

How digital therapeutic alliances influence the perceived helpfulness of online mental health Q&A: An explainable machine learning approach

Yinghui Huang^{1,2}, Hui Liu^{3,4}, Maomao Chi⁵, Sujie Meng^{3,4} and Weijun Wang^{3,4}

Abstract

Objective: This study investigates the role of digital therapeutic alliance (DTA) in predicting and explaining the perceived helpfulness of responses on online mental health Q&A platforms.

Methods: This study constructs a large dataset of 19,682 Q&A interactions from online mental health Q&A platforms, employs natural language processing, explainable machine learning, and causal inference methods to identify and understand the factors, particularly DTA, that influence the perceived helpfulness of human counselors' responses to mental health questions.

Results: The machine learning-based model for predicting perceived helpfulness demonstrated strong performance, achieving an root mean square error of 0.8234 and a mean absolute percentage error of 22.7288%. The explanatory analysis revealed that peripheral path-related language cues, such as counselor engagement (e.g., word count and response time), had the highest predictive power. Additionally, central path-related language cues, such as those linked to the DTA—specifically emotional bonds and therapeutic tasks—significantly influenced perceived helpfulness and were positively impacted by counselor engagement.

Conclusion: This study integrates DTA and elaboration likelihood model theories to propose a computational framework for understanding and predicting the perceived helpfulness of responses in online mental health Q&A platforms. Findings offer theoretical insights into the mechanisms of perceived helpfulness and practical guidance for optimizing platform design, training counselors, and improving user satisfaction through targeted language strategies.

Keywords

Digital therapeutic alliance, elaboration likelihood model, explainable machine learning, mental health Q&A, perceived helpfulness

Received: 11 August 2024; accepted: 20 March 2025

Introduction

Mental health risks are among the top ten global risks. In China, according to the “2023 National Mental Health Development Report,” the detection rate for depression risk is 10.6%, and for anxiety risk, it is 15.8%. Among adults, those aged 18–24 are at particularly high risk of depression, with a detection rate of 24.1%. However, due to economic burdens, time constraints, and physical distances, only 12% of individuals with mental illnesses receive treatment.¹ Traditional psychological counseling methods, constrained by limited resources, struggle to meet widespread demand, particularly among vulnerable

¹Research Institute of Digital Governance and Management Decision

Innovation, Wuhan University of Technology, Wuhan, Hubei Province, China

²School of Management, Wuhan University of Technology, Wuhan, Hubei Province, China

³Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China

⁴Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, Wuhan, China

⁵School of Economics and Management, China University of Geosciences, Wuhan, China

Corresponding authors:

Sujie Meng, School of Psychology, Central China Normal University, Wuhan 430079, China.

Email: psymsj1003@163.com

Weijun Wang, School of Psychology, Central China Normal University, Wuhan 430079, China.

Email: wangwj@mail.ccnu.edu.cn



groups, where access to psychological therapy and social support is often impeded.²

In recent years, the emergence of online mental health communities such as TalkLife (TalkLife.com) and Yixinli (xinli100.com) has created new possibilities for addressing mental health issues. For example, Yixinli is one of China's largest online mental health platforms, with nearly 40 million users. Through its Q&A section, users can share their psychological troubles and receive guidance and support from counselors.^{3,4} The text-based online counseling model, a form of collaboration between users and counselors, has been proven to be as effective as traditional face-to-face counseling in addressing depression, anxiety, and emotional issues.⁵ This model provides an alternative for individuals who feel unsafe discussing their distress and privacy in traditional counseling settings or over the phone. Online asynchronous text counseling through messaging platforms allows for non-real-time interactions, offering users flexibility in timing and lower psychological barriers while also breaking geographical constraints and enhancing the accessibility of mental health support.^{6,7} This counseling format enables interactions at convenient times, reducing scheduling conflicts and providing a recordable dialogue history that facilitates reflection and tracking progress.⁷ Exploring the effectiveness of text-based asynchronous mental health Q&A services is crucial for providing more timely and effective mental health support.⁷

Perceived helpfulness refers to the extent to which users believe a system supports them in achieving their goals,^{8,9} and it is particularly important in the context of online mental health platforms. Perceived helpfulness is commonly determined by the quality of information and responses provided by the platform.¹⁰ However, in the field of mental health, perceived helpfulness is influenced not only by explicit informational content but also by the complex mechanisms through which language cues shape users' emotional and cognitive responses. This complexity underscores the limitations of relying on a single theoretical model to fully understand the dynamics of perceived helpfulness.

According to the elaboration likelihood model (ELM), users process information through either a central or peripheral route depending on their level of involvement and the complexity of the information.¹¹ Through the central route, users engage in deep analysis of the information's content, while the peripheral route involves reliance on simpler cues, such as source credibility or text length.¹² Therefore, ELM provides a robust theoretical framework for explaining how information influences users' perceived helpfulness on online mental health platforms.

Perceived helpfulness in mental health platforms is also significantly shaped by the digital therapeutic alliance (DTA). DTA extends the traditional therapeutic alliance concept by emphasizing the emotional bonds, task alignment, and goal collaboration established between users and digital tools or interventions.^{13,14} Incorporating DTA

into the central-route framework of ELM allows for a more comprehensive understanding of how language cues influence users' information processing and perceptions of helpfulness, particularly in mental health settings. Therefore, this study integrates the perspectives of ELM and DTA to investigate how language cues, through both central and peripheral routes, impact perceived helpfulness.

Additionally, natural language processing (NLP) methods can learn data structures from written and spoken language, providing valuable feedback in psychotherapy.¹⁵ For example, language inquiry and word count (LIWC) can be used to automatically detect therapeutic alliance, which are significant predictors of psychotherapy outcomes.¹⁶ The application of NLP in large-scale open-text mental health Q&As offers new pathways for optimizing online mental health Q&A platforms.

Therefore, this study aims to explore the impact mechanisms of DTA on the perceived helpfulness of online mental health Q&As from the perspective of the ELM. By integrating large-scale mental health Q&A texts with explainable machine learning methods, this study examines the perceived helpfulness of mental health Q&A information, offering theoretical and practical guidance for optimizing online mental health platforms.

Text-Based counseling in mental health communities

Although psychotherapy and social support are effective, many disadvantaged groups encounter barriers when seeking help. For instance, the shortage of counseling resources limits their access to therapy and counseling services.¹⁷ A possible solution is to transition from traditional one-on-one to a one-to-many counseling, maximizing counseling resources without sacrificing effectiveness. In China, the emergence of online mental health communities such as Yixinli (xinli001.com) and Jiandanxinli (jiandanxinli.com) exemplifies effective practice. For instance, Yixinli is one of the largest online mental health platforms in China, serving nearly 40 million users seeking psychological assistance.^{18,19} This platform adopts an open Q&A model where users anonymously post their psychological distress and challenges to receive guidance and support from platform counselors. These counselors include licensed mental health professionals and trained volunteers who can freely choose which questions to answer. Unlike traditional one-on-one matching models, counseling provided by Yixinli is more similar to a public forum, allowing users to receive diverse responses and advice from multiple counselors. This open interaction model allows the platform to efficiently address high consultation demand while maintaining flexibility and accessibility. Although the platform does not impose fixed treatment cycles or stages, users can address their psychological concerns gradually through ongoing questions and consultations. Additionally, the platform emphasizes user anonymity, reducing psychological

barriers and promoting broader participation. Counselors respond using their professional identities, and users can choose whether or not to view detailed information about the counselors to ensure consultation quality and credibility.

Online text-based counseling represents collaboration between users and counselors to identify issues and develop solutions. Studies have shown that text-based counseling is as effective as traditional face-to-face counseling in addressing a range of issues, including depression, anxiety, and emotional problems.²⁰ This type of counseling offers an alternative for those who feel unsafe discussing their pain and privacy with strangers in traditional counseling environments or over the phone.^{21,22} However, compared to in-person counseling, text-based counseling still faces challenges such as communication delays, lack of non-verbal cues, and issues related to the environment and connectivity.^{7,22}

Notably, users may prefer real-time support when seeking help, even from trained volunteers or artificial intelligence agents.⁶ However, interactions within online mental health communities are often asynchronous. Therefore, this study will use the Yixinli platform as a case study to explore the effectiveness of DTA in text-based asynchronous mental health Q&A services.

