Master Thesis

An exploratory study using text mining methods to investigate online therapists' feedback texts for the use of empathy.

Luis Matthias Behrendt-Richter

University of Twente

03.07.2023

Study: Positive Clinical Psychology & Technology

Track: Text Mining

Faculty: Behavioral Management & Social Sciences

Examination committee: Dr P.M. Ten Klooster

Dr E.J. De Bruin

Abstract

Introduction: The increasing demand for psychological healthcare services has led to the development of electronic mental health (eMental health) technologies, which provide treatment options for mental health issues via the Internet. However, the extent to which eMental health can contribute to the development of a working alliance between client and therapist is still questioned. This study explored the potential of natural language processing (NLP) methods, specifically text mining, to investigate therapists' use of empathy in online therapy feedback texts.

Methods: A secondary analysis was conducted based on the data from a randomized control trial (RCT), which explored the effectiveness of online-delivered cognitive behaviour therapy for adolescents with insomnia. Online feedback data from seven therapists were analysed using an exploratory text-mining approach. The Orange Data Mining and the Linguistic Inquiry and Word Count (LIWC-22) software were used for investigating the feedback for underlying topics, sentiments, the application of empathy and inter-therapist language style matching.

Results: The results of the topic modelling analyses revealed three strongly overlapping topics, all with a functional focus on sleep patterns and advice. The sentiment analyses indicated very similar and predominantly neutral sentiments across the seven therapists. The LIWC-22 analysis suggested a similar level of empathy expressed within the seven corpora of the seven therapists. Finally, the results of the language style matching analysis showed a very high linguistic similarity between all seven therapists, with language style matching scores ranging between 0.89 and 0.98.

Discussion: The results of the topic modelling, sentiment analyses and LIWC analyses suggest that the seven therapists in the current study communicated in a very similar, neutral, and functional style concerning their written feedback to their clients and used a similar amount of empathetic language. These similarities in the results could be attributed to the strict protocols within the settings of an RCT study. Therefore, future studies should investigate feedback data from real-world eMental health applications to explore the expression of empathy in a therapeutic context further. The current study suggests that NLP methods provide promising approaches to studying the application of empathy in a therapeutic context and proposes future research at an automated and larger scale.

Table of Content

Introduction

The Meaning of eHealth and eMental Health for Society

In recent decades, healthcare systems have faced the challenge of meeting increased demands for healthcare services while contending with a decrease in the number of professional health workers (Plaß et al., 2021). In order to support the healthcare sector, a research field known as eHealth has emerged in recent years, which focuses on the development of technological interventions to modify unhealthy behaviour and promote healthy behaviours (Van Gemert-Pijnen et al., 2018). The emergence of eHealth has revolutionized healthcare delivery, providing individuals with a range of flexible, cost-effective, and accessible treatment options (Andersson & Titov, 2014). A benefit of these treatment options is the autonomy that the technology provides to patients and healthcare professionals in managing medical problems (Flinsenberg, 2020). In addition to a reduction in human communication errors (Muthiah et al., 2019) eHealth also has the advantage of scalability, which allows healthcare providers to deliver interventions remotely and simultaneously to a large number of patients, thus preventing hospitalisation and reducing staff costs when applied in a timely manner (Duettmann et al., 2021). Thereby, eHealth offers a promising approach to address the challenges faced by modern healthcare systems. The scope of eHealth covers all health-related digital information as well as technological interventions that potentially support patients and healthcare professionals in their daily activities (Van Gemert-Pijnen et al., 2018). Within this scope, subfields of eHealth emerged that specialised in the development of technological support for particular fields of healthcare.

One such subfield is eMental health, which focuses on developing technologies that support patients with mental health conditions (Ellis et al.,2021). Christensen and Griffiths (2003, p.3) defined this field as "mental health services and information delivered or enhanced through the internet and related technologies". Through channels such as smartphone applications, informational websites, telehealth and chat portals, self-help interventions as well as virtual and blended therapies can be provided. (Ellis et al., 2021). Various studies have demonstrated the efficacy of eMental health interventions. Blended therapy, for example, combines online therapy with conventional face-to-face therapy and showed significant improvements for individuals suffering from a range of conditions including depression and the debilitating effects of post-traumatic stress disorder (PTSD) (Kooistra et al, 2020). Further, virtual, solely online-based exposure therapy was proven to be effective in treating PTSD as well as other anxiety disorders (Carl et al., 2019). Finally, studies on self-help interventions, where no therapeutic guidance is necessary, provide evidence of reduced anxiety and depression symptoms (Al-Alawi et al. (2021).

Specific and Nonspecific Treatment Mechanisms in the Therapeutic Context

Concerning the implementation of psychotherapy in an online therapeutic context, especially in the beginning traditional face-to-face therapy and diagnostic tests were literally translated and transferred into an online environment (Luteijn & Barelds, 2019). However, the effects of traditional face-to-face therapy are assumed to be based on both specific and non-specific mechanisms, with especially the latter being challenging to transfer to the online environment without adaptation.

On the one hand, specific mechanisms are treatment methods that are directly applied to a specific illness pattern such as, for example, cognitive behaviour therapy for anxiety treatment or exposure therapy for PTSD (Butler et al., 2021). These mechanisms are designed to directly target the patient's symptoms and to lead to quick improvement of the client's mental health (Leichsenring et al., 2018). However, specific mechanisms may not be effective for all patients as they do not consider underlying client issues, which is why some patients may require more personalized or integrated approaches (Butler et al., 2021).

On the other hand, there are non-specific mechanisms, most of which are related to the therapist’s interpersonal skills (Hadjistavropoulos et al., 2017). Examples are empathy, warmth, and the building of a therapeutic alliance, all of which do not focus on a specific illness pattern but intend to foster a good relationship between therapist and patient (Hadjistavropoulos et al., 2017). Non-specific mechanisms are particularly helpful for patients who are hesitant or reluctant to seek treatment (Watson et al., 2016), and they are significant predictors of treatment adherence according to a study by Flückiger and colleagues (2018). On top, the therapeutic relationship accounts for up to 30% of the variance in treatment outcomes (Flückiger et al., 2018). However, nonspecific mechanisms may not lead to significant improvement in the patient's mental health on their own and may need to be combined with specific mechanisms to be effective (Watson et al., 2016).

The Meaning of Empathy for Working Alliance Development

Although non-specific mechanisms alone do not account for the success of psychotherapy, they still are assumed to have a major influence on the degree to which the therapeutic goals are achieved. A good working alliance between therapist and client forms the basis for trust and cooperation in therapy, which can positively impact treatment adherence and treatment results (Watts et al., 2020). A meta-analysis of more than 200 studies by Horvath et al. (2011) underlines this since it found a moderate correlation between treatment outcomes and the formation of a working alliance regardless of the therapy orientation, the rating perspective, the alliance measure and the time of assessment. Therefore, forming and maintaining a good working alliance is important for therapists and clients as it has a relevant impact on the success of therapy.

Empathy is considered an essential non-specific mechanism and an important skill for therapists when it comes to building a good working alliance with clients (Rogers, 1975). The early pioneer of empathic research Carl Rogers (1975, p. 2) defined empathy as "to perceive the internal frame of reference of another with accuracy and with the emotional components and meanings which pertain thereto as if one were the person, but without ever losing the 'as if' condition". To express this understanding for another person in psychotherapy, empathy demands the therapist to comprehend the client's experiences, emotions, and perspectives, without judgment or bias (Rogers, 2019). Furthermore, empathy not only positively impacts the therapist's recognition of the patient's needs and concerns, but also helps therapists to convey a sense of security and trust to the client in order to form an interpersonal connection. A study by Nienhuis et al. (2018) confirms the importance of empathy in forming a therapeutic alliance. In their meta-analysis of more than 50 studies, a moderate relationship was found between empathy and the development of a therapeutic alliance (Nienhuis et al., 2018). A study by Moyers and Miller (2013) found similar results as it demonstrated empathic listening as a predictor of client satisfaction due to a significant decrease in client resistance in a counselling setting. Finally, also Elliot et al. (2011) conducted a meta-analysis in which empathy was identified as a significant predictor for the success of several therapy forms, including cognitive behavioural therapy, and psychodynamic therapy.

Different Approaches to Measure Empathy in a Psychotherapeutic Context

The investigation of therapist use of empathy can be conducted through several approaches including self-report questionnaires, observer ratings or natural language processing (NLP). Self-report questionnaires are developed to assess the therapists' subjective and self-perceived estimation of their ability to understand the clients' emotions and experiences. An example of a self-report empathy questionnaire is the Toronto Empathy Questionnaire (TEQ) designed by Spreng, McKinnon, Mar, and Levine (2009). The TEQ is designed to measure different facets of empathy like perspective-taking or empathic concern from the perspective of the therapist in a therapeutic setting (Spreng et al., 2009).

Furthermore, observer ratings are designed to measure the therapists' actual non-verbal and verbal expressions of empathy, such as nodding, maintaining eye contact, and using reflective listening skills (Greenberg et al., 2001). This observation can take place from two different perspectives, either from the client’s point of view or from a third-party perspective. An example of a client-related observer rating is the empathy scale of the Barrett-Lennard Relationship Inventory (BLRI), which assesses the client's perception of the therapist's use of positive regard, empathy, and congruence (Greenberg et al., 2001). Greenberg and colleagues (2001) concluded that client-based observer ratings are a better predictor of therapy outcomes than third-party observer ratings or self-assessments carried out by the therapist.

