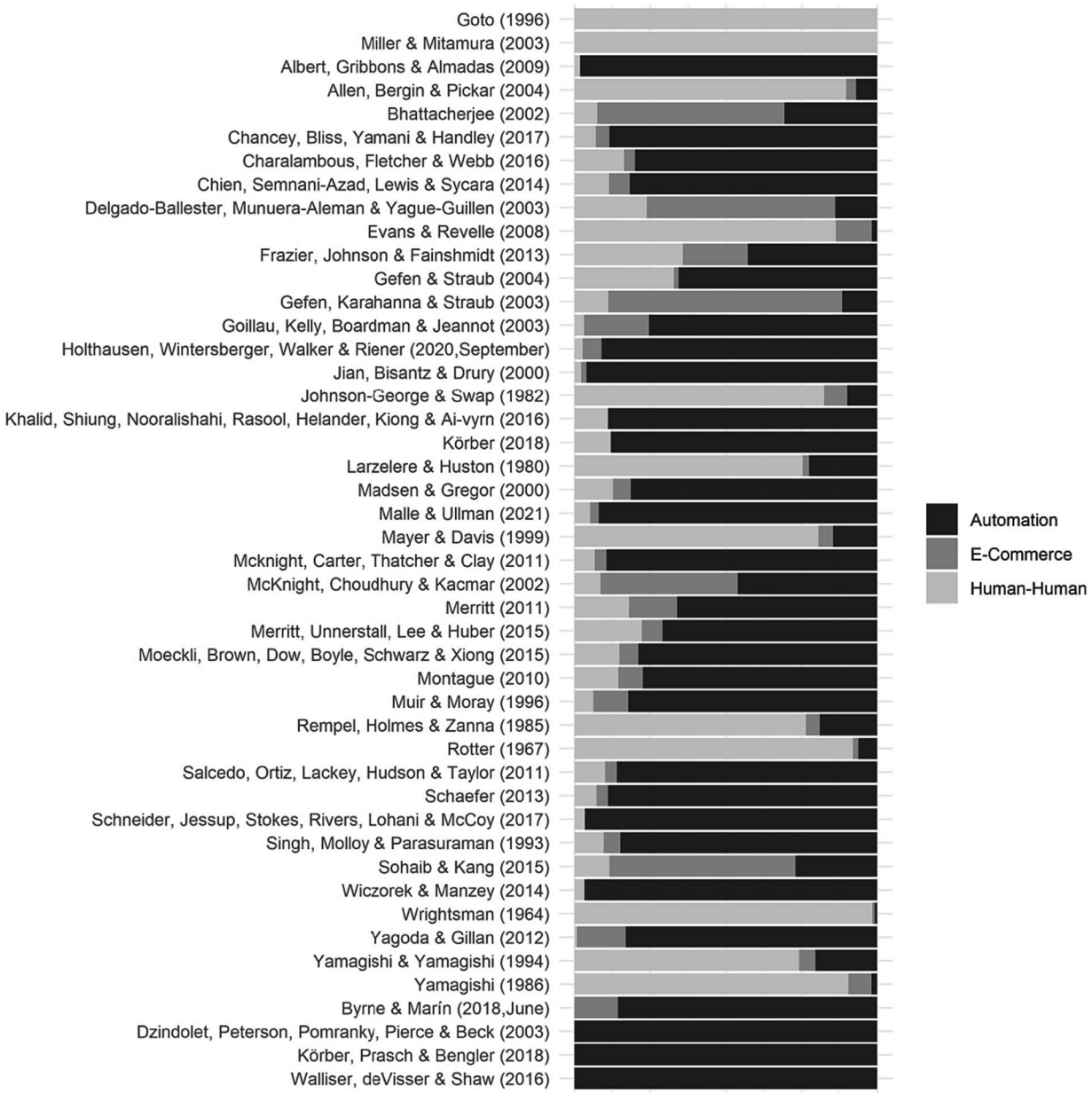

FIGURE 5
Trust questionnaires' domain composition (containing words most related to which domains of trust).

purpose (e.g., role), process (e.g., dependable), and performance (e.g., reliable) dimensions of trust in automation was apparent in some of the words, as well as the concept that trust may also have more moral dimensions (Mayer et al., 1995; Sheridan, 2019). Another important consideration is the attributes of the trustee. In Figure 5, we provided a questionnaire composition map to show what percentage of each questionnaire included words most unique to trust in automation, human-human, or e-commerce. The figure showed that some e-commerce questionnaires had a percentage of automation-related words, and some automation questionnaires had a percentage of human-human related words. In instances where the trustee is more human-like, it might be appropriate to select a questionnaire with higher human-human percentage. Nonetheless, this was based on the objective quantitative analysis, researchers should carefully assess whether or not a questionnaire is appropriate for measuring trust in a certain domain.

Moreover, identifying the *layer* of trust is important; whether the researcher is trying to assess people's general propensity to trust (i.e., dispositional), trust in a specific situation (i.e., situational), or trust based on previous experiences (i.e., learned). In Figure 6, we provided a map for understanding the composition of the questionnaires. Based on the research questions, the selected questionnaire items can be dispositional, learned, situational, or a combination (Merritt and Ilgen, 2008). It is important in this step to understand the nature of each trust layer. Measuring dispositional trust would be most appropriate for studies of individual differences, particularly when measuring trust across different cultures, as people from different cultures may have different perceptions of trust. Moreover, learned