Effectiveness of mental health Q&As

In recent years, there has been a growing use of online mental health Q&A platforms for expressing concerns, seeking social support, and obtaining therapy.¹⁰ Consequently, researchers have increasingly focused on the effectiveness of mental health Q&A communities. One study highlighted online support's value for French university students' well-being, suggesting that institutional promotion could enable digital interventions to replace traditional services.²³ Another quantitative study explored why and how young people use the static, asynchronous online mental health Q&A platform hosted by KOOTH.com, finding that such platforms provide a supportive and positive environment for young people with mental health and emotional needs, fostering greater connectivity with others.²⁴ Additionally, a systematic review indicated that online mental health Q&A communities offer practical resources for young people to obtain health information and emotional support.²⁵

Perceived helpfulness is the extent to which an individual believes that using a specific system will help them achieve their goals and improve job performance.^{8,9} For online mental health platforms, it is often described as the degree to which users believe the platform assists them in achieving mental health-related goals. This is evidenced by help-seekers continuing to provide interactive feedback after receiving responses. Researchers examined the perceived helpfulness of the mental health Q&A platform and its impact on users' mental health outcomes, finding that the quality of information and responses in help-seeking posts significantly influenced perceived helpfulness.¹⁰

Elaboration likelihood model and digital therapeutic alliance

The perceived helpfulness of online mental health platforms is not solely influenced by the quality of the information provided but also by how users process and interpret that information. The ELM, proposed by Petty and Cacioppo in 1986, is a widely accepted framework that explains how individuals process information and form attitudes. According to the ELM, individuals process information through two distinct routes: the central route and the peripheral route. The central route involves careful and thoughtful consideration of the content of the message, while the peripheral route relies on secondary cues such as the credibility of the source or the emotional factors of the message.¹¹ In the context of online mental health Q&A platforms, the central route might involve users deeply engaging with the quality and relevance of the information provided by counselors, while the peripheral route might involve evaluating the credibility of the platform or the emotional tone of the response.

The DTA extends the concept of the traditional therapeutic alliance to digital platforms, focusing on the relationship between users and digital therapeutic interventions. It refers to the relationship between clients and digital tools or interventions designed to provide mental health support.²⁶ The DTA encompasses various dimensions, including emotional connection, therapeutic tasks, and goal consistency.^{13,14} Research has shown that a strong DTA can significantly enhance therapy outcomes, similar to traditional face-to-face therapeutic alliance.^{13,27} In the context of online mental health Q&A platforms, the DTA can be operationalized through language cues that reflect emotional bond (e.g., empathy and warmth), task alignment (e.g., clarity of advice and actionable steps), and goal consistency (e.g., alignment between user needs and counselor responses). Psychotherapy's effectiveness largely depends on the therapeutic alliance.²⁸ A series of studies have shown that digital psychological therapeutic alliance are comparable to those established in face-to-face psychological therapy.^{13,14} A meta-analysis also found that DTA can improve therapy outcomes, with a moderate overall effect size.²⁷ Moreover, DTA are positively correlated with engagement,²⁹ and engagement can enhance the therapeutic effects of DTA.³⁰

This study integrates the ELM model and DTA to provide a comprehensive understanding of the perceived helpfulness of online mental health Q&A responses. By incorporating DTA-related language cues into the ELM framework, we aim to explore how both central and peripheral routes influence users' perceptions of helpfulness. Specifically, we explored how the DTA dimensions influenced perceived helpfulness through the central and peripheral routes, and the interaction between central and peripheral cues. This integration of ELM and DTA provides a novel theoretical

framework for understanding the mechanisms underlying perceived helpfulness in online mental health Q&A platforms. It also offers practical insights for optimizing the quality of mental health support services by highlighting the importance of both content quality and relational aspects in digital therapeutic interactions.

Natural language processing methods in the field of mental health

NLP is a subset of artificial intelligence that learns data structures from written and spoken language. Given the clinical relevance of human language in psychotherapy and its recordable nature, NLP has become a key tool that provide therapeutic feedback to clinicians and patients.¹⁵ LIWC is a tool that applies NLP. LIWC reports the percentage of words in a text that belong to categories defined by its dictionary, including grammatical, psychological, and content categories.^{31,32} In addition to classifying and counting words, LIWC provides a statistical overview of word usage in predefined and psychologically meaningful categories.³³ LIWC and similar tools enabled researchers to analyze language use patterns in patients with depression and other mental health conditions.²⁰

The relatively widespread application of NLP in text-based psychotherapy includes the automatic detection of therapeutic alliance. Therapeutic alliance is a significant predictor of psychotherapy outcomes, including online text counseling.¹⁶ However, the current gold standard assessment of the alliance is based on self-report and observer-based qualitative coding of therapeutic interactions, which are labor-intensive and time-consuming.¹⁵ Advances in LIWC and therapeutic language analysis provide a path for automatically identifying therapeutic alliance.^{15,29} Some existing studies have explored language cues of the therapeutic alliance. For example, the frequency of first-person pronouns used by therapists and the syntactic transformations in a patient's speech are potential indicators of the therapeutic alliance.^{12,15} The level of language style matching (LSM), defined as the similarity in the ratio of function words in a dyadic interaction, reflects the extent to which conversation partners automatically coordinate their language styles to achieve shared goals between the patient and therapist, making it an effective predictor of the therapeutic alliance.³⁴ Affect words,^{12,35} differentiation words,¹² and linguistic markers predict therapeutic alliance,³⁶ while weaker alliances correlate with denial and reflective words.³⁷

Research questions

The ELM, widely used in management research, explains how language cues shape decision-making via central (in-depth analysis) and peripheral (heuristic cues) routes. The central route involves individuals' in-depth analysis and evaluation of the content, while the peripheral route

depends on features such as the credibility of the source or the attractiveness of the presentation.¹¹ Accepting information from online mental health Q&A platforms is essentially a decision-making process that involves evaluating the information source and analyzing the content in-depth.³⁸ Thus, ELM's applicability can be extended to mental health scenarios, helping us understand how users decide to adopt mental health information based on different cues.

Currently, the factors affecting the perceived helpfulness of online Q&A platforms include answer length, answer quality, and the credibility of the information source.¹⁰ From the perspective of ELM theory, peripheral route-related cues often include the credibility of the information source, length of the information, etc., while central route cues include the quality of the answers, etc. Overall, although both routes affect individual attitudes and behaviors, research indicates that the central route tends to be more influential.³⁹ However, central route variables—particularly how language cues integrate with mental health-specific factors—remain understudied in online Q&A contexts. To fill this research gap, this study introduces a widely used concept in the field of mental health Q&As: DTA. Research indicates that therapeutic alliance, by establishing stable psychological therapy relationships, broadly affect the emotional connection between respondents and counselors, thereby producing high-quality therapeutic outcomes.²⁸ Extended to the online mental health domain, DTA theory has also been proven to affect the effectiveness of psychological interventions.⁴⁰ By incorporating key variables of the DTA model, such as emotional bonds, therapeutic tasks, and goal consistency,¹⁴ into the central route cues of ELM, we can more comprehensively understand users' information processing and decision-making behavior on mental health Q&A platforms. Moreover, our study will consider peripheral route influencers commonly used in previous research, such as text length.⁴¹ Thus, this study aims to optimize ELM from the perspective of DTA, specifically by analyzing emotional bonds, therapeutic tasks, and goal consistency, to enhance the cues of the central and peripheral routes of ELM.

Given that the overall DTA and its segmented dimensions have not been clearly established as having significant predictive and explanatory power regarding the perceived helpfulness of mental health platforms, we propose the first research question:

RQ1: To what extent does DTA predict and explain perceived helpfulness in mental health Q&A?

From the perspective of ELM theory, how do the central and peripheral routes affect perceived helpfulness? Most previous studies have explored the effects of these two routes in isolation. However, some researchers point out that the boundaries between these two routes are sometimes

blurred, highlighting the need to examine the interactions between central and peripheral cues.⁴² A study on social media privacy decisions indicated that peripheral cues modulate the relationship between central cues and self-disclosure intentions.⁴³ However, whether the modulating effect of peripheral cues applies to DTA and its segmented dimensions is still unknown. Therefore, we pose the second research question:

RQ2: How do DTA and its dimensions (emotional bonds, therapeutic tasks, goal consistency) interact to shape perceived helpfulness in mental health Q&As?

Considering that NLP technology makes it possible to quantify DTA and other language elements related to the effectiveness of mental health Q&As based on language cues, we first built a Q&A dataset containing 19,682 messages from users' seeking help and counselors' responses. The perceived helpfulness of the counselor's replies serves as an indicator of the quality of counselor Q&A. We used computational linguistics and explainable artificial intelligence (XAI) methods suitable for large-scale Q&A text analysis. We established an explainable predictive model for counselor reply perceived helpfulness based on DTA-related language cues and explored the effects of these language cues on the perceived helpfulness of mental health Q&As.

Methods

Data collection

This study employed a quantitative design to investigate how DTA influence the perceived helpfulness of responses on online mental health Q&A platforms. Specifically, the study integrates NLP, explainable machine learning, and causal inference methods to propose a computational framework for predicting the perceived helpfulness.

The study was conducted over a five-month period, from November 2022 to March 2023, using data from the YiXinLi platform, one of China's largest online mental health platforms. YiXinLi is widely used across China and offers anonymous psychological Q&A services to millions of users. This platform was selected due to its robust user base and the availability of high-quality, large-scale Q&A data relevant to the study objectives.

All counselors on the platform hold nationally recognized psychological counseling licenses in China, with relevant educational background and clinical experience. The YiXinLi platform requires counselors to submit valid credentials and go through a rigorous screening process to ensure their professional competence. Additionally, the platform provides regular professional training for counselors, covering the latest psychological counseling techniques, ethical guidelines, and client service skills to enhance their professional competence

and service quality. The counselors included in the dataset for this study all have at least three years of experience in online or offline psychological counseling and have received high ratings and positive feedback on the platform, indicating a high level of professional competence and service quality.