Finally, NLP methods offer language-based measures which analyse the language used by clients and therapists during a therapy session (Goldberg et al., 2020). Especially machine learning techniques can be used to identify patterns associated with a specific topic like empathy or working alliance (Litvak et al., 2016). This approach is still in its early stages but has already shown promise as an objective and less biased method to measure underlying concepts within language data (Goldberg et al., 2020). For example, in the study of Goldberg and colleagues (2020), a machine learning algorithm learned the association between therapeutic alliances and patient ratings based on the linguistic content of more than 1200 therapy conversations. In the end, the artificial intelligence could modestly predict the working alliance ratings, which among others also assessed the client's perception of empathy (Goldberg et al., 2020).

Although the concept of empathy can be investigated through different perspectives, the NLP method, which makes use of machine learning techniques and artificial intelligence, offers an interesting approach to exploring natural language data for the application of therapists' use of empathy. Although NLP methods cannot take non-verbal empathy skills into account, it provides researchers with a more objective perspective on therapists' use of empathy than observer ratings. Nevertheless, despite some promising findings of Goldberg et al. (2020) and Litvak et al. (2016), more research is still needed to explore the usefulness of NLP methods for measuring and studying empathy in different therapeutic settings.

Text Mining

NLP methods have been increasingly used in a scientific context to explore large amounts of text data in recent years. According to Chowdhary (2020, p.1), NLP is an "area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things". Therefore, NLP researchers focus on understanding human beings' use of language in order to develop techniques and tools that can create an understanding of the natural language for computer systems. Based on that understanding, computers are then able to analyse and manipulate natural language (Chowdhary, 2020).

One NLP method which is especially used to analyse the content of online therapy conversations is called text mining (Veiga, 2019). Text mining can be defined as "the process of extracting interesting information and knowledge from unstructured text" (Hotho et al., 2005, p.1). Through the application of machine learning and computational linguistics techniques, researchers can transform unstructured text data into structured data like tables and graphs (Feldman & Sanger, 2007). Different computational linguistics techniques like sentiment analysis or topic modelling analysis can extract information about underlying concepts within the text or the author's use of specific sentiments (Feldman & Sanger, 2007). For example, topic modelling analysis generates probabilistic models to discover hidden themes based on the vocabulary used and its occurrence in a document (Tong & Zhang, 2016). Sentiment analysis is an unsupervised NLP method analysing feelings, thoughts, and opinions based on a previously developed dictionary and the number of specific words used (Öztürk & Ayvaz, 2018).

Text Mining Approaches for Measuring Empathy in a Therapeutic Context

In recent years, a few studies have already used text mining methods to analyse text-based conversations for the presence of empathy. Nonetheless, the study of Mawani and Nderu (2020) is so far the only study that used text-mining to explore empathy in a therapeutic setting. In their study, text mining was used to analyse the content of counsellor-client messages exchanged on an online web-based counsellor platform. The researchers used a combination of machine learning and text analytics to explore and identify empathic language and expressions of empathy within the messages. Based on a lexical analysis, an artificial intelligence (AI) algorithm was then created which included empathic chat functionalities. Thereby the AI could generate targeted empathic responses based on the learned definition of empathy from the counsellor-client messages (Mawani & Nderu, 2020).

Furthermore, a study by Litvak, Otterbacher, Ang, and Atkins (2016) explored the relationship between social and linguistic behaviour as well as the role of empathy in facilitating interactions on social media with a text-mining approach. The study utilized the IRL David Scale, which is an automated language analysis tool that has been validated in previous studies and has been shown to accurately measure language characteristics related to mental health conditions, such as depression and anxiety (Rude et al., 2004). In the study by Litvak et al. (2016), the IRL David Scale, as well as the Linguistic Inquiry and Word Count (LIWC-22) software, were used to analyse the language use of participants with high and low levels of trait empathy during a social interaction task. The results showed that participants with high levels of the trait empathy scored higher on the social processing and emotional perception scales of the LIWC-22 analysis, indicating a more socially and linguistically engaged communication style in their language use. The findings suggest that automated text-mining approaches, like the LIWC-22 analysis, may have the potential to detect empathy in texts.

Cognitive Behavioural Therapy for Insomnia and the Role of the Therapists

A study by De Bruin et al. (2015) offers a data set which is useful to further investigate the potential of NLP methods for the exploration of therapist empathy in the online therapeutic context. The study aimed to investigate the impact of asynchronous online therapeutic text feedback on specific outcome measures such as insomnia severity and sleep quality in internet-based cognitive behavioural therapy for adolescents with insomnia. Cognitive Behavioural Therapy for Insomnia (CBTi) is a treatment approach that focuses on changing undesired thoughts and behaviours which are related to sleep problems (De Bruin et al. 2015). In the study, participants were provided with an online CBTi program designed for adolescents over a time span of 6 weeks. Results indicated that the participants achieved significant improvements in sleep outcomes due to the CBTi treatment (De Bruin et al., 2015) and suggested that online therapy may be as effective as face-to-face-delivered feedback in this context. However, the topics and sentiments expressed in the specific feedback responses by the individual therapists were not a factor of interest in the study by De Bruin et al. (2015) to determine the efficacy of online CBTi for adolescents with insomnia. Therefore, this study focuses on the feedback texts that were created by seven therapists in the frame of the study of De Bruin et al. (2015).

Research Questions

Using the data from De Bruin et al. (2015), this study explores the extent to which therapists' online feedback can be meaningfully analysed with NLP methods for underlying topics and sentiments and the extent to which the two empathy elements "social processing" and "emotional perception" (Litvak et al., 2016) emerge from the written feedback. Thereby, the following research questions (RQ) are stated:

RQ 1: What latent topics emerge from the written feedback?

RQ 2: To which extent can different sentiments be discovered in the feedback provided by different therapists?

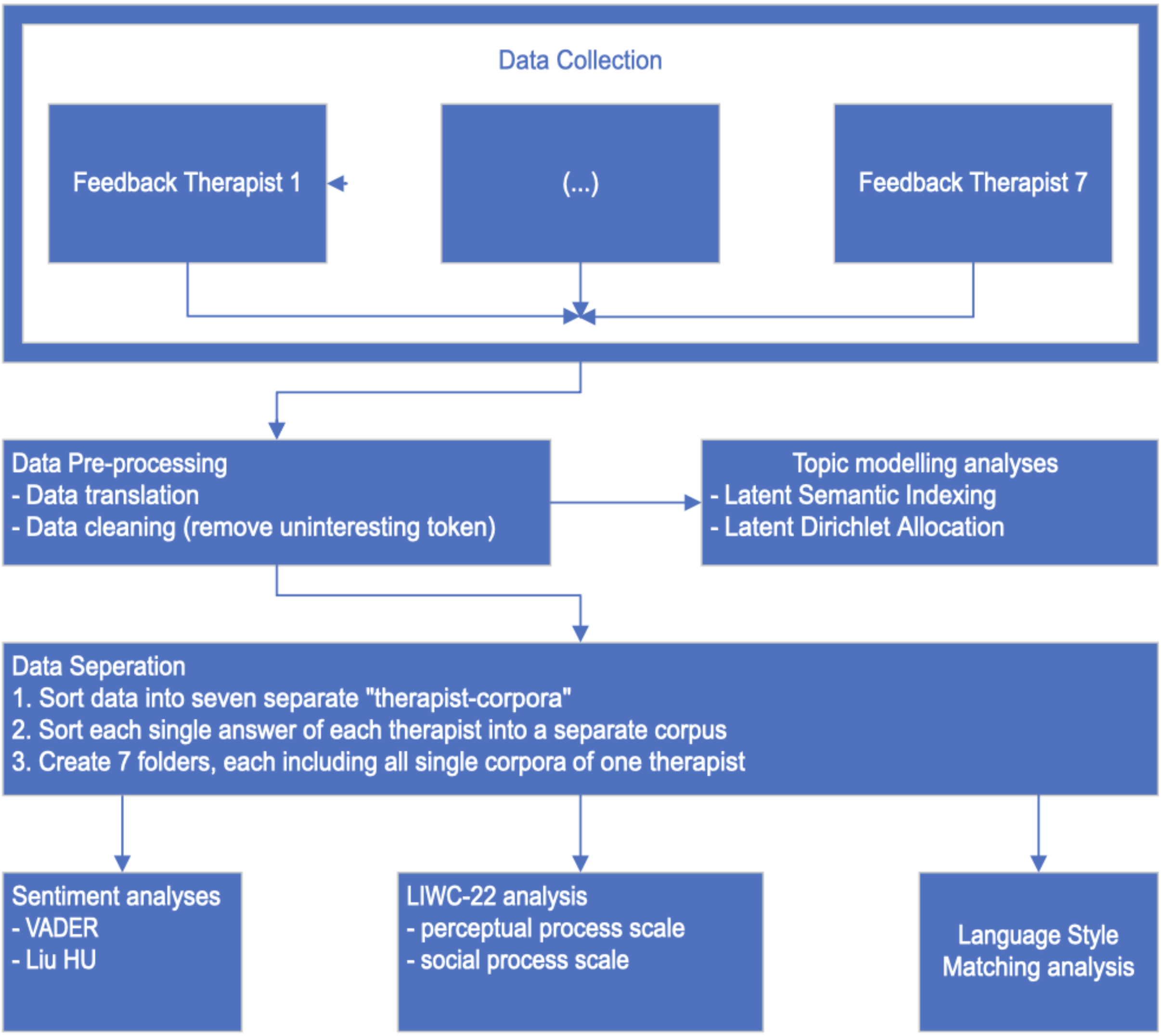
RQ 3: To what extent does the occurrence of the two empathy elements "social processing" and "emotional perception" differ between different therapists?