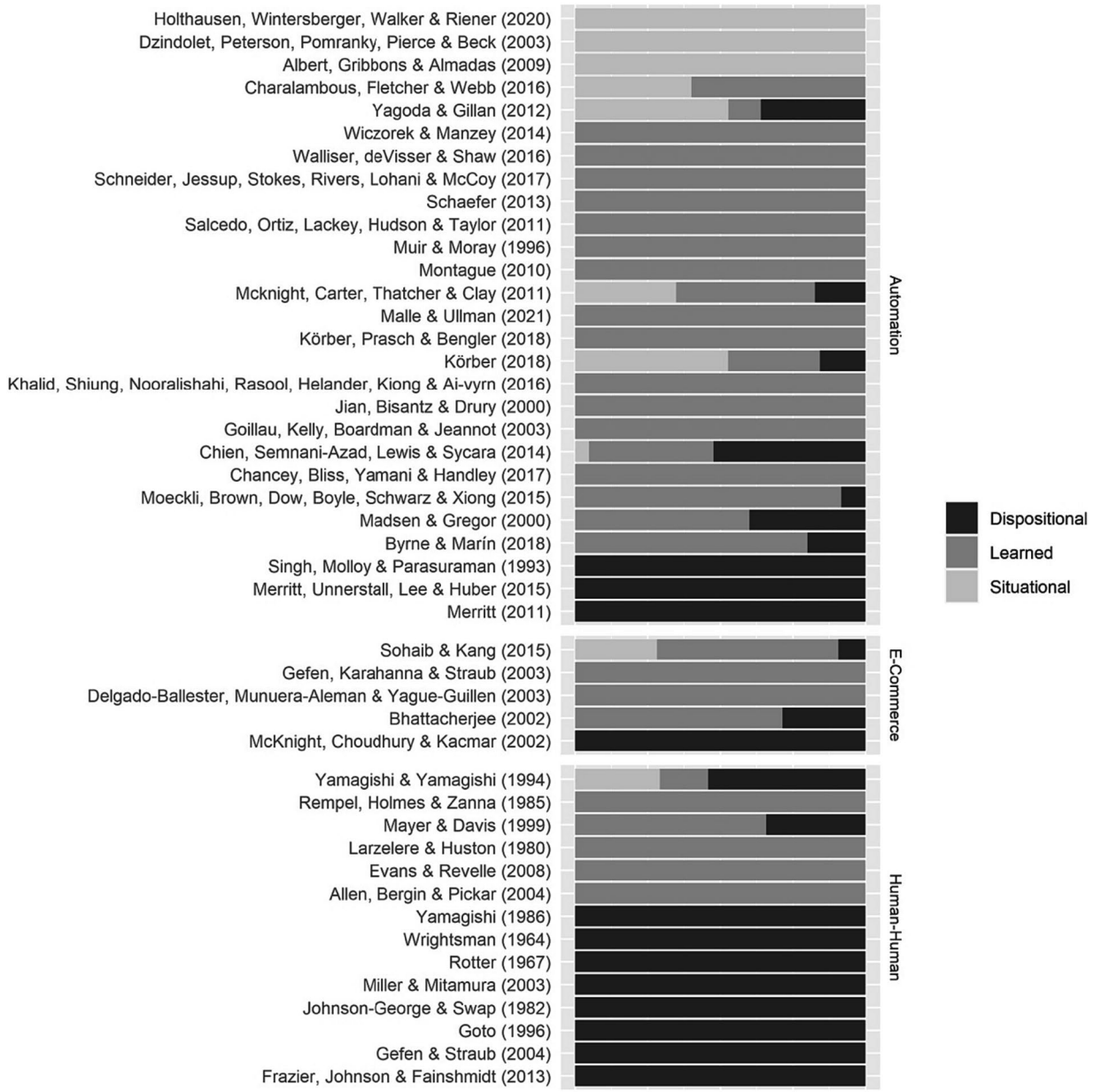

FIGURE 6
Trust questionnaires' layers of trust composition of each questionnaire item.

trust would be for studies of how interactions with an agent affect trust, and situational trust would be for measuring trust in a specific event. For example, if researchers are interested in evaluating the users' trust in automation in specific of interactions, a questionnaire that mainly consists of situational trust items would be suitable [e.g., Holthausen et al. (2020)] whereas if they wish to assess the persons' propensity to trust, a questionnaire like that mainly consists of dispositional trust item [e.g., Merritt (2011)] would be better suited.

Furthermore, the results revealed another element of the questionnaires' semantic characteristics and selection criterion: questionnaire items spread. The *spread* of the questionnaire items is an important criterion of selection: whether the research question and nature of the study focus on one or a few of the dimensions of trust (e.g., purpose, process, or performance information for forming a person's trust in automation (Lee and See, 2004), or rational and relational dimensions of trust in others (Lewis and Weigert, 2012)). This can be qualitatively determined by visually assessing the specific questionnaire items' distribution across the semantic space, or quantitatively by the spread values in Table 2 that were calculated as the mean of Euclidian distances from a questionnaire's centroid in the semantic space. When selecting a questionnaire, if researchers are interested in trust as a moderator or control variable (Fuchs and Diamantopoulos, 2009) or only focusing on a single aspect of trust (e.g., the performance of a particular automated system), then picking a narrower spread of trust scale can be appropriate (e.g., Chancey et al. (2017) for trust in automation which focuses on the ability and dependability dimensions of trust, Goto (1996) for trust in humans which measures trust relative to social distance, or McKnight et al.