Data were extracted from the YiXinLi platform's Q&A section, where users can anonymously post their questions and receive mental health services from platform counselors. The dataset includes 10,903 unique help-seeking questions and 19,682 counselor responses. The number of upvotes for each response served as the score of perceived helpfulness, which was used as the target variable in the machine learning models. The upvote data are objective and extensive, effectively reflecting users' actual recognition of and satisfaction with the responses. Table 1 provides a comprehensive statistical analysis of the mental health Q&A data used in this study.

We pre-processed the mental health quiz text in the above dataset. The Chinese NLP tool 'jieba' provides an easy-to-use interface to Chinese corpora and lexical resources, and is able to perform tasks such as analysis, deletion of stop words, and keyword extraction.⁴⁴ We utilized this tool to segment the mental health quiz text described above. A lexicon for the mental health domain was constructed using Gensim and an embedded representation of mental health topics was derived. We used the Chinese Complex Sentiment Analysis Tool (Cnsenti) to identify seven types of sentiment statistics (positive, joy, sadness, anger, fear, disgust, and surprise) from the Q&A texts. cLIWC (Chinese Language Inquired Word Count) was used to extract psycholinguistic cues from Q&A texts statistical features. Building on our previous study,¹⁸ this study used linguistic inquiry and word count (LIWC) to count thematic coherence, linguistic style similarity, and affective similarity in Q&A texts to characterize the therapeutic alliance in mental health Q&A texts. We confirmed the accuracy of our feature extraction results by manually examining 20 random Q&A texts related to mental health. All extracted feature values were validated.

Feature engineering

In this research, we explore the concept of perceived helpfulness from the public's perspective on the quality of mental health Q&As. Using computational linguistics and explainable machine learning methods suitable for large-scale discourse analysis of Q&As, we constructed a predictive model. We utilized language cues from the users' queries, counselors' responses, and their synchronous interactions to predict and explain the perceived helpfulness of mental health Q&As. This study includes features such as LIWC, thematic consistency, language style, and emotional similarity. LIWC is a text-based analysis tool that captures the psychological dimensions of language use.³³ It provides a reliable method to measure and capture the language behavior of visitors and counselors,

Table 1. Descriptive statistics of psychological Q&A data.

	Helpfulness	Number of answers	Total counselees helped by the counselor	Reward amount in RMB	Number of views of the question	Word length of the helpseeker's question	Word length of the counselors' answer
Mean	3.90	10.07	1676.68	9.10	268.93	175.73	514.03
Std	2.39	11.75	3013.34	14.06	322.24	89.59	323.94
Min	1.00	1.00	0.00	0.00	4.00	9.00	3.00
Max	25.00	87.00	19,253	100.00	4931.00	321.00	2970.00

including emotional expression, psychological states, social relationships, cognitive processes, and the vocabulary and sentence structures used.³³ Thematic consistency reflects the coherence of the topics discussed within the text,⁴⁵ helping us understand whether the user-counselor Q&As revolve around the same theme. LSM measures the similarity of language styles within the text.⁴⁶ It helps us understand whether the language styles of users and counselors are consistent and whether this match affects the public's evaluation of the perceived helpfulness of mental health Q&As. Emotional similarity reflects whether the emotions of the user and counselor in the dialogue are consistent. This feature, extracted using emotional analysis methods, can help us understand how the emotional interactions between users and counselors affect perceived helpfulness. Overall, we selected the aforementioned language cues based on their availability in large-scale mental health Q&A text analysis and their potential role in predicting perceived helpfulness. In the machine learning models constructed below, these language cues form a key feature set that can predict and explain the perceived helpfulness of mental health Q&As.

Machine learning and causal inference methods

We use perceived helpfulness as the output target for our prediction model and employ 'Language Inquiry and Word Count' and therapeutic alliance features as inputs, utilizing explainable machine learning techniques to develop an interpretable prediction model. These models are designed to evaluate the perceived helpfulness within mental health Q&A interactions.

Consistent with best practices for comparative machine learning,⁴⁶ we implemented five regression algorithms spanning distinct inductive biases: Regularized linear models (Ridge, LASSO) to handle multicollinearity; support vector regression (SVR) for kernel-based non-linear modeling; and tree-based ensembles (random forest—RF, XGBoost) to capture complex feature interactions. This diversity ensures robustness against model class limitations while enabling Shapley additive explanations (SHAP)-based interpretation of predictor effects.⁴⁷ Ridge and LASSO address collinearity and overfitting; Ridge limits

model coefficients through a penalty term to reduce noise sensitivity,⁴⁸ while LASSO allows coefficients to shrink to zero, achieving automatic variable selection and enhancing interpretability.⁴⁸ Support vector regression, an application of support vector machine, effectively handles nonlinear and high-dimensional data problems and has been used in predicting perceived helpfulness.⁴⁹ Ensemble methods like random forest and XGBoost improve prediction accuracy and prevent overfitting by combining multiple decision trees and applying regularization terms, respectively.^{50,51} We used recursive feature elimination with cross-validation and XGBoost model fitting to select effective features from the original feature set, filtering out 338 significant features influencing the prediction of perceived helpfulness. Additionally, we used GridSearchCV to select the best-performing hyperparameters from a set of parameters. To further enhance prediction accuracy, interpretability, and generalization performance, we employed sensitivity analysis and feature pruning methods. We selected an important subset of features from the effective feature set and retrained the model using these features.

To gain a deeper understanding of how language behaviors related to users, counselors, and their synchronous interactions influence perceived helpfulness, we applied explainable machine learning and causal inference methods. Shapley values, widely used in cooperative game theory, represent the responsibility a feature holds for changes in model output.⁴⁷ SHAP values significantly enhance the transparency of machine learning and have been applied in many research and industrial scenarios.⁵² Thus, we initially used SHAP values to interpret our constructed prediction models, helping us understand the contribution of each feature to the prediction outcome. Recently, the double machine learning (DML) method has become a popular causal inference approach, estimating treatment effects through a two-stage machine learning process.⁵³ When dealing with a large number of confounders or variables needing control, traditional statistical methods may be inadequate, especially when it is difficult to model the variables adequately through parametric functions. The DML method, leveraging machine learning technology, effectively identifies high-dimensional or non-parametric

relationships and has been widely used in network scenarios. To further investigate the interactive effects of different language behaviors on perceived helpfulness, we opted for the econml framework developed by Microsoft.⁵⁴ Using its linear DML method, we estimated the conditional average treatment effect (CATE) under specific conditions.⁵⁵ CATE measures the average effect of treatment on the target variable under a specific condition or subgroup.⁵⁶ To ensure our estimates were reliable and robust, we also conducted robustness tests using the Dowhy library, which helps confirm whether our causal effect estimates are influenced by unobserved confounding factors. The Dowhy library provides tools to simulate different scenarios and check whether the estimates remain stable. In summary, by combining explainable machine learning with DML models, we not only predict perceived helpfulness but also gain a deeper understanding of the causal relationships behind it. This provides a more comprehensive perspective, helping us better understand the underlying language behavior mechanisms in online mental health Q&As.

Model evaluation

In this study, as the output target of the prediction model, perceived usefulness is calculated from the number of upvotes that users on the platform have on the counselor's responses. Specifically, the number of upvotes received for each counselor response is considered as the perceived usefulness score of that response. The number of upvotes not only reflects the extent to which the questioner recognizes the response, but also includes other users' evaluations on the quality of the response. Therefore, as an indicator of perceived usefulness, the number of upvotes has a high degree of objectivity and validity. The predictive performance of the models constructed in this study was evaluated using four widely applied metrics in regression models, including root mean square error (RMSE) and mean absolute percentage error (MAPE), as defined in Equations (1) and (2). RMSE measures the difference between predicted and actual values, while MAPE quantifies prediction accuracy in percentage terms, providing a relative measure of the prediction error.

Where represents the total number of observations in the mental health Q&A dataset, \hat{y}_i is the actual perceived helpfulness of the observation in the dataset, and is the predicted perceived helpfulness of the i observation from the regression model. Overall, through the methods and processes described, we can explore the mechanisms by which language behavior influences perceived helpfulness in online mental health Q&As.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2} \quad (1)$$

$$\text{MAPE} = \frac{1}{n} \sum_i \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2)$$

The main methods and research questions of this study are presented in Figure 1. Specifically, we first utilize computational linguistics methods and machine learning regression algorithms to construct interpretable prediction models for the perceived helpfulness of mental health Q&As. Based on this, we use sensitivity analysis and SHAP methods to identify key language cues influencing the perceived helpfulness of mental health Q&As. Finally, we analyze the interaction effects between different types of language cues using DML methods and conduct robustness tests. Through this large-scale data experimental process, we aim to understand the language behavior mechanisms behind online mental health Q&As.