Methods

Text-mining is an iterative and exploratory process that consists of several subsequent steps from specifying the study design, over data collection and data pre-processing to finally the analyses performed in this study. Figure 1 shows the pipeline of tasks performed as well as their chronological order.

Figure 1

Flowchart



Note. VADER stands for Valence Aware Dictionary and Sentiment Reasoner. LIWC stands for Linguistic Inquiry and Word Count.

Study Design

The current study represents a secondary analysis of text data derived from the study by De Bruin et al. (2015). Their study was part of a larger trial that was approved by the medical ethical committee of the Academic Medical Center in Amsterdam and registered with the International Standard Randomized Controlled Trial Number Registry (ISRCTN33922163). The trial involved one treatment arm in which participants received CBTi via online sessions and additional asynchronous text-based online feedback that was delivered by seven therapists. The present study aims to employ text-mining techniques to inductively explore the feedback texts composed by the seven therapists. Specifically, this study explores the therapists' feedback for underlying concepts, investigates the extent to which different sentiments can be discovered in the text data and examines the extent to which the online therapists scored differently on the two empathy scales "social processing" and "emotional perception" described by Litvak et al. (2016). To accomplish these objectives, a qualitative exploratory text mining approach was used to analyse the vast amounts of textual data composed in the asynchronous therapy sessions.

Data Collection

In the study by De Bruin et al. (2015), adolescents with diagnosed insomnia were identified through the completion of an online screening questionnaire as well as a one-hour face-to-face diagnostic interview with a sleep therapist. Eligible participants as well as their parents provided informed consent, and parents received a booklet explaining the CBTi treatment protocol and its goals. A total of 57 participants received CBTi from seven therapists over a time span of six weeks. The treatment was designed to be completed independently by the adolescents, and therapists provided individual feedback to their clients on a weekly basis. The feedback messages written by the therapists were then collected and anonymized to ensure participant confidentiality. The messages mainly included weekly feedback about the client's bedtimes, personal advice, and instructions for exercises. On average, the individual messages were 330 words and 1403 signs long. For more detailed information see Table 1.

The seven therapists involved in the data collection for this study were trained specifically in providing online therapy to children with insomnia via an asynchronous text-based chat portal. Four of the therapists held a certification from the Nederlandse Vereniging voor Slaap Waak Onderzoek, while one was certified by the European Sleep Research Society and one was an expert in sleep research and cognitive behavioural therapy (De Bruin et al., 2015). Throughout the data collection period, the therapists were supervised on a weekly basis by an independent expert in sleep therapy. This supervision aimed to ensure the quality and consistency of the written feedback data collected from the therapists.

Table 1

Numerical characteristics of the feedback messages of the seven therapists

	Messages Total	Average words per message	Average characters per message
Therapist 1	35	134	574
Therapist 2	121	384	1640
Therapist 3	211	388	1637
Therapist 4	197	316	1344
Therapist 5	197	342	1464
Therapist 6	96	324	1395
Therapist 7	23	225	949

Materials

The current study employed several tools and techniques to analyse the data collected from the online CBTi intervention. Microsoft Word was used to open and sort the data and explore the individual therapists' number of responses and average word signs used. Furthermore, Microsoft Word was used to write exclusion lists, which are text files including lists of uninteresting words for the analysis. Thereby, these lists defined which words were not considered by the analysis programs. As the data were written in Dutch but not understood by the researcher, the messages corpus was translated from Dutch to English, by employing Google Translate (Google Translate, 2023). A dedicated Python (version 3) function was written to sort the data into separate corpora by therapist (Van Rossum & Drake, 2009). To conduct the natural language processing (NLP) analyses, the program Orange Data Mining version 3.0 was utilized (Demsar et al., 2013). Orange is a Python-based open-source platform for analysing and visualising text data and is mostly used in social sciences, healthcare, and marketing to gain insights from unstructured text data.

In order to conduct additional Sentiment analysis focussed specifically on assessing the therapists' use of empathy, the Linguistic Inquiry and Word Count (LIWC-22) software was used (Boyd, Ashokkumar, Seraj & Pennebaker, 2022). LIWC-22 is a text analysis software that identifies and categorizes words based on their linguistic features and was developed by James W. Pennebaker and his colleagues in the 1990s based on their research on the relationship between language use and psychological states. By analysing the frequency of words occurring in a text and considering the context, the software can assign the composer's word choice to more than 22 predefined categories like emotional tone, cognitive processes, and personal concerns.

Compared to the topic modelling and sentiment analysis options in Orange, LIWC-22 offers a more nuanced and extensive analysis of the content of text-based therapy conversations based on predefined psychological categories. While Orange allows for text cleaning, pre-processing, visualization, unsupervised and exploratory topic modelling and sentiment analysis, it does not provide the same level of detail in the analysis of the language used by therapists. LIWC-22, therefore, offers an additional standardized and validated measure of linguistic features that can be used to compare different texts and language samples. In this study, LIWC-22 was used to investigate the specific use of empathic language used by the therapists based on the two scales “social processing” and “emotional perception”. Additionally, it was used to perform an inter-therapist language style matching (LSM) analysis to further explore the extent to which the used language of the individual therapists overlapped with each other.

Data Pre-Processing

The current study analysed a corpus of text data that was intended for human interpretation, including grammatical elements such as punctuation, accents, function words, and capitalization. These elements of natural language facilitate effective communication between individuals by conveying information in a coherent and comprehensible manner. However, NLP methods are not as proficient in discerning between meaningful and irrelevant textual elements as humans are, necessitating the implementation of rigorous pre-processing of raw text data.

Text pre-processing is the process of preparing text data for utilization in text mining algorithms (Wesslen, 2018). This process involves standardizing text data and eliminating text, characters, and symbols that are not germane to the analysis (Danubianu, 2015). The cleaning of data during pre-processing streamlines text data into a suitable form for text mining algorithms, thereby augmenting the performance and quality of models derived from text mining procedures (Danubianu et al., 2015).

For the current study, the data was pre-processed by excluding all unnecessary information from the corpus and translating all text from Dutch into English. The data was then uploaded into Orange and inspected iteratively through the use of word clouds, which show the most used words in the text. Based on the visualization through the word clouds, additional irrelevant tokens such as numbers, individual letters, punctuation marks or the names of the clients that were not of interest for further analysis could be identified. The identified words, numbers, letters, and symbols were listed in a text file which later served as an exclusion list for Orange's pre-processing function. The final cleaned data was then checked again in a word cloud. Figure 1 shows a word cloud of the raw data on the left side and a word cloud of the pre-processed data on the right side.

Figure 1

Word cloud before (left) and after (right) pre-processing

Note. Examples of words that were excluded due to the pre-processing were numbers like "12" or single characters like "p" that were not of interest in the frame of this study.

The topic modelling analyses were conducted based on the whole pre-processed corpus. Afterwards, to conduct the two sentiment analyses as well as the LIWC-22 analyses, the pre-processed data was divided into seven corpora, each including only the answers of one therapist. Next, the single messages of each therapist were again divided into single documents in order to consider them individually and to define outliers. Thereby every analysis can be conducted for each set of answers for each therapist separately. In the end, the results provided the basis to compare the single therapists with each other and with the whole group to answer RQ2.

Topic Modelling Analysis

First, a topic modelling analysis was conducted with the entire pre-processed corpus. Topic modelling identifies hidden themes or topics in a corpus. By creating mathematical models, the topic modelling analysis can calculate representations of the words distributed in each text file. These models can then be used to compare topics across multiple documents. The first step of conducting a topic modelling analysis is to define the number of topics inside the text. A topic is defined by the computer program by analysing the frequency of words as well as their appearance in relation to other words in a document. Since the research is exploratory, this step usually starts with an estimation of the number of topics of interest in the text by optimizing the log perplexity and topic coherence parameters. The topic coherence score measures the meaningfulness of the topics generated by the model. Higher coherence scores indicate that the topics are more coherent and meaningful. The log perplexity measures how well the model predicts the observed data. Lower perplexity scores indicate that the model fits the data better. If the coherence and perplexity values of different numbers of topics point to a similar optimal number of topics, the interpretability of the different topics is also considered.

Two topic modelling approaches were chosen, namely Latent Semantic Indexing (LSI) and Latent Dirichlet Allocation (LDA). LSI is a mathematical technique used to identify the underlying semantic relationships between words in a corpus. It does this by creating a low-dimensional representation of the original higher-dimensional space of word frequencies, in which similar words are closer together and dissimilar words are farther apart. LSI is based on the assumption that words that are semantically related will tend to co-occur frequently in documents, and therefore their low-dimensional representation will be closer together (Foltz et al., 1998). LDA is a statistical method which assumes that each document is a mixture of topics and that each topic is a probability distribution over words. LDA iteratively assigns words to topics based on their co-occurrence patterns and estimates the topic probabilities based on the words that appear in each document. Thereby LDA can uncover the most probable topic structure of the corpus (Blei, Ng, & Jordan, 2003).

Based on the selected number of topics, topic modelling analyses provide output lists entailing the most frequent words which together define a topic, which then needs to be interpreted and labelled by the researcher.

Sentiment Analysis

Sentiment analysis is an NLP method which explores the emotional tone as well as the different sentiments expressed within a corpus (Gupta et al., 2021). Therefore, the text is explored for words and phrases that determine underlying emotional valence like positive, negative, or neutral emotions. Most sentiment analysis approaches use semantic lexicons or dictionaries, which include pre-built lists of words and phrases that are associated with specific emotional valences. This lexicon is used to automatically classify the sentiment of a corpus based on the presence and frequency of words that are defined through the semantic lexicon (Gupta et al., 2021). For this research two different sentiment analyses are conducted namely the Valence Aware Dictionary and Sentiment Reasoner (VADER) and the Liu Hu sentiment lexicon.