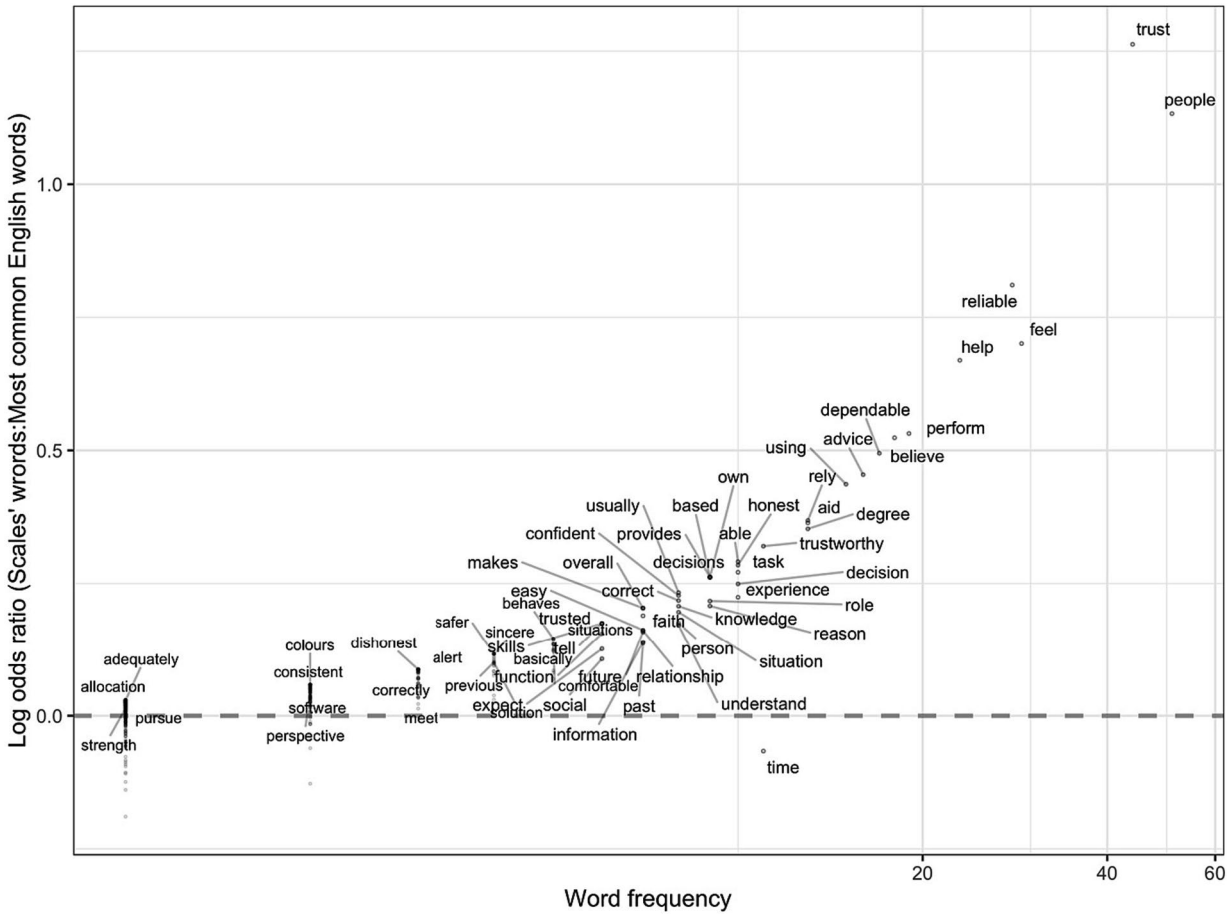

FIGURE 7
The log odds ratio of the words in the trust questionnaires to the words in the English language. The x-axis represents words frequency and is displayed on a logarithmic scale (base 10) to allow for the visualization of a wide range of values where the spacing between points increases exponentially as you move further right. The y-axis represents the log odds ratio of each word: the higher the word on the y-axis, the more unique the word is to trust questionnaires.

(2002) for trust in e-commerce which measures tendency in particular). If researchers are interested in assessing various characteristics and layers of trust in the study, then a broader spread of trust scale should be considered (e.g., Schaefer (2013) for trust in automation, Mayer and Davis (1999) for trust in humans, or Sohaib and Kang (2015) or e-commerce).

In addition, evaluating the trade-off between the *number of questionnaire items* and *sampling frequency* is critical (Kohn et al., 2021). If trust needs to be measured multiple times for its dynamic characteristic, using a few or single itemed questionnaire might provide a quick trust measurement and minimal interruptions to the continuity of a study participant's experience (Körber, 2018). However, one item might be limited and not measure the different dimensions of trust (Lee and See, 2004). If the research objective requires a more detailed assessment of trust, then multi-item questionnaires are recommended [e.g., Yagoda and Gillan (2012)]. This is particularly important in situations where different layers of trust need to be measured at different times of a study (i.e., dispositional trust before the study, situational trust during the study). In combination with the questionnaire composition analysis, the researcher can make an informed decision regarding the questionnaire selection with the right number of questions that meets the research needs.

Finally, the supplemented web app implementation provides an interactive interface to compare, contrast and select the questionnaire most appropriate based on the considerations provided above. For a more detailed explanation and description of how to use the app for questionnaire exploration and selection see Appendix B.

4.5. Limitations and further research

This study has several limitations. First, some of the identified words as part of the trust lexicon (e.g., "feel" or "believe") may be more an artifact of the measurement method and our ability to elicit self-report from lay-people (i.e., not trust scholars) through a scaled question (e.g., "how much do you feel…"). Framing questions as such is a common means to measure and quantify attitudes (Michael, 1979). Because attitudes have emotional, cognitive, and behavioral dimensions, asking questions about feelings enhances construct validity (Schwarz and Bohner, 2007) Nonetheless, trust is fundamentally an attitude, and not an intention, feeling, belief, or behavior (Lee and See, 2004).