Results

Performance analysis of perceived helpfulness prediction models of mental health Q&As

Performance differences Among algorithm models for predicting perceived helpfulness of mental health Q&As. We constructed prediction models for the perceived helpfulness of mental health Q&As using seven types of regression prediction algorithms: Linear regression (LR), Ridge Regression (Ridge), Lasso Regression (Lasso), SVR, neural network (NN), RF, and XGBoost. We evaluated the predictive performance of each algorithm on the mental health Q&A data. The results are shown in Table 2, with RF performing best across all metrics: Its RMSE is the lowest, indicating the smallest prediction error; the MAE is the lowest, indicating the smallest median absolute error; and the MAPE is relatively low, indicating a smaller relative error in predictions. Conversely, Lasso regression performs the worst across all metrics: Its RMSE, MAE, and MAPE are the highest, indicating the largest prediction error. LR, Ridge Regression, SVR, NN, and XGBoost perform between RF and Lasso in various metrics. NN and SVR outperform LR and Ridge Regression in RMSE and MAPE values but perform slightly worse in MAE values. XGBoost performs well on some metrics, with the lowest MAPE value and a lower MAE value, second only to RF, indicating the model's relatively smaller median absolute error. Additionally, this study compared the predictive performance of different algorithms when using the single counselor involvement feature versus the combination of counselor involvement and DTA features. Overall, across all algorithms used, the combination of counselor involvement and DTA features outperforms using the single counselor involvement feature alone in nearly all performance metrics. Therefore, DTA features help improve prediction accuracy.

Feature impact analysis on perceived helpfulness prediction of mental health Q&As. We used the XGBoost model to predict

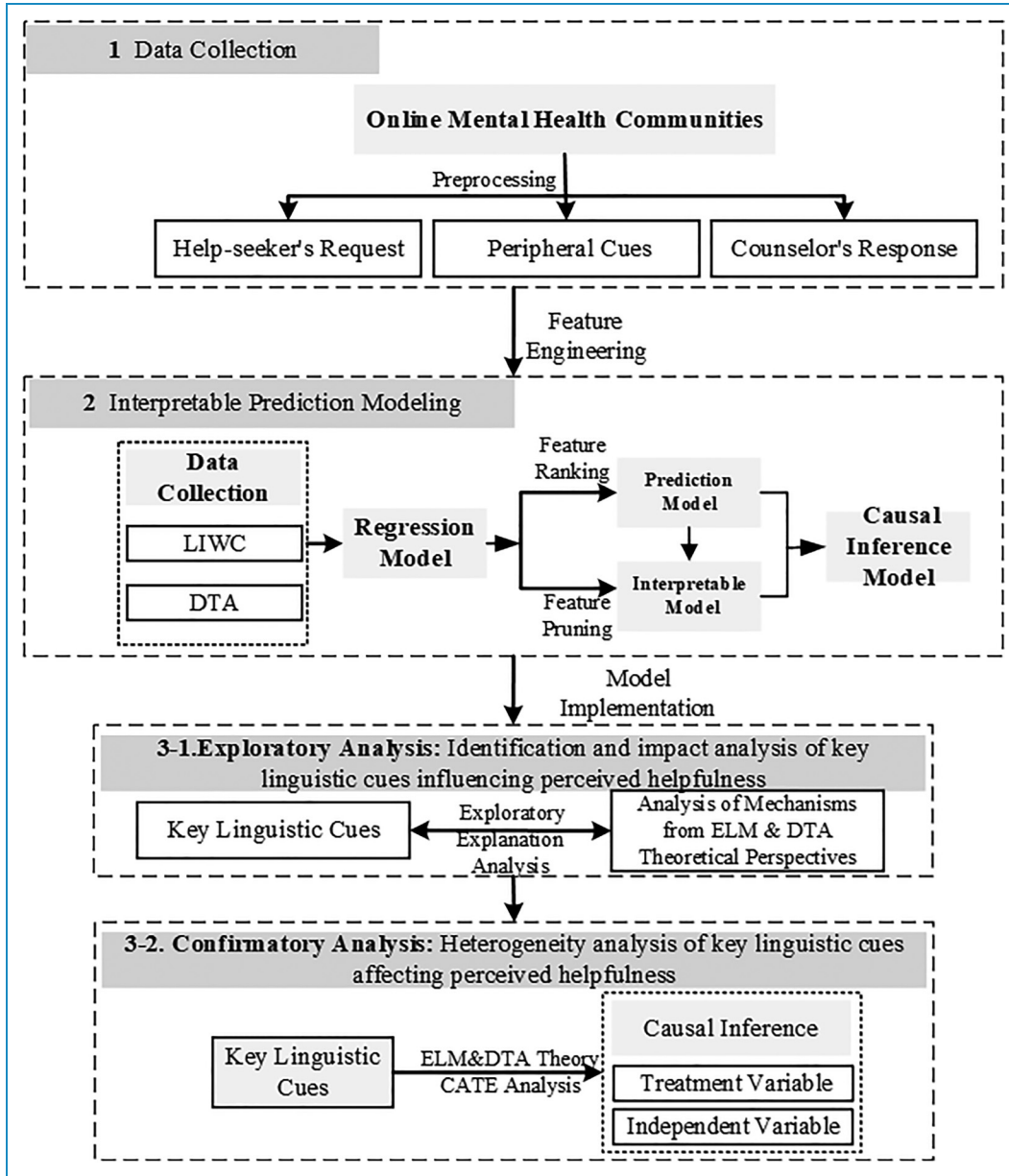


Figure 1. Experimental and methodological flow of this study.

perceived helpfulness and selected five categories of features influencing the perceived helpfulness of mental health Q&As: “Emotional Bond,” “Therapeutic Task,” “Goal Consistency,” “Control Factor,” and “Counselor Involvement.” To evaluate the impact of these features on prediction performance, we first conducted feature ablation experiments on these five features. The experimental results are shown in Table 3. “Counselor Involvement” is the most important feature, with excellent predictive performance when used alone ($R^2 = 0.1114$, $RMSE = 0.8406$, $MAPE = 23.2327\%$), and generally improves prediction performance when combined with other features. This indicates that the “Counselor Involvement” feature contains key information on client

satisfaction. The predictive performance of “Emotional Bond,” “Therapeutic Task,” “Goal Consistency,” and “Control Factor” features is poor when used alone.

We further observed the incremental predictive impact of features and the prediction effectiveness of feature combinations through feature ablation experiments, with results shown in Table 3. Firstly, from the perspectives of RMSE and MAPE, when we combine the four factors of “Emotional Bond,” “Therapeutic Task,” “Goal Consistency,” and “Counselor Involvement,” the RMSE value is 0.8175 and the MAPE value is 0.225, representing the best performance among all combinations. When we add the “Control Factor” to this combination, the RMSE and MAPE values slightly increase to 0.8175 and

Table 2. Performance of prediction algorithms,

Metrics	RMSE		MAE		MAPE (%)	
	Counselor involvement	Counselor involvement + DTA	Counselor involvement	Counselor involvement + DTA	Counselor involvement	Counselor involvement + DTA
LR	0.828	0.825	0.5365	0.5342	32.49	32.2
Ridge	0.8279	0.8155	0.5371	0.52459	32.50	31.25
Lasso	0.8932	0.88	0.7126	0.7014	38.14	36.95
SVR	0.8467	0.834	0.5073	0.4951	29.69	28.40
NN	0.7915	0.779	0.4611	0.4496	29.05	27.85
RF	0.7656	0.753	0.4307	0.4186	28.31	27.10
XGBoost	0.8577	0.845	0.4338	0.4219	24.71	23.52

RMSE: root mean square error; MAPE: mean absolute percentage error; LR: linear regression; SVR: support vector regression; NN: neural network; RF: random forest

0.2255, respectively. This indicates that adding the “Control Factor” does not significantly improve the model’s incremental predictive effect. Secondly, regarding the best-performing feature combination, the combination of “Emotional Bond,” “Therapeutic Task,” “Goal Consistency,” and “Counselor Involvement” has the smallest prediction error, with the lowest RMSE and MAPE values, indicating superior performance in the prediction model. Finally, regarding the incremental predictive impact of features, when using only the “Counselor Involvement” feature for prediction, its RMSE value is 0.8406 and the MAPE value is 0.2323. When combined with other features, both RMSE and MAPE values decrease, demonstrating that “Counselor Involvement” has certain interactive effects with other features, enhancing the model’s prediction accuracy.

Interpretability analysis of mental health Q&A perceived helpfulness prediction model

To further enhance the interpretability of the machine learning prediction model, we introduced a feature refinement process to improve the model’s interpretability and generalizability. As shown in Figure 1, we selected features with significant impact on predictions, thereby reducing dimensionality and eliminating redundant information. The horizontal axis represents the order of feature importance, incorporating the Top-N features into the prediction model sequentially, while the vertical axis represents MAPE. Notably, features beyond the 20th feature show minimal changes in MAPE. While therapist identity was not explicitly modeled, linguistic style standardization through Counselor Involvement and DTA features—validated in

prior telepsychotherapy research—mitigated nesting effects.⁵⁷ Our bootstrap sensitivity analysis further confirmed model stability across therapist subgroups.