VADER Analysis

VADER is a lexicon and rule-based sentiment analysis tool which assigns positive, negative, or neutral scores to each word in the corpus, based on its emotional valence (Hutto & Gilbert, 2014). Originally, it was developed to explore the sentiments in social media texts. The results of the VADER analysis display three sentiment scores namely positive, negative and neutral which consider all sentiment scores of all words in a corpus and sum up to 100%. Thereby each sentiment score represents the proportion of text that falls into each sentiment category (Hutto & Gilbert, 2014).

Liu Hu Analysis

Liu Hu's sentiment analysis tool is a lexicon and rule-based method which also originally was designed to analyse the sentiment of social media texts (Abayomi-Alli et al., 2022). However, unlike the VADER analysis, the Liu Hu method provides only a single sentiment score for a corpus which ranges from -100 (extremely negative) to +100 (extremely positive). Like VADER, this score also considers the identification of positive, negative, and neural words, but also takes whole phrases into account and combines the scores into a single number. Therefore, a score of 0 indicates a neutral sentiment, while scores above 0 are considered positive and scores below 0 are considered negative (Abayomi-Alli et al., 2022).

LIWC-22 Analysis

The LIWC-22 software makes use of linguistic rules which identify and categorise words from a corpus based on content, syntax, and function. Thereby text can be classified into various linguistic categories (Pennebaker, Boyd, Jordan, & Blackburn, 2015). Based on this categorisation, the researcher can examine the frequency of specific words and the different types of language used within a text in order to explore the psychological and social processes underlying communication. This quantitative and objective measure of language use can identify patterns and relationships that may not be directly observable from a qualitative analysis of the text (Pennebaker, Boyd, Jordan, & Blackburn, 2015).

The LIWC-22 analysis provides two scales that are of particular interest for identifying language related to empathy namely "Perceptual processes" and "Social processes". The perceptual processes scale measures the extent to which a text reflects the author's attention to sensory and perceptual experiences. By considering aspects like seeing, hearing, and feeling, this scale can provide insights into the extent to which a therapist is attuned to the sensory experiences of the client. Therefore, within a corpus, the perceptual process scale takes "Visual", "Auditory", and "Sensory/Physical" categories into account. Furthermore, the social process scale explores the corpus for the author's focus on social relationships and interactions by considering categories like "Family", "Friends", and "Positive Emotion". This scale can provide insights into the therapist's level of engagement with the client and the quality of the therapeutic relationship (Pennebaker, Boyd, Jordan, & Blackburn, 2015).

Both scales were developed based on a combination of expert judgment and statistical analysis on research about psychological processes within language use and were validated through inter-rater reliability and convergent validity. The scores on these two scales can be interpreted as the extent to which the therapists considered the clients' sensory experiences as well as their social relationships within their answers which is considered an important factor in the application of empathy as well as the building of a therapeutic relationship. The two analyses evaluate a corpus by calculating the percentage of words that fit into the respective category, so the resulting scores range from 0-100%. The higher the score is, the greater the emphasis on the relevant linguistic categories (Pennebaker, Boyd, Jordan, & Blackburn, 2015). For orientation, in a previous study, Valdés (2010) used five different scales of the LIWC-22 analysis to explore a dialogue between a therapist and a client and could categorise 88.53% of around 15,000 words. Nonetheless, the results of the LIWC-22 analysis are not generalisable, and for the interpretation of the numbers the context in which the data was written needs to be considered to make a description of the author's linguistic style (Valdés, 2010).

Language Style Matching Analysis

Besides the linguistic classification, the LIWC-22 software allows for the computation of a language style matching (LSM) score which measures the degree of linguistic similarity between two individuals based on their use of language (Gonzales, Hancock, & Pennebaker, 2010). LSM compares the proportions of words that fall in each of the 22 linguistic categories and calculates a score that reflects the degree of similarity between their language styles. This provides a unique insight into the level of rapport or alignment between two individuals in a range of different conversations from romantic relationships to business negotiations. The LSM score ranges from 0 to 1, with Higher scores indicating a greater linguistic similarity between two individuals and lower scores indicating greater linguistic differences. An important consideration for the evaluation of the LSM analysis is the context in which the text data was created. For example, a high score can either indicate high levels of rapport or agreement between two individuals or a lack of diversity in their communication style. Furthermore, a low score can either be interpreted as two different approaches to communication within a dialogue or as a lack of engagement or attention during the conversation (Gonzales, Hancock, & Pennebaker, 2010).

For the frame of this study, the pairwise comparison LSM option was chosen and applied to the seven individual therapists' corpora to explore the LSM between each dyadic combination of therapists possible.

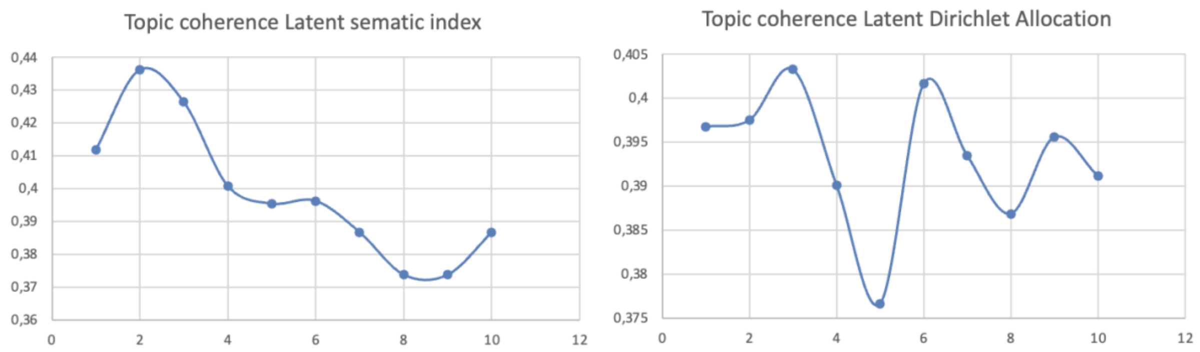
Results

Topic Modelling

The first research question intended to explore the underlying topics of the whole corpus including all answers of all seven therapists. Based on the LSI analysis, the coherence score increased from 0.412 for one topic to 0.436 for two topics but then decreased again for higher numbers of topics (see Figure 2). The LDA model showed an increasing trend in the coherence score, reaching a maximum of 0.403 for three topics (see Figure 2).

Figure 2

Coherence scores in relation to the number of topics for LSI analysis and LDA analysis

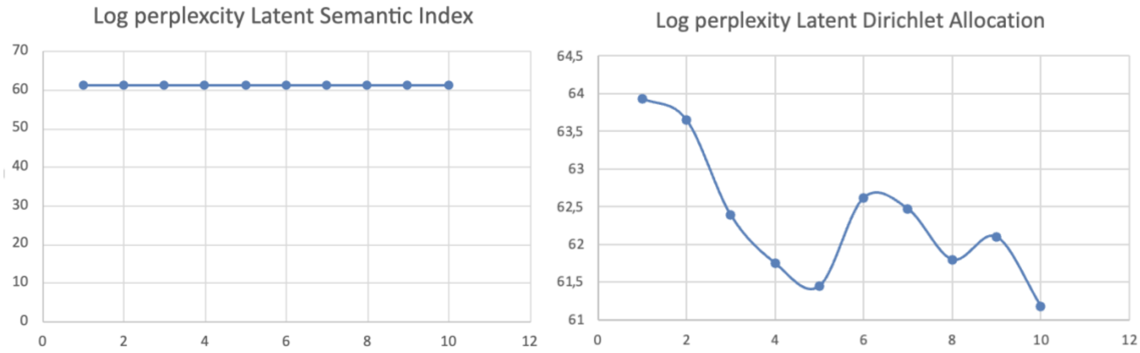


Note. The horizontal axes of both graphs indicate the number of topics, and the vertical axes indicate the topic coherence scores.

The log perplexity score remained constant at 61.2 for all numbers of topics in the LSI analysis, indicating that adding more topics did not improve the model fit (see Figure 3). This output indicated that the analysed data may not have enough variation in the topics mentioned. For the LDA model, the log perplexity decreased from 63.932 for one topic to a minimum of 61.18 for five topics but then remained rather constant for higher numbers of topics (see Figure 3).

Figure 3

Log perplexity in relation to the number of topics for LSI analysis and LDA analysis



Note. The horizontal axes of both graphs indicate the number of topics, and the vertical axes indicate the log perplexity scores.

Across the coherence scores and log perplexity values, the optimal number of topics for both LSI and LDA models appeared to fluctuate around three topics. Table 2 shows the three topics from the LSI as well as the LDA analysis. Both topic modelling analyses resulted in similar topics. The content of the individual topics from both, the LSI and the LDA analysis could not be easily labelled as distinct topics as most frequently included words tended to overlap to a great extent. The results showed that the majority of the words in the different topics were directly related to the sleep discipline focus of the primary study, since they centred around sleeping characteristics and functional instructions to the clients with insomnia, such as the words: "sleep", "bedtime", "awake", "morning" and "week". Nonetheless, some words could be identified in the topics that may have been written with an empathic intention by the therapists, such as the words: "good" and "greetings" (See Table 2). However, investigating the single answers of the therapists, suggested that the word "good" was mostly used in terms like "good morning" and "good luck" and the word "greetings" was most often used at the end of a message as a farewell. It can thereby be concluded that these two frequently used words appeared most often in the context of general politeness, which does not rule out an empathetic intention of the therapists.