Second, in the questionnaire composition analysis, we show what each questionnaire measures, in terms of domain and layers of trust,

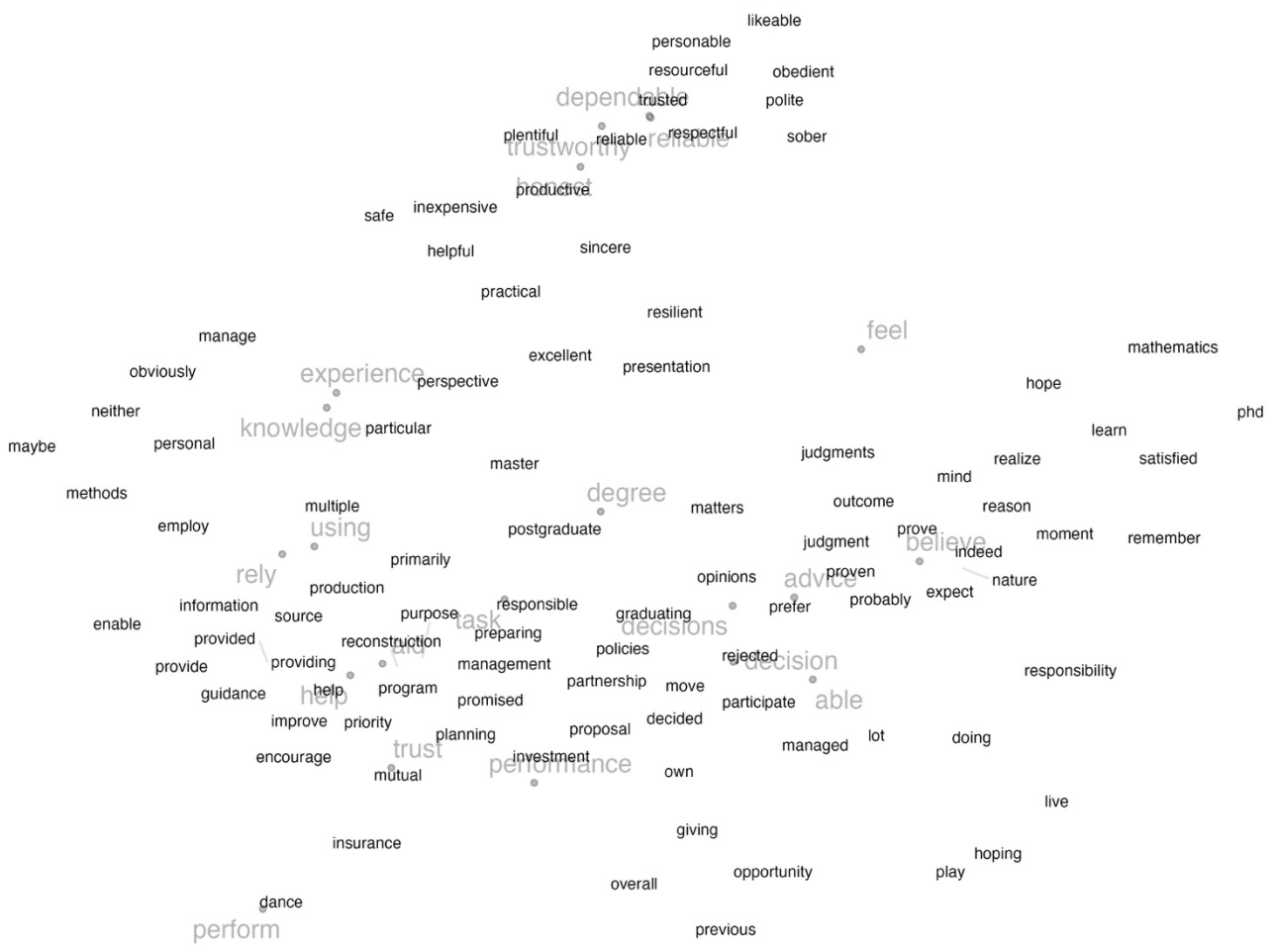

FIGURE 8
A lexicon of trust-related words. The words in light gray are the most common in the trust questionnaires, while the rest of the words are their closest neighbors in the high-dimensional GloVe embeddings.

while remaining agnostic as to whether or not these questionnaires measure them well. Researchers should self-assess and investigate further the validity of each measure for their research task at hand, as is the standard practice of scientific rigor. Furthermore, the questionnaire composition analysis does not precisely reflect the effect of the number of items. That is, a single-item questionnaire would be 100% dispositional, learned, or situational. But that does not necessarily mean that it is the best questionnaire to measure that specific trust layer. In addition, the trust layer categorization was based on the specific questionnaire's purpose, however, the same question can be used to assess different layers of trust depending on the context and the time it was administered (i.e., before, during, or after an interaction) (Merritt and Ilgen, 2008).

Third, our categorization of the trust questionnaires' domains is rather generic. For example, some automation trust questionnaires are targeted at trust in automated vehicles specifically whereas others are targeted at trust in automation in general. We tackle this in the analysis by removing system-specific terms (e.g., robot, vehicle, website), however, questionnaire specificity remains an important consideration in questionnaire selection. A questionnaire that assesses trust in an assistive robot might have more human-like questions that do not necessarily translate to trust in an automated vehicle.

Fourth, the similarity found between the trust in automation and trust in e-commerce questionnaires may have been due to having similar theoretical origins; both categories assess trust in some type of technology or trust in an entity mediated by technology (Ghazizadeh et al., 2012) – to understand trust-related decisions such as reliance or purchasing. We are not claiming that this is a novel finding, and indeed assessing the history of these questionnaires would lead to a similar insight. However, our approach reveals this relationship through a quantitative, systematic analysis that shows researchers across multiple domains have similarly operationalized the construct of trust.

Fifth, our approach to characterizing the similarities and differences between questionnaires was data-driven. Data-driven approaches have been proven useful in expanding knowledge and extracting scientific relationships years in advance of their discovery (Tshitoyan et al., 2019). Yet, trust is a complex, multifaceted construct and future work should incorporate a theory-driven approach to safeguard the theoretical underpinnings of trust, expand trust theory,

and build on existing measures (McCroskey and Young, 1979; Long et al., 2020).

Sixth, our analyses leveraged one method of text analysis: word embeddings. Although this approach has demonstrated effectiveness and has been widely employed in various applications, it is important for future studies to explore alternative methodologies, such as topic modeling. Hierarchical topic modeling, in particular, can prove valuable, as it preserves the structure of the word-item-questionnaire and offer a different perspective on the data.