Model refinement involves a series of optimization steps to improve the performance of the machine learning model. This can include tuning hyperparameters, selecting more optimal algorithms, optimizing model architecture, and more. After feature pruning and ablation, we further refined the model to enhance its predictive and interpretive performance. We compared the performance of the refined model and found that the XGBoost prediction algorithm’s evaluation metrics showed improvements: RMSE decreased from 0.826 to 0.823, an improvement of 0.003; MAPE decreased from 22.8% to 22.7%, an improvement of 0.1%. Based on the refined model, we analyzed the impact of language cues from three sources—help-seeker, counselor, and their interaction synchrony—on the perceived helpfulness of responses and explored the interactive effects between language cue elements. In order to further improve the interpretability of machine learning prediction models, we introduced the SHAP value method. SHAP value is based on cooperative game theory and can quantify the contribution of each feature to the model output. Specifically, SHAP value measures the marginal contribution of a feature to the prediction result when all possible feature combinations are considered. Figure 3 shows the distribution of SHAP values for different features. The horizontal axis represents the SHAP value, which represents the contribution of the feature to the model prediction results; the vertical axis lists the features, which are sorted from top to bottom according to the average absolute value of the SHAP value. The higher the feature, the greater the impact on the model output. Each point in the figure represents a

Table 3. Impact of feature sets on Q&A usefulness prediction model performance.

Feature type	Feature set	Evaluation metrics			
		R ²	RMSE	MAE	MAPE
Single feature	Emotional bond	0.0124	0.8855	0.4213	0.2526
	Therapeutic task	0.0787	0.9257	0.4563	0.2718
	Goal consistency	0.0212	0.8999	0.4725	0.2673
	Counselor involvement	0.1114	0.8406	0.3954	0.2323
	Control variable	0.0605	0.9164	0.8397	0.2707
Combined features	Emotional Bond + therapeutic Task + goal Consistency + cCounselor involvement	0.1602	0.8175	0.376	0.225
	Therapeutic task + goal Consistency + counselor involvement	0.1582	0.8184	0.3787	0.226
	Therapeutic task + goal Consistency + control Variable + counselor Involvement	0.1568	0.819	0.3777	0.2266
	Emotional bond + goal consistency + counselor involvement	0.1513	0.8216	0.3805	0.2279
	Goal consistency + control variable + counselor involvement	0.1513	0.8213	0.3869	0.2292
	Emotional bond + goal consistency + control variable + counselor involvement	0.151	0.8217	0.3828	0.2283
	Goal consistency + counselor involvement	0.1484	0.8229	0.3853	0.2294
	Emotional bond + therapeutic task + control variable + counselor involvement	0.1365	0.8286	0.3881	0.2291
	Emotional bond + therapeutic task + counselor involvement	0.132	0.8309	0.3857	0.2284
	Therapeutic task + control variable + counselor involvement	0.1303	0.8316	0.3862	0.2299
	Emotional bond + control variable + counselor involvement	0.1287	0.8322	0.3927	0.2322
	Control variable + counselor involvement	0.1256	0.8333	0.3975	0.2337
	Emotional bond + counselor involvement	0.1253	0.834	0.393	0.2316
	Therapeutic task + counselor involvement	0.1134	0.8398	0.3819	0.2286

Note: R² Score: The R² score is the coefficient of determination of the model, reflecting its predictive capability. A higher score indicates that the model can better explain the variability in the observed data. In this table, a higher R² score indicates better model performance. RMSE: Root mean square error is the average difference between the model's predicted values and the actual values. A lower RMSE value indicates a higher accuracy of the model predictions. MAE: Mean absolute error is the average absolute difference between the model's predicted values and the actual values. A lower MAE value indicates smaller prediction errors. MAPE: Mean absolute percentage error is the average percentage difference between the model's predicted values and the actual values. A lower MAPE value indicates smaller relative errors in the model.

sample, and the color represents the size of the feature value. Red indicates a larger feature value, and blue indicates a smaller feature value. The position of the point reflects the positive or negative impact of the feature value on the prediction result: the right side indicates that the feature value has a positive impact on the perceived usefulness, and the left side indicates a negative impact. Through the SHAP chart, we can intuitively

identify which features play a key role in predicting perceived usefulness and understand how these features specifically affect the model's prediction results. This has important guiding significance for further optimizing counselors' language use and improving the service quality of online mental health Q&A. We used SHAP values to explain the machine learning model's prediction results, helping us understand the

contribution of each feature to the model output. For the impact of counselor involvement-related language cue features on perceived helpfulness, we obtained the SHAP values of relevant language cues, as shown in Figure 3. The length of counselor replies (Inv_Couns_WordCount) and average words per sentence (Inv_Couns_WordPerSentence) show that as these feature values increase, the model output value increases; conversely, smaller feature values decrease the model output value, indicating a positive impact on perceived helpfulness. Conversely, the greater the time difference in help-seeker responses (Inv_Response_time_difference), the stronger the negative impact on perceived helpfulness.

Similarly, for the impact of therapeutic task consistency-related language cue features on perceived helpfulness, we obtained the SHAP values, as shown in Figure 2(b). The use of prepositions in counseling (Couns_CP_Preps), phrases ending in prepositions (Couns_CP_PrepEnd), insight-related vocabulary (Couns_CP_Insight), causal relationship vocabulary (Couns_CP_Cause), conjunctions or subordinating conjunctions (Couns_CP_Conj), and cognitive mechanism vocabulary (Couns_CP_CogMech) are all positively associated with perceived helpfulness. When counseling includes these language cues, counselees are more likely to perceive the counseling as effective, possibly because these cues help provide clear logical thinking, insight, and cognitive autonomy. From the perspective of the DTA, we found that the use of auxiliary verbs (DTA_LSSpreps_AuxVerb) is negatively associated with perceived helpfulness. This may suggest that excessive use of auxiliary verbs may introduce uncertainty or ambiguity, thereby reducing the clarity and effectiveness of counseling. Overall, our data reveal that in mental health Q&As, specific language cues, especially those related to cognitive processing, can significantly influence counselees' perceptions of counseling effectiveness. This provides a new perspective for further understanding the interaction between counselors and visitors.

For the impact of emotional bond-related language cue features on perceived helpfulness, we obtained the SHAP values, as shown in Figure 3(c). Specifically, when counselor disclosures related to anger-related vocabulary (Couns_Emo_Anger), sadness-related vocabulary (Couns_Emo_Sad), and fear-related vocabulary (Couns_Emo_Fear) increase, the model output value decreases. Conversely, when counselor disclosures related to happiness-related vocabulary (Couns_Emo_Happiness), the total number of emotion-related vocabulary (Couns_Emo_Affect), and kindness-related vocabulary (Couns_Emo_Goodness) increase, the model output value increases. Notably, anger-related vocabulary (Couns_Emo_Anger) is the most important feature, but its impact on model output is not unidirectional and is moderated by other features. Vocabulary related to anxiety (Couns_EmoAnx) also shows a similar situation and requires further analysis. Overall, the emotional disclosures of counselors, whether positive or negative, impact help-seekers' perceptions of the usefulness of counseling

responses. The direction and extent of this impact may be influenced by specific contexts and factors.

Finally, for the impact of control variable-related language cue features on perceived helpfulness, we obtained the SHAP values. As shown in Figure 3(d), we conducted an in-depth analysis of the SHAP values for each feature of the control variables. Among these, whether a bounty is set (bounties) is the most important feature, with its value positively correlated with model output. This means that when help-seekers set a bounty for a question, the model output value increases accordingly. Additionally, the number of answers to a question (Number_of_answers) does not have a unidirectional impact on model output, but may be moderated by other factors. These analysis results reveal the importance of specific features in control variables within the model prediction, as well as their relationships with model output.

Study of interaction effects between factors influencing perceived helpfulness of mental health Q&As

In further causal forest regression analysis, we focused on the independent language cue variables of "Emotional Connection" and "Therapeutic Task Consistency," as well as the impact of counselor involvement on perceived helpfulness. The estimated results of our proposed causal forest regression model are presented in Table 4. Model 1 assesses the impact of emotional connection and therapeutic task-related importance variables on Q&A usefulness. Specifically, the emotional connection-related language cue variables include Couns_Emo_Affect, Couns_Emo_Disgusting, and DTA_MS_Fear, i.e., the frequency of emotional and disgust words used in counselor replies, and the similarity ratio of fear emotion words used by counselors and help-seekers. The therapeutic task-related language cue variables include Couns_CP_Preps, Couns_CP_Discrep, Couns_CP_PrepEnd, Couns_CP_Insight, Couns_CP_Conj, and DTA_LSSpreps_Preps, i.e., the prepositions, discrepancy words, end-of-sentence prepositions, foreground words, conjunctions used in counselor replies, and the similarity in preposition use between counselors and help-seekers. Models 2 and 3 include emotional bond-related importance variables and different mediating variables, such as Couns_WordCount and Inv_Response_time_difference, i.e., the word frequency in counselor replies and the time difference between help-seeker questions and counselor replies. In Table 4, Model 1 shows the main effects, i.e., the impact pathways from emotional connection and therapeutic task-related language cues to comment usefulness are significantly positive. Model 2's results indicate a significantly positive interaction coefficient between counselor reply length and emotional connection-related variables ($\beta = 4.486, p < 0.001$), while Model 3's results indicate a significantly negative interaction coefficient between help-seeker questions and counselor answer time difference and emotional connection-related variables ($\beta = -2.283, p < 0.05$).