Table 2

Results of the Latent Semantic Index (left) and the Latent Dirichlet Allocation analysis (right) based on 3 topics.

Topics	Terms (Probability) of the Latent Semantic Index	Terms (Probability) of the Latent Dirichlet Allocation
Topic 1	sleep (0.017362), week (0.010611), bed (0.010611), time (0.007152), awake (0.006573), minutes (0.006166), bedtimes (0.005895), good (0.00495), asleep (0.004284), efficiency (0.003844)	week (0.010611), sleep (0.017362), awake (0.004284), bedtimes (0.005895), bed (0.008618), minutes (0.006166), efficiency (0.003844), time (0.007152), going (0.002781), last (0.003188)
Topic 2	week (0.010611), times (0.003856), sleep (0.017362), time (0.007152), thoughts (0.001445), minutes (0.006166), morning (0.002171), lie (0.001845), next (0.00296), little (0.002347)	sleep (0.017362), week (0.010611), bed (0.010611), minutes (0.006166), awake (0.004284), time (0.007152), bedtimes (0.005895), good (0.00495), asleep (0.004284), efficiency (0.003844)
Topic 3	sleep (0.017362), week (0.010611), times (0.003856), diaries (0.003305), awake (0.006573), little (0.002347), bed (0.00859),	sleep (0.017362), bed (0.008618), week (0.010611), times (0.007152), bedtimes (0.005895), good (0.00495), times

good (0.00495), regards (0.000573), greetings (0.000856)	(0.003856), awake (0.006573), go (0.004109), asleep (0.004284)

Note. Green-coloured words in the Latent Semantic Index results show positively identified sentiments. Red-coloured words show negatively identified sentiments.

Sentiment Analysis

The second research question investigated the different sentiments expressed within the feedback of the single therapists.

VADER Analysis

The results of the VADER sentiment analysis showed that the different therapists predominantly expressed neutral sentiment with an average of 80.6% of the words being neutral in valence (see Table 3). The average negative sentiment score of 0.033 indicated a relatively low level of negative expression within the corpora and the average positive sentiment score of 0.160 showed that 16% of the words in the seven corpora were categorised as positive sentiments expressed by the different therapists. Notably, there was very little variability between both the average sentiment scores and their variances across the different therapists. For the positive sentiment scores the variance ranged from -1.03% to 2.67% resulting in a total variance of 3.7% from the mean of 0.160. The total variance of the negative sentiment scores suggested even less variability between the single therapists with a score of 2.2%. Finally, also the total variance of the neutral scores showed similar results with a score of 4.5%.

Table 3

Outcome Table Vader Sentiment Analysis

Name	Positive Sentiments	Negative Sentiments	Neutral Sentiments
Therapist 1	0.187 (2.67%)	0.031 (-0.24%)	0.782 (-2.43%)

Therapist 2	0.152 (-0.83%)	0.033 (0.00%)	0.815 (0.87%)
Therapist 3	0.150 (-1.03%)	0.030 (-0.34%)	0.820 (1.37%)
Therapist 4	0.150 (-1.03%)	0.023 (-1.04%)	0.827 (2.07%)
Therapist 5	0.156 (-0.43%)	0.045 (1.16%)	0.799 (-0.73%)
Therapist 6	0.158 (-0.23%)	0.044 (1.06%)	0.798 (-0.83%)
Therapist 7	0.169 (0.87%)	0.028 (-0.54%)	0.803 (-0.33%)
Average score	0.160	0.033	0.806

Note. The percentage variation of the individual scores from the average scores of the respective sentiment category can be found in the brackets.

Liu Hu Analysis

The Liu Hu sentiment analysis showed similar results to the VADER analysis. The sentiment scores of the different therapists again indicated a mostly neutral use of words within the corpora with an average score of 0.467 across all therapists (see Table 4). These results suggested an even more neutral sentiment detection than the VADER analysis. Again, the total variance between the different therapists was very low with a score of 1.546.

Table 4

Outcome Table Liu Hu Sentiment Analysis

Name	Sentiment Scores
Therapist 1	1.032 (0.565)
Therapist 2	0.204 (- 0.263)
Therapist 3	0.048 (-0.419)
Therapist 4	0.107 (-0,360)
Therapist 5	0.054 (-0,413)
Therapist 6	0.233 (-0.234)
Therapist 7	1.594 (1.127)

Average score	0.467

Note. The variation of the individual scores from the average scores of the Liu Hu Sentiment analysis can be found in the brackets.

Based on the results of both the VADER sentiment analysis and the Liu Hu sentiment analysis the findings suggested an overall very similar and neutral tone across the different therapists with slightly more positive than negative sentiment expressions used.

LIWC-22 Analysis

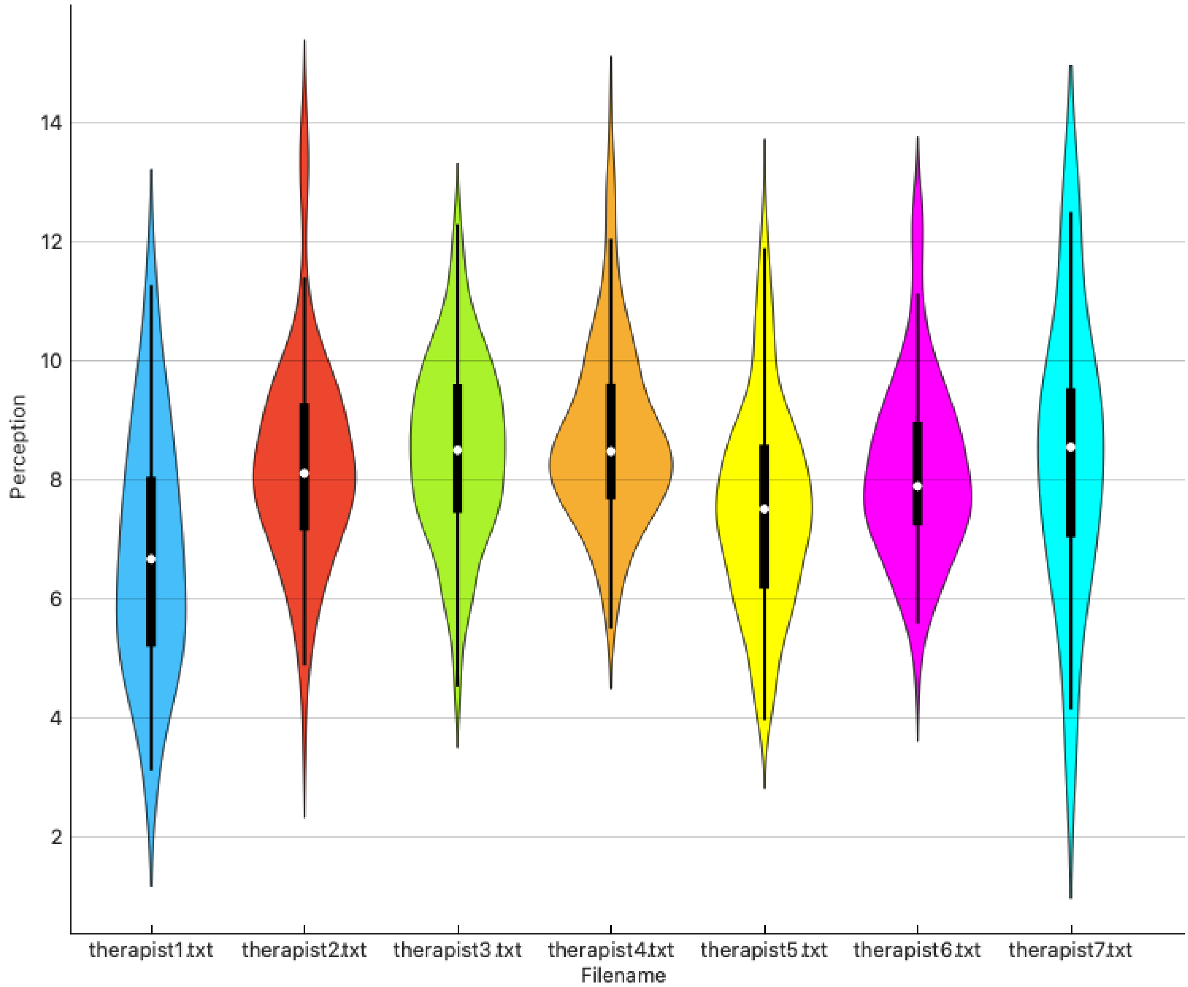
In order to investigate the differences in the use of empathy between the seven therapists and to answer research question three, the LIWC-22 analysis was conducted to explore the “Perceptual processes” and “Social processes” scale.

Perception Processes

The results of the LIWC-22 perceptual process scale showed the extent of attention to sensory and perceptual experiences detected within the seven corpora. The violin plots (see Figure 4) display the distribution of scores for each therapist. The median scores ranging from approximately 6.8% to 8.8%, indicated a modest percentage of words as perceptual language within the seven corpora. The narrow range of therapists' median scores is in line with the similarity between therapists based on the results of the VADER and the Liu Hu sentiment analysis. However, some differences could be observed in the data concerning the variance of the scores considering for instance the highest observed scores of therapist 2 and therapist 4 (14.1%) and the lowest score of therapist 1 (3.1%). Also, differences in the distribution and the density of the scores can be observed indicated by the different shapes of the single violin plots. Nonetheless, the small differences in density and distribution suggested little differences in perceptual processes language used between the therapists.