Seventh, we are potentially missing some questionnaires – a more comprehensive review might have revealed more relevant questionnaires, such as studies that focus on information credibility that may be related to trust (Renn and Levine, 1991; Fogg and Tseng, 1999; McCroskey and Teven, 1999). However, the methods used in this paper are scalable and could be easily applied to an expanded corpus if new questionnaires are developed or to explore a broader conception of trust. Futhermore, one limitation of the study is the inclusion of questionnaires in the corpus that were developed *ad hoc*, relying primarily on face validity, or were not empirically validated (Gutzwiller et al., 2019). This may lead to potential issues for item and questionnaire selection.

Finally, one important limitation of this study is that it only included questionnaires developed in English. Excluding trust questionnaires that might have been developed in another language has important implications for advancing trust theory and methods of trust measurement across languages and cultures, but also for generalizability of the trust lexicon.

5. Conclusion

This study demonstrates the potential of text analysis in understanding trust questionnaires in different contexts, which provides a systematic method to quantify similarities and differences for further survey development and questionnaire selection.

The analyses conducted were at the word, item, and questionnaire levels. Each highlighted important considerations of questionnaire development and selection. The word-level analysis showed the most common words and themes that emerged from the trust questionnaires literature and produces a trust lexicon. This has implications for questionnaire development and understanding of trust in conversational speech and public attitudes on social media (Pang and Lee, 2008; Li et al., 2020). Furthermore, the item and questionnaire analyses provided higher-level insights into questionnaire items composition, and questionnaire items spread across the semantic space, both of which are important considerations for questionnaire selection.

While this study focused on text-based trust questionnaires, this approach can be extended to more specific domains; such as estimating drivers' trust in self-driving vehicles through speech using the developed trust lexicon, similar to previous work on emotion classification (Aman and Szpakowicz, 2007) and the analysis of open-ended survey responses (Lee and Kolodge, 2018).

Overall, word-embedding text analysis is a useful way to understand the sentiments and emotions associated with words. The resulting semantic space of trust words provides a way to compare and select trust questionnaires. In addition, the resulting lexicon of trust-related words can be used in natural language processing to understand trust attitudes through conversations between people, and between people and technologies in different domains.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

AA and JL conceived of the presented idea. AA and ML conducted the questionnaires literature review. ML and EC contributed significantly to the planning and direction of the paper. AA conducted the analysis and developed the web app. All authors discussed the results and contributed to the final manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1192020/full#supplementary-material

References

Ajzen, I., and Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Upper Saddle River, NJ: Prentice Hall.

Albert, W., Gribbons, W., and Almadas, J. (2009). Pre-conscious assessment of trust: a case study of financial and health care web sites. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 53, 449–453. doi: 10.1177/154193120905300603

Allen, K., Bergin, R., and Pickar, K. (2004). Exploring trust, group satisfaction and performance in geographically dispersed and co-located university technology commercialization teams. Education that works: the NCIIA 8th annual meeting, University of Florida; Gainesville, FL, USA.

Alsaid, A., and Lee, J. D. (2022). The DataScope: a mixed-initiative architecture for data labeling. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 66, 1559–1563. doi: 10.1177/1071181322661356

Alsaid, A., Lee, J. D., Noejovich, S. I., and Chehade, A. (2023). The effect of vehicle automation styles on drivers' emotional state. *IEEE Trans. Intell. Transp. Syst.* 24, 3963–3973. doi: 10.1109/TITS.2023.3239880

Alsaid, A., Lee, J. D., Roberts, D. M., Barrigan, D., and Baldwin, C. L. (2018). Looking at mind wandering during driving through the windows of PCA and t-SNE. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 62, 1863–1867. doi: 10.1177/1541931218621424

Altszyler, E., Sigman, M., Ribeiro, S., and Slezak, D. F. (2016). Comparative study of LSA vs word2vec embeddings in small corpora: a case study in dreams database. *Conscious Cogn.* 56, 178–187. doi: 10.1016/j.concog.2017.09.004

Aman, S., and Szpakowicz, S. (2007). "Identifying expressions of emotion in text" in *Text, speech and dialogue*. eds. V. Matoušek and P. Mautner (Berlin, Heidelberg: Springer), 196–205.

Benedek, J., and Miner, T. (2002). Measuring desirability: new methods for evaluating desirability in a usability lab setting. Proceedings of usability professionals association, ScienceOpen. Orlando. 8–12.

Bhattacherjee, A. (2002). Individual trust in online firms: scale development and initial test. *J. Manag. Inf. Syst.* 19, 211–241. doi: 10.1080/07421222.2002.11045715

Byrne, K., and Marín, C. (2018). Human trust in robots when performing a service. 2018 IEEE 27th International conference on enabling technologies: infrastructure for collaborative enterprises (WETICE). IEEE. Paris, France

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 81–105. doi: 10.1037/h0046016

Cassell, J., and Bickmore, T. (2000). External manifestations of trustworthiness in the interface. *Commun. ACM* 43, 50–56. doi: 10.1145/355112.355123

Chancey, E. T., Bliss, J. P., Yamani, Y., and Handley, H. A. H. (2017). Trust and the compliance-reliance paradigm: the effects of risk, error bias, and reliability on trust and dependence. *Hum. Factors* 59, 333–345. doi: 10.1177/0018720816682648

Charalambous, G., Fletcher, S., and Webb, P. (2016). The development of a scale to evaluate trust in industrial human-robot collaboration. *Int. J. Soc. Robot.* 8, 193–209. doi: 10.1007/s12369-015-0333-8

Chien, S.-Y., Semnani-Azad, Z., Lewis, M., and Sycara, K. (2014). "Towards the ddevelopment of an inter-cultural scale to measure trust in automation" in *Cross-cultural design*. ed. P. L. P. Rau (Cham: Springer), 35–46.