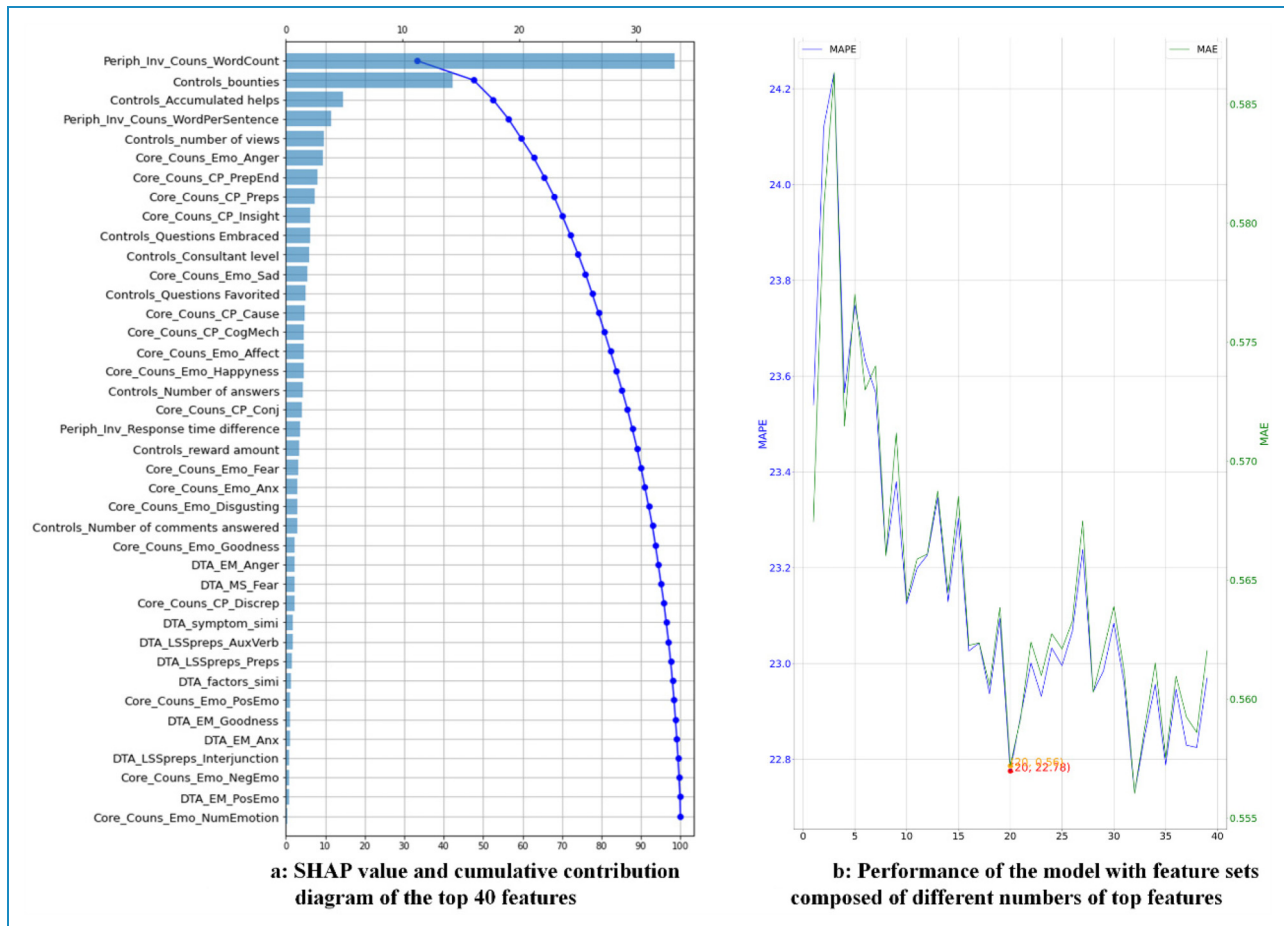


Figure 2. Feature importance ranking and mean absolute percentage error (MAPE) values under different numbers of top features based on XGBoost algorithm.

Discussion

Principal findings

This study constructed an evaluation model for assessing the effectiveness of counselor responses in a mental health Q&A context. It explored the impact of different language cues on perceived helpfulness from the perspectives of the DTA and the refined possibility model theory. In analyzing the predictive performance of the mental health Q&A perceived helpfulness model, we evaluated seven different regression prediction algorithms. Among them, RF and XGBoost performed well in evaluation metrics. Consistent with prior research on depression diagnosis,⁵⁸ the XGBoost model has shown excellent performance in predicting mental health conditions. Additionally, our findings align with previous studies emphasizing the importance of feature selection in clinical prediction models.⁵⁹ Through model refinement, we further improved the predictive performance of the XGBoost algorithm, with both RMSE and MAPE decreasing. In the feature impact analysis, we used the XGBoost model to assess the relative

predictive impact of five key features. The results showed that “Counselor Involvement” was the most important feature, demonstrating strong predictive performance, and further improving overall prediction accuracy when combined with other features. Conversely, the features “Emotional Bond,” “Therapeutic Task,” “Goal Consistency,” and “Control Factor” performed relatively weaker predictive power when used independently. Feature ablation experiments further confirmed that the combination of “Emotional Bond,” “Therapeutic Task,” “Goal Consistency,” and “Counselor Involvement” had the lowest RMSE and MAPE values, indicating the best predictive effect of this combination. This highlights the key role of the DTA in the field of online mental health Q&A, a form of online psychological support service. These findings are consistent with previous research and others in the field of digital psychological therapy on the DTA.¹⁴ Overall, these analysis results provide insights on how to optimize the mental health Q&A perceived helpfulness prediction model.

To further investigate the underlying mechanisms, we used the SHAP value analysis to examine the contribution

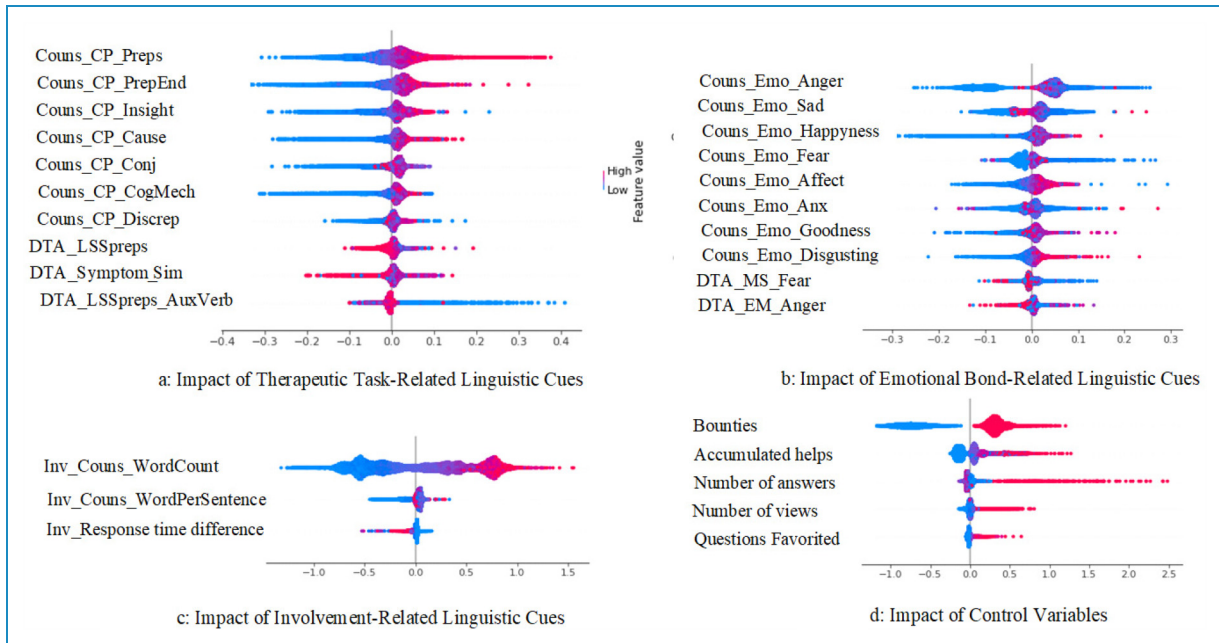


Figure 3. Summary of Shapley additive explanations (SHAP) values showing the impact of different types of language cues on perceived helpfulness. (Note: The above diagram shows the impact of different features on model output. The horizontal axis represents the SHAP value, indicating the contribution size of each feature to the model output. The vertical axis lists the features, sorted from top to bottom by the average absolute value of their SHAP values, with higher positions indicating more important features. Each point in the diagram represents a sample, with the color indicating the feature value—redder points indicate larger values, and bluer points indicate smaller values. The horizontal position of each point reflects the SHAP value, with positions further to the right indicating a larger positive impact, and positions further to the left indicating a larger negative impact. The vertical position indicates feature importance, with higher positions denoting greater importance and lower positions denoting lesser importance. The width of each point represents the number of samples, with wider points indicating more samples at that location).

of each feature to the model output. The analysis results showed that language cues related to ‘Counselor Involvement,’ such as the length of counselor replies and the average number of words per sentence, were positively related to perceived helpfulness. This finding is consistent with previous research suggesting that user involvement and cognitive effort can affect their perception of information service usefulness.⁶⁰ Similarly, language cues related to “Emotional Bond,” particularly the usage patterns of emotional vocabulary usage by counselors, affected the perceived helpfulness of counseling, consistent with the findings of previous research,⁶¹ who found that the use of emotional vocabulary is closely related to the effectiveness of counseling. Overall, these findings reveal that in mental health Q&As, different sources of language cues significantly impact perceptions of counseling effectiveness. By integrating LIWC with existing information systems and digital therapeutic research, we provide a new perspective for further understanding the interaction between counselors and visitors.