Figure 4

Violin plot displaying the result of LIWC-22 Perceptual processes scale

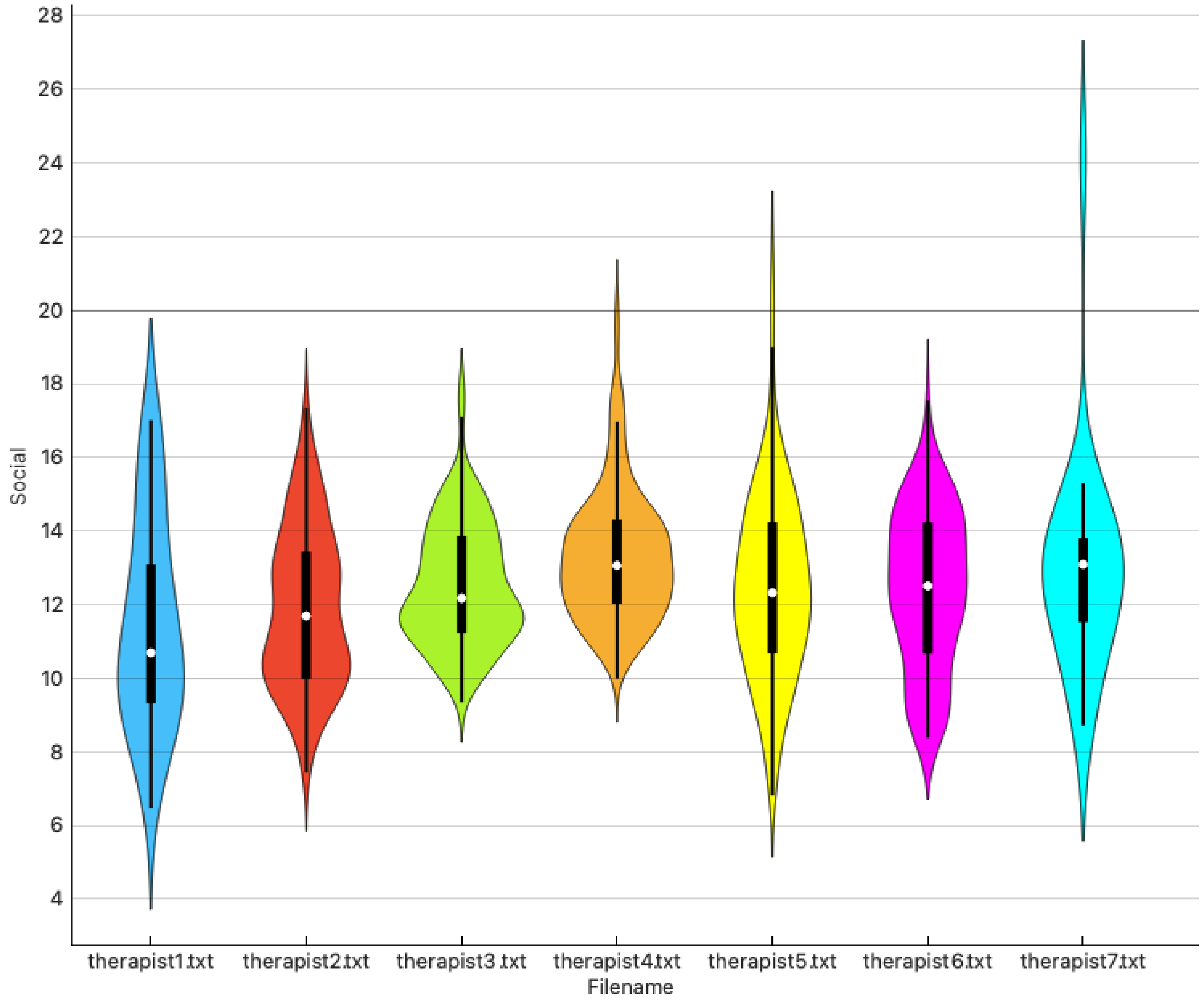


Social Processes

The second analysis based on the LIWC-22 software aimed to examine the extent to which therapists utilized social language in their interactions with clients. As depicted in Figure 5, the results revealed that the therapists exhibited similar average levels of social language application with clients. In fact, all therapists demonstrated a moderate level of social language use in their feedback, with similar median scores ranging from 10.5% to 13%. As previously observed in the perceptual process analysis, the individual violin plots of the seven corpora displayed some intra-therapist differences in terms of the maximum score, distribution, and density.

Figure 5

Violin plot displaying the result of LIWC-22 Social processes scale



Language Style Matching Analysis

To further explore the similarities in the feedback styles of the different therapists the LSM analysis was performed with the seven corpora. The results of the pairwise comparison LSM analysis are presented in a matrix in Table 5. The LSM values ranged from 0.89 to 0.98, indicating very high levels of language style matching between the seven therapists. This indicates that the different therapists showed a great linguistic similarity, which is in line with the very similar results of the different therapists in the VADER, the Liu Hu and the LIWC-22 semantic analysis.

Table 5

Matrix of the pairwise Language Style Matching scores

Filenames	Therapist 1	Therapist 2	Therapist 3	Therapist 4	Therapist 5	Therapist 6
Therapist 1	-					
Therapist 2	0.91	-				
Therapist 3	0.91	0.95	-			
Therapist 4	0.89	0.94	0.96	-		
Therapist 5	0.91	0.98	0.93	0.93	-	
Therapist 6	0.90	0.97	0.94	0.94	0.96	-
Therapist 7	0.89	0.96	0.96	0.96	0.95	0.95

Discussion

This study used different text-mining methods to explore underlying topics of CBTi therapists' online feedback as well as to detect differences in the expressed sentiments and use of empathic language between seven therapists. First, the topic modelling analyses both identified three latent topics within the written feedback of the seven therapists. The discovered topics of the LSI as well as the LDA analysis overlapped to a great extent and showed mainly functional CBTi-related word choices by therapists. Furthermore, the sentiment analyses identified mainly neutral sentiments and slightly more positive than negative sentiments within the seven corpora. Regarding the use of empathy, the results of both, the social processing as well as the emotional perception scale indicated a modest use of empathy within and slight differences between the seven therapists. Finally, the results of the LSM analysis indicated very high similarity between the seven therapists' language styles.

Overall, it can be concluded, that the text-mining methods used in this study were apparently useful for meaningfully analysing the feedback data of the seven therapists and for investigating the seven therapists' use of empathy.

Topic Modelling Analyses

The results of the LSI and LDA topic modelling analysis identified three underlying topics within the whole data set. Since in both analyses, the words which define each topic overlapped to a great extent, and the log perplexity score of the LSI analysis showed no dependency on the number of topics, the results of the topic modelling analyses suggested little variation in topics underlying the therapists' feedback. The identified topics in both the LSI and LDA models mostly included rather functional word choices related to sleep, bedtime, and morning routines, which can be assumed to be directly related to the focus of the CBTi study of De Bruin et al. (2015). However, the identification of some words like "good", and "greetings" can be interpreted as an empathic intention on behalf of the therapists. Nonetheless, considering the context in which the two words were used, their application is more likely indicative of general politeness, which, however, does not exclude an empathic intention.

The finding that the word "good" appears in the outcome table of the LSI analysis as both a positive and negative sentiment is interesting and warrants further exploration. The paper of Liu and Zhan (2012) explains such cooccurrence of the same word in different topics by underlining that the LSI analysis considers the context in which a word appears within a corpus of interest. Therefore, the same word can occur in a positive and negative context within the same corpus. Compared to other topic modelling analyses, this context consideration can be seen as a strength of the LSI analysis due to its wide range of applicability to different text types. Other text mining techniques like the LDA analysis can associate a word only with either a positive or negative sentiment (Liu & Zhan, 2012). Therefore, it can be concluded, that the word "good" occurred in the data of the study of De

Bruin et al. (2015) in both a positive and a negative context. Phrases like "good night" and "good job" appear to create a positive context whereas phrases like "not feeling good" or "not too good" are considered as negative sentiments by the LSI analysis within the seven corpora.

Overall, the topic modelling analyses provided some initial insights into the underlying topics of the data from the study of De Bruin et al. (2015). Although three topics seemed to underly both LSI and LDA models, the identified overlapping topics were very similar and mostly of functional nature. The little variation within the results can be attributed to the original CBTi study of De Bruin et al. (2015). The data for the study were collected as part of an RCT study, which assumes a certain standardization for the therapists' messages. Therefore, the results of the topic modelling analyses are in line with the data collection conditions. Nonetheless, future research should investigate the application of topic modelling analyses for non-controlled and non-standardized therapeutic text conversations to better explore the NLP method's sensitivity.

Sentiment Analyses

The outcome table of the VADER analysis indicated a predominant use of neutral language by all therapists as well as a moderate use of positive sentiment and a low use of negative language. The Liu Hu sentiment analysis also detected mostly neutral sentiments in the different corpora and only minor differences between the therapists' expressions of sentiment. The very low variance between the individual scores of the seven therapists in the outcome tables of both sentiment analyses indicated a very similar writing style within the seven corpora which is in line with the findings from the topic modelling analyses.

Concerning the sentiment analyses, it needs to be mentioned that the VADER as well as the Liu Hu algorithm were designed to explore social media texts and therefore might not be sensitive enough to detect more nuanced sentiment in a therapeutic context (Hutto & Gilbert, 2014; Hu, Minqing & Bing Liu, 2004). This might be the case since the algorithms, as well as their lexical features, are developed and validated based on micro-blog context

(Hutto & Gilbert, 2014). Compared to the micro-blog social media context, the feedback messages of the therapists were longer and written in a more intimate setting than a social media post. Nonetheless, the results of the two sentiment analyses are in line with the other NLP methods' results as they indicate a very low variance within seven therapists' language use.