Chiou, E. K., and Lee, J. D. (2021). Trusting automation: designing for responsivity and resilience. *Hum Factors* 65, 137–165. doi: 10.1177/00187208211009995

Chita-Tegmark, M., Law, T., Rabb, N., and Scheutz, M. (2021). Can you trust your trust measure?. Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (New York, USA: ACM), 92–100.

Davies, M., and Gardner, D. (2010). *A frequency dictionary of contemporary american English*. Routledge. London

Delgado-Ballester, E., Munuera-Aleman, J. L., and Yague-Guillen, M. J. (2003). Development and validation of a brand trust scale. *Int. J. Mark. Res.* 45, 35–54. doi: 10.1177/147078530304500103

Domeyer, J., Venkatraman, V., Price, M., and Lee, J. D. (2018). Characterizing driver trust in vehicle control algorithm parameters. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 62, 1821–1825. doi: 10.1177/1541931218621413

Dzindolet, M., Peterson, S., Pomranky, R., Pierce, L., and Beck, H. (2003). The role of trust in automation reliance. *Int. J. Hum. Comput. Stud.* 58, 697–718. doi: 10.1016/S1071-5819(03)00038-7

Evans, A. M., and Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *J. Res. Pers.* 42, 1585–1593. doi: 10.1016/j.jrp.2008.07.011

Fast, E., Chen, B., and Bernstein, M. S. (2017). Lexicons on demand: neural word embeddings for large-scale text analysis. IJCAI'17: proceedings of the 26th international joint conference on artificial intelligence. AAAI Press. Washington, DC

Fogg, B. J., and Tseng, H. (1999). The elements of computer credibility. Proceedings of the SIGCHI conference on human factors in computing systems the CHI is the limit - CHI '99, (New York: ACM Press), 80–87.

Frazier, M. L., Johnson, P. D., and Fainshmidt, S. (2013). Development and validation of a propensity to trust scale. *J. Trust Res.* 3, 76–97. doi: 10.1080/21515581.2013.820026

Fuchs, C., and Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: conceptual issues and application guidelines. *Die Betriebswirtschaft* 69, 195–210.

Gefen, D., Karahanna, E., and Straub, D. W. (2003). Trust and tam in online shopping: an integrated model. *MIS Q.* 27, 51–90. doi: 10.2307/30036519

Gefen, D., and Straub, D. W. (2004). Consumer trust in B2C e-commerce and the importance of social presence: experiments in e-products and e-services. *Omega* 32, 407–424. doi: 10.1016/j.omega.2004.01.006

Ghazizadeh, M., Lee, J. D., and Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cogn. Tech. Work* 14, 39–49. doi: 10.1007/s10111-011-0194-3

Goillau, P., Kelly, C., Boardman, M., and Jeannot, E. (2003). *Guidelines for trust in future atm systems-measures*. Brussels, Belgium: EUROCONTROL, the European Organisation for the Safety of Air Navigation

Goto, S. G. (1996). To trust or not to trust: situational and dispositional determinants. *Soc. Behav. Pers.* 24, 119–131. doi: 10.2224/sbp.1996.24.2.119

Grant, M. J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf. Libr. J.* 26, 91–108. doi: 10.1111/j.1471-1842.2009.00848.x

Gutzwiller, R. S., Chiou, E. K., Craig, S. D., Lewis, C. M., Lematta, G. J., Hsiung, C. P., et al. (2019). Positive bias in the 'Trust in Automated Systems Survey'? An examination of the Jian et al. (2000) scale. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 63, 217–221. doi: 10.1177/1071181319631201

Hoff, K. A., and Bashir, M. (2014). Trust in automation: integrating empirical evidence on factors that influence trust. *Hum. Factors* 57, 407–434. doi: 10.1177/0018720814547570

Holthausen, B. E., Wintersberger, P., Walker, B. N., and Riener, A. (2020). Situational trust scale for automated driving (STS-AD): development and initial validation. International conference on automotive user interfaces and interactive vehicular applications, 40–47. Association for Computing Machinery. New York

Hu, W., Zhang, J., and Zheng, N. (2016). Different contexts lead to different word embeddings. COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers. The COLING 2016 Organizing Committee. Osaka, Japan

Hubert, M., Rousseeuw, P. J., and Vanden Branden, K. (2005). ROBPCA: a new approach to robust principal component analysis. *Technometrics* 47, 64–79. doi: 10.1198/004017004000000563

Jeong, H., Park, J., Park, J., Pham, T., and Lee, B. C. (2018). "Analysis of trust in automation survey instruments using semantic network analysis" in *Advances in human factors and systems interaction. AHFE 2018. Advances in intelligent systems and computing*. ed. I. Nunes (Cham: Springer)

Jian, J.-Y., Bisantz, A. M., Drury, C. G., and Llinas, J. (2000). Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* 4, 53–71. doi: 10.1207/S15327566IJCE0401_04

Johnson-George, C., and Swap, W. C. (1982). Measurement of specific interpersonal trust: construction and validation of a scale to assess trust in a specific other. *J. Pers. Soc. Psychol.* 43, 1306–1317. doi: 10.1037/0022-3514.43.6.1306

Khalid, H. M., Shiung, L. W., Nooralishahi, P., Rasool, Z., Helander, M. G., Kiong, L. C., et al. (2016). Exploring psycho-physiological correlates to trust: implications for human-robot-human interaction. Proceedings of the human factors and ergonomics society annual meeting. SAGE Publications Sage CA: Los Angeles, CA

Kohn, S. C., de Visser, E. J., Wiese, E., Lee, Y.-C., and Shaw, T. H. (2021). Measurement of trust in automation: a narrative review and reference guide. *Front. Psychol.* 12:604977. doi: 10.3389/fpsyg.2021.604977

Konopka, T. (2019). *Uniform manifold approximation and projection*. R package version, 0.2.7.0

Körber, M. (2018). Theoretical considerations and development of a questionnaire to measure trust in automation. Congress of the International Ergonomics Association, Springer, Cham. 13–30.