Additionally, cognitive effort and cognitive involvement are key terms describing the level of mental activity and attention individuals devote to a specific task. High levels of cognitive involvement are associated with more positive

emotional responses and better task performance.⁶² In the field of mental health, cognitive involvement is crucial for therapeutic outcomes and patient satisfaction.^{63,64} The number of words spoken by counselors and the time difference in responses may represent their cognitive involvement and effort, both of which have been shown to be related to therapeutic outcomes.⁶⁵ In the context of Chinese culture, which emphasizes collectivism and relationship-building, these linguistic markers may also reflect the cultural values of harmony and indirect communication.^{66,67} In the study of interaction effects between factors influencing perceived helpfulness of mental health Q&As, we conducted an in-depth analysis of the impact of the independent variables “Emotional Connection” and “Therapeutic Task” on perceived helpfulness, as well as the interactive effects of counselor involvement. The results showed that the number of words spoken by counselors and the timeliness of their response enhance the impact of these independent variables on perceived helpfulness to varying degrees. For example, for “Emotional Connection,” both the number of words spoken by counselors and the time difference in responses have a significant positive effect on its relationship with perceived helpfulness. These findings reveal significant effective factors in the interactive process

Table 4. Random forest regression results for perceived helpfulness of mental health Q&As

Variable type	Variables	Model 1	Model 2	Model 3
Therapeutic task	Couns_CP_Preps	7.384***	7.389***	7.394***
	Couns_CP_Discrep	6.777***	6.787***	6.782***
	Couns_CP_PrepEnd	3.260***	3.221***	3.293***
	Couns_CP_Insight	3.398***	3.390***	3.388***
	DTA_LSSpreps_Preps	3.148**	3.155**	3.151**
Affective bonds	Couns_Emo_Affect	3.322***	3.361***	3.325***
	Couns_Emo_Disgusting	3.061**	3.080**	3.091**
	DTA_MS_Fear	2.880**	2.887**	2.898**
	(Couns_Emo_Affect + Couns_Emo_Disgusting + DTA_MS_Fear)×Couns_WordCount		4.486***	4.489***
	(Couns_Emo_Affect + Couns_Emo_Disgusting + DTA_MS_Fear)×Inv_Response_time_difference			−2.283*

Note: *indicates $p < 0.05$, **indicates $p < 0.01$, *** indicates $p < 0.001$ significance level.

between counselors and help-seekers in online psychological Q&As, providing new perspectives for further research.

Theoretical implications

This study offers several theoretical implications. First, our research enriches the theoretical foundation of perceived helpfulness in mental health Q&As from the perspectives of the DTA and the refined possibility model, providing a theoretical reference for future online mental health Q&A research. Second, although an increasing number of people join mental health Q&A communities to seek support,¹⁰ only a few studies have explored the effectiveness of online mental health Q&As. Our research fills this gap by constructing a prediction model for perceived helpfulness in mental health Q&As, thereby enhancing understanding of the antecedents of perceived helpfulness. Counselor involvement has a significant predictive effect on perceived helpfulness, and when combined with the factors “Emotional Bond,” “Therapeutic Task,” “Goal Consistency,” and “Counselor Involvement,” the predictive effect is best. This makes online mental health Q&A platforms a bridge between counselors and visitors, effectively enhancing the perceived helpfulness of online Q&As.²⁶ Third, this study uses the SHAP value method to deeply understand the contribution of each feature to model output, finding that under different factors, different language cues contribute to perceived helpfulness to varying degrees, showcasing the complexity of predicting counseling effectiveness. Finally, this study demonstrates how to construct models


for perceived helpfulness using computational linguistics, XAI, and causal inference machine learning methods suitable for large-scale Q&A text analysis. These practices help expand research on online mental health Q&A platforms.

Practical implications

This study provides practical references for platform managers, counselors, and users. First, our research finds that perceived helpfulness can be predicted by counselor involvement. Therefore, platform managers need to standardize counselor levels, provide regular training to enhance counselor involvement, and encourage the formation of a reliable therapeutic alliance between counselors and users, thereby maximizing the effectiveness of psychological Q&A platforms. Second, this study finds that different language cues have varying predictive effects on perceived helpfulness. For example, regarding “Counselor Involvement,” the length of counselor replies and the average number of words per sentence are positively related to perceived helpfulness. In contrast, for language cues related to “Emotional Bond,” the pattern of emotional vocabulary usage by counselors directly affects the perceived helpfulness of counseling. The findings from this study have significant implications for clinical practice, particularly in the integration of digital tools and therapeutic alliance in mental health care. Firstly, counselors should aim to provide detailed and timely responses and positive emotional language in their responses to enhance perceived helpfulness. Clinical training programs for counselors could be adapted to emphasize these language

strategies, fostering stronger therapeutic relationships and improving patient outcomes. These skills can be integrated into both face-to-face therapy and online platforms to improve overall therapy effectiveness. Secondly, given that DTA plays a crucial role in enhancing therapeutic outcomes, digital platforms can integrate elements of DTA into their designs. For instance, chatbots or other automated systems can be programmed to exhibit empathy and create a sense of emotional connection with users.⁶⁸ Furthermore, considering the cultural context of collectivism in China, counselors may need to adopt more empathetic and indirect communication styles to build stronger therapeutic alliance. For instance, in collectivist cultures, more indirect and empathetic communication may be preferred, while in individualist cultures, direct and clear communication may be more effective.

ORCID iD

Sujie Meng  <https://orcid.org/0009-0004-6068-3779>

Statements and declarations

Author Contributions/CRedit

YH conceptualized the paper and analyzed the data. HL and MC provided constructive suggestions. SM wrote the original and modified version of manuscript. WW contributed to the preparation and revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

The authors received the following financial support for the research, authorship, and/or publication of this article: This project is funded by the National Natural Science Foundation of China (Award Number: 72204095), the Humanities and Social Science Young Scientist Program sponsored by the Ministry of Education of the People's Republic of China (Award Number: 22YJC880022), the China National Center for Mental Health and Prevention, China Education Development Foundation, Ministry of Education Student Service and Quality Development Center (Award Number: XS24A010), and Major Project of Philosophy and Social Science Research in Jiangsu Universities (Award Number: 2024SJZD068).

Conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Mo R, Fang ZZ and Fang JD. How to establish a digital therapeutic alliance between chatbots and users: The role of relational cues. *Adv Psychol Sci* 2023; 31: 669–683.
2. Sharma A, Lin IW, Miner AS, et al. Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat Mach Intell* 2023; 5: 46–57.
3. Huang Y, Liu H, Li S, et al. Effective prediction and important counseling experience for perceived helpfulness of social question and answering-based online counseling: An explainable machine learning model. *Front Public Health* 2022; 10: e817570.
4. Zhou J, Zuo M and Ye C. Understanding the factors influencing health professionals' online voluntary behaviors: evidence from YiXinLi, a Chinese online health community for mental health. *Int J Med Inf* 2019; 130: e103939.
5. Antoniou M, Estival D, Lam-Cassettari C, et al. Predicting mental health status in remote and rural farming communities: Computational analysis of text-based counseling. *JMIR Form Res* 2022; 6: e33036.
6. Hoermann S, McCabe KL, Milne DN, et al. Application of synchronous text-based dialogue systems in mental health interventions: Systematic review. *J Med Internet Res* 2017; 19: e267.
7. Navarro P, Bambling M, Sheffield J, et al. Exploring young people's perceptions of the effectiveness of text-based online counseling: mixed methods pilot study. *JMIR Ment Health* 2019; 6: e13152.
8. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information of information technology. *MIS Q* 1989; 13: 319–340.
9. Kwon D, Jeong P and Chung D. An empirical study of factors influencing the intention to use robo-advisors. *J Inf Knowl Manag* 2022; 21: 2250039.
10. Li J, Liu D, Wan C, et al. Empirical study of factors that influence the perceived usefulness of online mental health community members. *PSYCH J* 2023; 12: 307–318.
11. Petty RE and Cacioppo JT. The elaboration likelihood model of persuasion. *Adv Exp Soc Psychol* 1986; 19: 123–205.
12. Jacques J. A corpus linguistic analysis of counseling alliance ruptures: a pilot study. 2022.
13. Heim E, Rotger A, Lorenz N, et al. Working alliance with an avatar: How far can we go with internet interventions? *Intern Interv* 2018; 11: 41–46.
14. Tremain H, McEnery C, Fletcher K, et al. The therapeutic alliance in digital mental health interventions for serious mental illnesses: Narrative review. *JMIR Ment Health* 2020; 7: e17204.
15. Ryu J, Heisig S, McLaughlin C, et al. A natural language processing approach reveals first-person pronoun usage and non-fluency as markers of therapeutic alliance in psychotherapy. *IScience* 2023; 26: e106860.
16. Fluckiger C, Del Re AC, Wampold BE, et al. The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy* 2018; 55: 316–340.
17. Sharma A, Lin IW, Miner AS, et al. Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat Mach Intell* 2023; 5: 46–57.
18. Huang Y, Liu H, Li S, et al. Effective prediction and important counseling experience for perceived helpfulness of social question and answering-based online counseling: an explainable machine learning model. *Front Public Health* 2022; 10: e817570.
19. Zhou J, Zuo M and Ye C. Understanding the factors influencing health professionals' online voluntary behaviors:

- evidence from YiXinLi, a Chinese online health community for mental health. *Int J Med Inf* 2019; 130: e103939.
20. Antoniou M, Estival D, Lam-Cassettari C, et al. Predicting mental health status in remote and rural farming communities: Computational analysis of text-based counseling. *JMIR Form Res* 2022; 6: e33036.
 21. Liu J and Gao L. Research on the characteristics and usefulness of user reviews of online mental health consultation services: A content analysis. *Healthcare* 2021; 9: e1111.
 22. Navarro P, Sheffield J, Edirippulige S, et al. Exploring mental health professionals' perspectives of text-based online counseling effectiveness with young people: Mixed methods pilot study. *JMIR Ment Health* 2020; 7: e15564.
 23. Montagni I, Cariou T, Feuillet T, et al. Exploring digital health use and opinions of university students: Field survey study. *JMIR Mhealth Uhealth* 2018; 6: e65.
 24. Prescott J, Hanley T and Gomez KU. Why do young people use online forums for mental health and emotional support? Benefits and challenges. *Br J Guid Couns* 2019; 47: 317–327.
 25. Hanley T, Prescott J and Gomez KU. A systematic review exploring how young people use online forums for support around mental health issues. *J Ment Health* 2019; 28: 566–576.
 26. Henson P, Wisniewski H, Hollis C, et al. Digital mental health apps and the therapeutic alliance: Initial review. *Bjpsych Open* 2019; 5: 15.
 27. Probst GH, Berger T and Fluckiger C. The alliance-outcome relation in internet-based interventions for psychological disorders: A correlational meta-analysis. *Verhaltenstherapie* 2019; 29: 182–195.
 28. Martin DJ, Garske JP and Davis MK. Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *J Consult Clin Psychol* 2000; 68: 438–450.
 29. Goldberg SB, Baldwin SA, Riordan KM, et al. Alliance with an unguided smartphone app: Validation of the digital working alliance inventory. *Assessment* 2022; 29: 1331–1345.
 30. Arndt A, Rubel J, Berger T, et al. Outpatient and self-referred participants: Adherence to treatment components and outcome in an internet intervention targeting anxiety disorders. *Internet Interv- Appl Inf Technol Ment Behav Health* 2020; 20: 100319.
 31. Pennebaker JW, Chung CK, Ireland M, et al. *The development and psychometric properties of LIWC2007*. Austin, TX: LiwcNet, 2007. <https://api.semanticscholar.org/CorpusID:180769814>
 32. Tausczik YR and Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 2010; 29: 24–54.
 33. Pennebaker JW, Boyd RL, Jordan K, et al. *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin, 2015.
 34. Van Doorn AK, Porcerelli J and Mueller-Frommeyer LC. Language style matching in psychotherapy: an implicit aspect of alliance. *J Couns Psychol* 2020; 67: 509–522.
 35. Yahya NH and Abdul RH. Linguistic markers of depression: insights from English-language tweets before and during the COVID-19 pandemic. *Lang Health* 2023; 1: 36–50.
 36. Vail AK, Girard JM, Bylsma LM, et al. Toward Causal Understanding of Therapist-Client Relationships: A Study of Language Modality and Social Entrainment. In: Proceedings of the 2022 International conference on multi-modal interaction, 2022. doi:10.1145/3536221.3556616
 37. Christian C, Barzilai E, Nyman J, et al. Assessing key linguistic dimensions of ruptures in the therapeutic alliance. *J Psycholinguist Res* 2021; 50: 143–153.
 38. Guo C, Guo X, Wang G, et al. What makes helpful online mental health information? Empirical evidence on the effects of information quality and responders' effort. *Front Psychol* 2022; 13: 985413.
 39. Cheng P, Wang W and Yang S. Doing the right thing: how to persuade travelers to adopt pro-environmental behaviors? An elaboration likelihood model perspective. *J Hosp Tour Manag* 2024; 59: 191–209.
 40. D'Alfonso S, Lederman R, Bucci S, et al. The digital therapeutic alliance and human-computer interaction. *JMIR Ment Health* 2020; 7: e21895.
 41. Fiske ST. Social cognition and social perception. *Annu Rev Psychol* 1993; 44: 155–194.
 42. SanJose-Cabezudo R, Gutierrez-Arranz AM and Gutierrez-Cillan J. The combined influence of central and peripheral routes in the online persuasion process. *Cyberpsychol Behav* 2009; 12: 299–308.
 43. Wang L, Hu HH, Yan J, et al. Privacy calculus or heuristic cues? The dual process of privacy decision making on Chinese social media. *J Enterp Inf Manag* 2020; 33: 353–380.
 44. Sun X, Liu Z and Huo X. Six-granularity based Chinese short text classification. *IEEE Access* 2023; 11: 35841–35852.
 45. Blei DM, Ng AY and Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003; 3: 993–1022.
 46. Caruana R and Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning 2006, Pune, India, 16–18 August 2018, pp.6.
 47. Lundberg SM and Lee SI. A unified approach to interpreting model predictions. In: 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, 4–9 December 2017, pp.30.
 48. Friedman J, Hastie T and Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010; 33: 1–22.
 49. Smola AJ and Schölkopf B. A tutorial on support vector regression. *Stat Comput* 2004; 14: 199–222.
 50. Liaw A and Wiener M. Classification and regression by randomForest. *R News* 2002; 2: 18–22.
 51. Chen T and Guestrin C. XGBoost: a scalable tree boosting system. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, 13–17 August 2016, pp.785–794. doi:10.1145/2939672.2939785
 52. Barredo Arrieta A, Diaz-Rodriguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 2020; 58: 82–115.

53. Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters. *Econom J* 2018; 21: 61–68.
54. Sharma A and Kiciman E. DoWhy: An end-to-end library for causal inference. *arXiv* 2020. <http://arxiv.org/abs/2011.04216>
55. Wager S and Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc* 2018; 113: 1228–1242.
56. Athey S and Imbens G. Recursive partitioning for heterogeneous causal effects. *Proc Natl Acad Sci USA* 2016; 113: 7353–7360.
57. Schwartzman CM and Boswell JF. A narrative review of alliance formation and outcome in text-based telepsychotherapy. *Pract Innov* 2020; 5: 128–142.
58. Chahar R, Dubey AK and Narang SK. Multiclass classification of mental health disorders using XGBoost-HOA algorithm. *SN Comput Sci* 2024; 5: 1167.
59. Bagherzadeh-Khiabani F, Ramezankhani A, Azizi F, et al. A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. *J Clin Epidemiol* 2016; 71: 76–85.
60. Agarwal R and Karahanna E. Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Q* 2000; 24: 665–694.
61. Kahn JH, Tobin RM, Massey AE, et al. Measuring emotional expression with the linguistic inquiry and word count. *Am J Psychol* 2007; 120: 263–286.
62. Slater MD. Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. *MEDIA Psychol* 2015; 18: 370–395.
63. Eysenbach G. The law of attrition. *J Med INTERNET Res* 2005; 7: e11.
64. Musiat P, Johnson C, Atkinson M, et al. Impact of guidance on intervention adherence in computerised interventions for mental health problems: A meta-analysis. *Psychol Med* 2022; 52: 229–240.
65. Althoff T, Clark K and Leskovec J. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Trans Assoc Comput Linguist* 2016; 4: 463–476.
66. Hall E. *Beyond Culture*. DoubleDay 1976.
67. Zhang L, Liu D, Li J, et al. Exploring linguistic features and user engagement in Chinese online mental health counseling. *Heliyon* 2024; 10: e38042.
68. Liu-Thompkins Y, Okazaki S and Li H. Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience. *J Acad Mark Sci* 2022; 50: 1198–1218.