In conclusion, it therefore can be said, that the VADER and the Liu Hu sentiment analyses proved useful for the exploration of different sentiments within therapeutic feedback text data.

LIWC-22 Analyses

The results of the LIWC-22 analysis were also in line with the topic modelling and sentiment analysis findings. The therapists' individual median scores on the perceptual process scale, as well as the median scores on the social process scale, again were in a very narrow range. Even though some outliers indicated that the therapists sometimes used more or less empathic language, the variance of the mean scores between the single therapists was very low.

The findings of the present study shed light on the expressions of empathy within seven therapists' online asynchronous feedback. While the results indicated a very similar level of empathy between the therapists, there was fluctuation observed within their individual responses. This fluctuation highlights the importance of further exploring the optimal degree of expressed empathy within the therapeutic context to individually improve the formation of a good working alliance and to enhance the treatment outcome for clients in general (Falkenström et al., 2014). The need to further explore the concept of empathy in an online therapeutic context is underlined by this study. The results of the LIWC-22 analyses showed, that even when the median scores on both scales are very narrow, there are still remarkable differences in the detection of empathy between the single messages when considering the individual maximum scores. Concluding, even in a controlled study setting it is not yet possible for therapeutic experts to constantly show an adequate level of empathy. Therefore, further investigation of the optimal level of empathy expression within both human and computer-based eMental health interventions holds promise for enhancing therapy effectiveness and adherence (Hadjistavropoulos et al., 2017).

In conclusion, the social processing and emotional perception scales of the LIWC-22 analysis successfully and reliably detected differences in the expression of empathy within the seven therapists' feedback messages. Although the differences in the median scores of the individual therapists were very small, noteworthy differences between the individual messages could be identified. Therefore, it is recommended for future research to continue using the LIWC-22 analysis as due to its sensitivity to various linguistic categories, this NLP method offers great potential for further investigating the optimal expression of empathy in a therapeutic text-based online context.

Language Style Matching Analysis

The outcome table of the language style matching analysis showed very high language style matching scores among the seven therapists, confirming very similar language styles. In combination with the low variation detected in the results of the topic modelling and the sentiment analyses as well as on the two LIWC-22 outcome scales, these findings suggest a highly protocolised approach for providing feedback within the study of De Bruin et al. (2015). This assumption is in line with the fact that the participating therapists of the RCT study took part in weekly supervision and intervision sessions with an additional independent expert for sleep therapy (De Bruin et al., 2015). Therefore, it is assumed, that the analysed feedback texts were composed under highly protocolized guidelines, which would correspond with the relatively strict procedures within the settings of an RCT study. Based on this standardized context, research about the application of empathy in a text-based therapeutic context might not be representative for the content in real-world therapy, since the extent to which the therapists had leeway to independently express empathy was unknown.

Concluding, due to possible further differences in therapeutic feedback within non-RCT real-world settings, the generalisability of the results of this study should be treated with caution.

Moreover, the degree to which the language between the seven single therapists matched with each other suggests that online feedback tasks could even be taken over by unguided web interventions based on a chatbot or artificial intelligence (Al-Alawi et al.,2021). In this context, other studies already proved that computer-generated feedback messages within web-based interventions lead to similar effectiveness and adherence as human-created feedback messages which were composed for the same web-based intervention for the treatment of depression symptoms (Kelders et al., 2015). However, the extent to which artificial intelligence can mimic nonspecific treatment mechanisms on a human level requires further research.

Overall, the results of the LSM analysis gave valuable insights into the extent to which the single therapists' feedback messages included similar language styles. The high level of linguistic agreement between the seven therapists can be attributed to the strict procedures within the settings of an RCT study. Therefore, future research should consider therapeutic feedback messages of non-RCT real-world settings. Finally, future research could also explore the optimal amount of empathy within online feedback, to optimise the applicability of computer-generated feedback for eMental health interventions and to investigate the extent to which an AI can express empathy.

Limitations

As the analysed data in the current study was secondary data from a Dutch RCT, several limitations need to be mentioned. First of all, the data was written in Dutch but analysed in English. Therefore, the translation of the data with google translate could have led to a loss of semantic meaning in the word choice of the therapists as well as the content of the messages as a whole. Furthermore, the data was created within a rather strictly protocolized setting which limits the generalisation of the results for real-world therapeutic settings.

Among other characteristics, RCT studies include strict treatment protocols, strict inclusion and exclusion criteria for the participant selection and specifically trained therapists (De Bruin et al., 2015). Therefore, in a real-world setting, the therapists' choice of treatment approach is expected to be more flexible since strategies, content and therapeutic techniques could be adapted individually to a client's needs. Also, non-RCT settings would include patients with a wider range of symptoms as well as comorbidities and therapists with varying levels of expertise, experience, and training backgrounds. Concluding, by considering more therapists with different expertise levels and a wider range of illness patterns under a less controlled setting, more insights about the use of empathy in a therapeutic online context could be made. Finally, the text mining methods VADER and Liu Hu were originally developed for the exploration of sentiments in social media texts which could have potentially limited their validity and sensitivity in investigating therapeutic texts.

Future Research

Based on the current findings and considering the limitations of this study, in the future, more real-world non-RCT study data should be taken into account, in general, to further develop text mining methods and specifically for more precise detection of empathy expression in a therapeutic context. Preferably, this data should then be derived from a clinical setting for example from several eMental health portals and analysed in the mother tongue in which it was composed. Furthermore, to identify specific words that can express empathy the best, the outcome of individual therapies as well as client-related observer ratings on empathy measurement scales like the BLRI should be considered in further research (Greenberg et al., 2001). Thereby, an NLP artificial intelligence could be created that is trained to recognize empathy more accurately than humans, as Mawani and Nderu (2020) have described in their article. A potential technology could then be developed, which can indicate to therapists the extent of expressed empathy in a feedback text before sending it, and can even present suggestions for modification to improve the text. A computer-based control instance could thus have a positive influence on the therapist-client relationship and at the same time even support the dynamics of the conversation so that no conflicting information is passed on to the client or follow-up information is always given in the correct order, even if the therapist has changed in the course of therapy.

Moreover, future research could even explore the optimal amount of empathy within online feedback, to investigate the applicability of computer-generated feedback for eMental health interventions and to optimise the extent to which an AI can express empathy.

Finally, since the specific NLP analyses possibilities of the Orange 3.0. text mining software are limited and not specifically targeted for psychological research; other text mining methods could be used for future research. For instance, one promising approach for more accurate and meaningful analysis is offered by Googles Bard artificial intelligence which already shows promising approaches in NLP (Google AI, 2023).

Conclusion

Empathy is often considered a critical success factor of psychotherapy. Nonetheless, little is known about the correct conveyance of empathy in the online therapeutic context. The empathy investigation in the frame of this study has shown, that even under very controlled circumstances, professional experts are not able to constantly convey an adequate level of empathy within their asynchronous online feedback. It is therefore important to continue researching the optimal expression and role of empathy in eMental health applications. The current study suggests that automated NLP methods offer the potential to further investigate the concept of empathy in this setting. This study pointed out that basic text mining methods like the LSI and the LDA sentiment analysis, which were not originally designed to explore therapeutic feedback texts, showed results in line with the results of specifically for this purpose developed analyses like the LIWC-22. Thereby it can be concluded that NLP methods offer great potential for the understanding of underlying concepts within a therapeutic context and by that to improve therapist-client relationships within asynchronous text-based eMental health interventions in general.

Reference List

Abayomi-Alli, A., Abayomi-Alli, O., Misra, S., & Fernandez-Sanz, L. (2022). Study of the Yahoo-Yahoo Hash-Tag Tweets Using Sentiment Analysis and Opinion Mining Algorithms. *Information*, *13*(3), 152. https://doi.org/10.3390/info13030152

AbuKhousa, E., Mohamed, N., & Al-Jaroodi, J. (2012). e-Health Cloud: Opportunities and Challenges. *Future Internet*, *4*(3), 621–645. https://doi.org/10.3390/fi4030621

Al-Alawi, M., McCall, R. K., Sultan, A., Al-Balushi, N., Al-Mahrouqi, T., Ghailani, A. A., Sabti, H. A., Al-Maniri, A., Panchatcharam, S. M., & Sinawi, H. A. (2021). Efficacy of a Six-Week-Long Therapist-Guided Online Therapy Versus Self-help Internet-Based Therapy for COVID-19–Induced Anxiety and Depression: Open-label, Pragmatic, Randomized Controlled Trial. *JMIR Mental Health*, *8*(2), e26683. https://doi.org/10.2196/26683

Andersson, G., & Titov, N. (2014). Advantages and limitations of Internet-based interventions for common mental disorders. *World Psychiatry*, *13*(1), 4–11. https://doi.org/10.1002/wps.20083

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin*. https://www.liwc.app

Butler, R., O'Day, E. B., Swee, M. B., & Horenstein, A. (2021). Cognitive Behavioral Therapy for Social Anxiety Disorder: Predictors of Treatment Outcome in a Quasi-Naturalistic Setting. *Behavior Therapy*, *52*(2), 465–477. https://doi.org/10.1016/j.beth.2020.06.002

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B. O., Emmelkamp, P. M. G., Asmundson, G. J., Carlbring, P., & Powers, M. B. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders*, *61*, 27–36. https://doi.org/10.1016/j.janxdis.2018.08.003

Chowdhary, K. R. (2020). Natural Language Processing. Springer eBooks, 603–649. https://doi.org/10.1007/978-81-322-3972-7_19

Danubianu, M. (2015). Step by step data preprocessing for data mining. A case study. In Proc. *Of the International Conference on Information Technologies (InfoTech-2015)* (pp. 117-124).