Körber, M., Prasch, L., and Bengler, K. (2018). Why do I have to drive now? Post hoc explanations of takeover requests. *Hum. Factors* 60, 305–323. doi: 10.1177/0018720817747730

Landauer, T. K., and Dumais, S. (2008). Latent semantic analysis. *Annu. rev. inf. sci. technol.* 3:4356. doi: 10.4249/scholarpedia.4356

Lankton, N. K., Harrison Mcknight, D., and Tripp, J. (2015). Technology, humanness, and trust: rethinking trust in technology. *J. Assoc. Inf. Syst.* 16, 880–918. doi: 10.17705/1jais.00411

Larzelere, R., and Huston, T. (1980). The dyadic trust scale: toward understanding interpersonal trust in close relationships. *J. Marriage Fam.* 42, 595–604. doi: 10.2307/351903

Lee, J. D., and Kolodge, K. (2018). Understanding attitudes towards self-driving vehicles: quantitative analysis of qualitative data. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 62, 1399–1403. doi: 10.1177/1541931218621319

Lee, J. D., Liu, S.-Y., Domeyer, J., and DinparastDjadid, A. (2019). Assessing drivers' trust of automated vehicle driving styles with a two-part mixed model of intervention tendency and magnitude. *Hum. Factors* 63, 197–209. doi: 10.1177/0018720819880363

Lee, J. D., and Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 1243–1270. doi: 10.1080/00140139208967392

Lee, J. D., and See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Hum. Factors* 46, 50–80. doi: 10.1518/hfes.46.1.50.30392

Lewis, J. D., and Weigert, A. J. (2012). The social dynamics of trust: theoretical and empirical research, 1985-2012. *Soc. Forces* 91, 25–31. doi: 10.1093/sf/sos116

Li, M., Alsaid, A., Noejovich, S. I., Cross, E. V., and Lee, J. D. (2020). Towards a conversational measure of trust. Available at: http://arxiv.org/abs/2010.04885.

Long, S. K., Sato, T., Millner, N., Loranger, R., Mirabelli, J., Xu, V., et al. (2020). Empirically and theoretically driven scales on automation trust: a multi-level confirmatory factor analysis. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 64, 1829–1832. doi: 10.1177/1071181320641440

Madsen, M., and Gregor, S. (2000). Measuring human-computer trust. Proceedings of eleventh Australasian Conference on Information Systems, Melbourne, Australia: AIS.

Malle, B. F., and Ullman, D. (2021). "A multidimensional conception and measure of human-robot trust" in *Trust in human-robot interaction: research and applications*. eds. C. S. Nam and J. B. Lyons (Amsterdam: Elsevier)

Mayer, R. C., and Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: a field quasi-experiment. *J. Appl. Psychol.* 84, 123–136. doi: 10.1037/0021-9010.84.1.123

Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An integrative model of organizational trust. *Acad. Manag. Rev.* 20, 709–734. doi: 10.2307/258792

McCroskey, J. C., and Teven, J. J. (1999). Goodwill: a reexamination of the construct and its measurement. *Commun. Monogr.* 66, 90–103. doi: 10.1080/03637759909376464

McCroskey, J. C., and Young, T. J. (1979). The use and abuse of factor analysis in communication research. *Hum. Commun. Res.* 5, 375–382. doi: 10.1111/j.1468-2958.1979.tb00651.x

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. *J. Open Source Softw.* 3:861. doi: 10.21105/joss.00861

Mcknight, D. H., Carter, M., Thatcher, J. B., and Clay, P. F. (2011). Trust in a specific technology: an investigation of its components and measures. *ACM Trans. Manag. Inf. Syst.* 2, 1–25. doi: 10.1145/1985347.1985353

McKnight, D. H., Choudhury, V., and Kacmar, C. (2002). Developing and validating trust measures for e-commerce: an integrative typology. *Inf. Syst. Res.* 13, 334–359. doi: 10.1287/isre.13.3.334.81

Merritt, S. M. (2011). Affective processes in human-automation interactions. *Hum. Factors* 53, 356–370. doi: 10.1177/0018720811411912

Merritt, S. M., and Ilgen, D. R. (2008). Not all trust is created equal: dispositional and history-based trust in human-automation interactions. *Hum. Factors* 50, 194–210. doi: 10.1518/001872008X288574

Merritt, S. M., Unnerstall, J. L., Lee, D., and Huber, K. (2015). Measuring individual differences in the perfect automation schema. *Hum. Factors* 57, 740–753. doi: 10.1177/0018720815581247

Meyer, J., and Lee, J. D. (2013). "Trust, reliance, and compliance" in *The Oxford handbook of cognitive engineering*. ed. A. Kirlik (Oxford: Oxford University Press), 109–124.

Michael, S. (1979). Attitude measurement: why and how. *Educ. Technol.* 19, 34–38.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations ofwords and phrases and their compositionality. Proceedings of the 26th International Conference on Neural Information Processing Systems. Curran Associates Inc. Red Hook, NY, United States

Miller, A. S., and Mitamura, T. (2003). Are surveys on trust trustworthy? *Soc. Psychol. Q.* 66:62. doi: 10.2307/3090141

Moeckli, J., Brown, T., Dow, B., Boyle, L. N., Schwarz, C., and Xiong, H. (2015). *Evaluation of adaptive cruise control interface requirements on the national advanced driving simulator*. Washington, DC: National Highway Traffic Safety Administration.

Mohammad, S. M., and Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Comput. Intell.* 29, 436–465. doi: 10.1111/j.1467-8640.2012.00460.x

Monroe, B. L., Colaresi, M. P., and Quinn, K. M. (2008). Fightin' words: lexical feature selection and evaluation for identifying the content of political conflict. *Polit. Anal.* 16, 372–403. doi: 10.1093/pan/mpn018

Montague, E. (2010). Validation of a trust in medical technology instrument. *Appl. Ergon.* 41, 812–821. doi: 10.1016/j.apergo.2010.01.009

Muir, B., and Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 429–460. doi: 10.1080/00140139608964474

Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* 2, 1–135. doi: 10.1561/1500000001

Patel, F. N. (2016). Large high dimensional data handling using data reduction. In 2016 international conference on electrical, electronics, and optimization techniques (ICEEOT) (IEEE. Chennai, India), 1531–1536.

Pennington, J., Socher, R., and Manning, C. D. (2014). GloVe: global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Association for Computational Linguistics. Doha, Qatar

Price, M., Lee, J. D., Dinparastdjadid, A., Toyoda, H., and Domeyer, J. (2017). Effect of vehicle control algorithms on eye behavior in highly automated vehicles. Paper presented at fourth international symposium on future active safety technology Sage Publication. Nara, Japan

R Development Core Team (2016). *R: a language and environment for statistical computing*. Vienna, Austria R Foundation for Statistical Computing

Rempel, J. K., Holmes, J. G., and Zanna, M. P. (1985). Trust in close relationships. *J. Pers. Soc. Psychol.* 49, 95–112. doi: 10.1037/0022-3514.49.1.95

Renn, O., and Levine, D. (1991). "Credibility and trust in risk communication" in *Communicating risks to the public*. eds. R. E. Kasperson and P. J. M. Stallen (Dordrecht: Springer Netherlands), 175–217.

Rotter, J. B. J. (1967). A new scale for the measurement of interpersonal trust. *J. Pers.* 35, 651–665. doi: 10.1111/j.1467-6494.1967.tb01454.x

Salcedo, J. N., Ortiz, E. C., Lackey, S. J., Hudson, I., and Taylor, A. H. (2011). Effects of autonomous vs. remotely-operated unmanned weapon systems on human-robot teamwork and trust. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 55, 635–639. doi: 10.1177/1071181311551130

Schaefer, K. E. (2013). *The perception and measurement of human-robot trust*. Ph.D. dissertation, University of Central Florida Orlando, Florida.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schneider, T. R., Jessup, S. A., Rivers, S., Lohani, M., and McCoy, M. (2017). *The influence of trust propensity on behavioral trust*. APS Boston, MA, USA.

Schwarz, N., and Bohner, G. (2007). "The construction of attitudes" in *Blackwell handbook of social psychology: intraindividual processes*. ed. A. Tesser (Malden, Massachusetts, USA: Blackwell Publishers Inc.), 436–457.

Sheridan, T. B. (2019). Individual differences in attributes of trust in automation: measurement and application to system design. *Front. Psychol.* 10:1117. doi: 10.3389/fpsyg.2019.01117

Silge, J., and Robinson, D. (2016). tidytext: text mining and analysis using tidy data principles in R. *J. Open Sour. Softw.* 1:37. doi: 10.21105/joss.00037

Singh, I. L., Molloy, R., and Parasuraman, R. (1993). Automation-induced "complacency": development of the complacency-potential rating scale. *Int. J. Aviat. Psychol.* 3, 111–122. doi: 10.1207/s15327108ijap0302_2

Sohaib, O., and Kang, K. (2015). Individual level culture influence on online consumer iTrust aspects towards purchase intention across cultures: a SOR model. *Int. J. Electron. Bus.* 12, 142–161. doi: 10.1504/IJEB.2015.069104

Takayama, L. (2009). Making sense of agentic objects and teleoperation: in-the-moment and reflective perspectives. 2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE. La Jolla, CA, USA

Tausczik, Y. R., and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* 29, 24–54. doi: 10.1177/0261927X09351676

Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., et al. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* 571, 95–98. doi: 10.1038/s41586-019-1335-8

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605. doi: 10.1007/s10479-011-0841-3

Walliser, J. C., De Visser, E. J., and Shaw, T. H. (2016). Application of a system-wide trust strategy when supervising multiple autonomous agents. *Proc. Hum. Fact. Ergon. Soc.* 60, 133–137. doi: 10.1177/1541931213601031

Wang, Y., Huang, H., Rudin, C., and Shaposhnik, Y. (2021). Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization. *J. Mach. Learn. Res.* 22, 9129–9201. doi: 10.48550/arXiv.2012.04456

Wickham, H. (2016). *Tidyverse: easily install and load*. "Tidyverse" packages.

Wickham, H., and Winston, C. (2019). *ggplot2: create elegant data visualisations using the grammar of graphics*. Available at: https://CRAN.R-project.org/package=ggplot2

Wiczorek, R., and Manzey, D. (2014). Supporting attention allocation in multitask environments: effects of likelihood alarm systems on trust, behavior, and performance. *Hum. Factors* 56, 1209–1221. doi: 10.1177/0018720814528534

Wrightsman, L. S. (1964). Measurement of philosophies of human nature. *Psychol. Rep.* 14, 743–751. doi: 10.2466/pr0.1964.14.3.743

Yagoda, R. E., and Gillan, D. J. (2012). You want me to trust a ROBOT? The development of a human-robot interaction trust scale. *Int. J. Soc. Robot.* 4, 235–248. doi: 10.1007/s12369-012-0144-0

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* 51, 110–116. doi: 10.1037/0022-3514.51.1.110

Yamagishi, T., and Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motiv. Emot.* 18, 129–166. doi: 10.1007/BF02249397

Yang, X. J., Schemanske, C., and Searle, C. (2021). Toward quantifying trust dynamics: how people adjust their trust after moment-to-moment interaction with automation. *Hum. Factors* 65, 862–878. doi: 10.1177/00187208211034716