De Bruin, E. J., Bögels, S. M., Oort, F. J., & Meijer, A. M. (2015). Efficacy of Cognitive Behavioral Therapy for Insomnia in Adolescents: A Randomized Controlled Trial with Internet Therapy, Group Therapy and A Waiting List Condition. Sleep, 38(12), 1913–1926. https://doi.org/10.5665/sleep.5240

De Bruin, E. J., & Meijer, A. M. (2017). The impact of online therapeutic feedback on outcome measures in Internet-CBTI for adolescents with insomnia. Sleep Medicine, 29, 68–75. https://doi.org/10.1016/j.sleep.2016.05.017

Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research* 14(Aug): 2349−2353.

Duettmann, W., Naik, M., Zukunft, B., Osmonodja, B., Bachmann, F., Choi, M., Roller, R., Mayrdorfer, M., Halleck, F., Schmidt, D., & Budde, K. (2021). eHealth in transplantation. *Transplant International*, *34*(1), 16–26. https://doi.org/10.1111/tri.13778

Elliott, R. J., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, *48*(1), 43–49. https://doi.org/10.1037/a0022187

Ellis, L. A., Meulenbroeks, I., Churruca, K., Pomare, C., Hatem, S., Harrison, R., Zurynski, Y., & Braithwaite, J. (2021). The Application of e-Mental Health in Response to COVID-19: Scoping Review and Bibliometric Analysis. *JMIR Mental Health, 8*(12), e32948. https://doi.org/10.2196/32948

Falkenström, F., Granström, F., & Holmqvist, R. (2014). Working alliance predicts psychotherapy outcome even while controlling for prior symptom improvement. *Psychotherapy Research*, *24*(2), 146–159. https://doi.org/10.1080/10503307.2013.847985

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

Flinsenberg, D. (2020). *Autonomy through technology: engaging executive functions in ADHD with an extended mind approach* (Master's thesis, University of Twente). http://essay.utwente.nl/81873/

Flückiger, C., Del Re, A., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, *55*(4), 316–340. https://doi.org/10.1037/pst0000172

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, *25*(2–3), 285–307. https://doi.org/10.1080/01638539809545029

Goldberg, S. B., Flemotomos, N., Martinez, V. D., Tanana, M. J., Kuo, P. X., Pace, B. P., Villatte, J. L., Georgiou, P. G., Van Epps, J., Imel, Z. E., Narayanan, S. S., & Atkins, D. C. (2020). Machine learning and natural language processing in psychotherapy research: Alliance as example use case. *Journal of Counseling Psychology, 67*(4), 438–448. https://doi.org/10.1037/cou0000382

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language Style Matching as a Predictor of Social Dynamics in Small Groups. *Communication Research*, *37*(1), 3–19. https://doi.org/10.1177/0093650209351468

Google AI. (2023). Google Bard. Retrieved from https://bard.google.com/.

Google Translate. (2023, May 23). Retrieved from https://translate.google.com/.

Greenberg, L. S., Elliott, R. J., Watson, J. C., & Bohart, A. C. (2001). Empathy. *Psychotherapy, 38*(4), 380–384. https://doi.org/10.1037/0033-3204.38.4.380

Gupta, R., Sameer, S., Muppavarapu, H., Enduri, M. K., & Anamalamudi, S. (2021). *Sentiment Analysis on Zomato Reviews*. https://doi.org/10.1109/cicn51697.2021.9574641

Hadjistavropoulos, H. D., Pugh, N. E., Hesser, H., & Andersson, G. (2017). Therapeutic Alliance in Internet-Delivered Cognitive Behaviour Therapy for Depression or Generalized Anxiety. Clinical Psychology & Psychotherapy, 24(2), 451–461. https://doi.org/10.1002/cpp.2014

Horvath, A. O., Del Re, A., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. Psychotherapy, 48(1), 9–16. https://doi.org/10.1037/a0022186

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. *Journal for Language Technology and Computational Linguistics*, 20(1), 19-62.

Hu, Minqing and Bing Liu (2004). Mining opinion features in customer reviews. In: AAAI, vol. 4, pp. 755–760.

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Kelders, S. M., Bohlmeijer, E. T., Pots, W. T. M., & Van Gemert-Pijnen, J. E. (2015). Comparing human and automated support for depression: Fractional factorial randomized controlled trial. *Behaviour Research and Therapy*, *72*, 72–80. https://doi.org/10.1016/j.brat.2015.06.014

Kooistra, L., Wiersma, J., Ruwaard, J., Neijenhuijs, K. I., Lokkerbol, J., Van Oppen, P., Smit, F., & Riper, H. (2019). Cost and Effectiveness of Blended Versus Standard Cognitive Behavioral Therapy for Outpatients With Depression in Routine Specialized Mental Health Care: Pilot Randomized Controlled Trial. *Journal of Medical Internet Research*, *21*(10), e14261. https://doi.org/10.2196/14261

Leichsenring, F., Abbass, A., Luyten, P., Hilsenroth, M. J., & Rabung, S. (2013). The Emerging Evidence for Long-Term Psychodynamic Therapy. *Psychodynamic Psychiatry*, *41*(3), 361–384. https://doi.org/10.1521/pdps.2013.41.3.361

Litvak, M., Otterbacher, J., Ang, C. S., & Atkins, D. (2016). Social and linguistic behavior and its correlation to trait empathy. *In Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES*) (pp. 128–137). The COLING 2016 Organizing Committee. Retrieved May 30, 2023, from https://www.acl web.org/anthology/W16-4314

Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In *Springer eBooks* (pp. 415–463). https://doi.org/10.1007/978-1-4614-3223-4_13

Luteijn, F., & Barelds, D. P. H. (2019). *Psychological Diagnostics in Health Care*.

Mawani, A., & Nderu, L. (2020). Towards an Online Empathy Assisted Counselling Web Application. *EAI Endorsed Transactions on Context-aware Systems and Applications*, *7*(22), 167792. https://doi.org/10.4108/eai.23-12-2020.167792

Moyers, T. B., & Miller, W. R. (2013). Is low therapist empathy toxic? *Psychology of Addictive Behaviors*, *27*(3), 878–884. https://doi.org/10.1037/a0030274

Muthiah, A., Ajitha, S., S, M. K., K, V. V., Kavitha, K., & Mallick, P. S. (2019). *Maternal ehealth Monitoring System using LoRa Technology*. https://doi.org/10.1109/icawst.2019.8923228

Nienhuis, J., Owen, J., Valentine, J. C., Black, S. R., Halford, T. C., Parazak, S. E., Budge, S. L., & Hilsenroth, M. J. (2018). Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: A meta-analytic review. *Psychotherapy Research*, *28*(4), 593–605. https://doi.org/10.1080/10503307.2016.1204023

Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, *35*(1), 136–147. https://doi.org/10.1016/j.tele.2017.10.006

Plaß, B., Prudhomme, C., & Ponciano, J. (2021). BIM ON ARTIFICIAL INTELLIGENCE FOR DECISION SUPPORT IN E-HEALTH. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B2-2021*, 207–214. https://doi.org/10.5194/isprs-archives-xliii-b2-2021-207-2021

Rogers, C. R. (1975), "Empathy: An Unappreciated Way of Being," The Counseling Psychologist, Vol. 5 No. 2, pp. 2-10.

Rogers, K., & Biesanz, J. C. (2019). Reassessing the good judge of personality. *Journal of Personality and Social Psychology*, *117*(1), 186–200. https://doi.org/10.1037/pspp0000197

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale Development and Initial Validation of a Factor-Analytic Solution to Multiple Empathy Measures. *Journal of Personality Assessment*, *91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Tong, Z., & Zhang, H. (2016). *A Text Mining Research Based on LDA Topic Modelling*. https://doi.org/10.5121/csit.2016.60616

Valdés, N. (2011). Analysis of Patient and Therapist Linguistic Styles During Therapeutic Conversation in Change Episodes, Using Linguistic Inquiry and Word Count (LIWC) 13. Faculty of Behavioural and Cultural Studies, Heidelberg University in Cooperation with the Pontificia Universidad Católica de Chile. *SciELO Journal of Subjetividad y Procesos Cognitivos*, 155.

Van Gemert-Pijnen, L., Kelders, S. M., Sanderman, R., & Kip, H. (2018). *EHealth Research, Theory and Development: A Multi-Disciplinary Approach*.

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Veiga, G. M. da S. (2019). *The application of sentiment analysis to a psychotherapy session: an exploratory study using four general-purpose lexicons*. Retrieved from http://hdl.handle.net/10400.12/7372.

Watson, J. C. (2016). The role of empathy in psychotherapy: Theory, research, and practice. In *American Psychological Association eBooks* (pp. 115–145). https://doi.org/10.1037/14775-005

Watts, S. W., Marchand, A., Bouchard, S., Gosselin, P., Langlois, F., Belleville, G., & Dugas, M. J. (2020). Telepsychotherapy for generalized anxiety disorder: Impact on the working alliance. Journal of Psychotherapy Integration, 30(2), 208–225. https://doi.org/10.1037/int0000223

Wesslen, R. (2018). Computer-assisted text analysis for social science: Topic models and beyond. *arXiv preprint arXiv*:1803.11045.