FUTURITY OF SOCIAL SCIENCE

DOI: https://doi.org/10.57125/FS.2026.09.20.01

How to cite: Li, W., & Lian, J. (2026). A comparative study of attitudinal resources in human and AI-generated TED-style leadership talks. *Futurity of Social Sciences, 4*(3), 4–26. https://doi.org/10.57125/FS.2026.09.20.01

A Comparative Study of Attitudinal Resources in Human and AI-Generated TED-Style Leadership Talks

Wei Li*

PhD in English Language and Literature, Associate Professor, College of International Studies, Southwest University, Chongqing, China, https://orcid.org/0009-0003-8586-8270

Jiayi Lian

Master's Candidate of Education in English Language Teaching, College of International Studies, Southwest University, Chongqing, China

Minquan County Senior High School, Henan, China, https://orcid.org/0009-0005-8306-5858

***Corresponding author:** womansky@swu.edu.cn*

Received: April 8, 2026 | **Accepted:** June 21, 2026 | **Published:** July 15, 2026

Abstract: This study compares the use of attitudinal resources in human TED talks and AI-generated TED-style texts from the perspective of appraisal theory. A corpus of 20 human TED talks on leadership and 20 AI-generated texts was constructed and manually annotated with UAM

CorpusTool. The analysis focused on the distribution and polarity of affect, judgment, and appreciation resources. The findings show some observable differences between the two types of texts. Human texts draw more on affective resources, suggesting that human speakers in this corpus tend to construct evaluation through personal feelings, lived experience, and emotional involvement. In contrast, AI-generated texts show a more balanced distribution of the three resource types, with judgment taking a slightly higher proportion. In terms of polarity, the AI corpus contained a higher proportion of positive resources. In contrast, the human corpus contained a higher proportion of negative resources related to failure, doubt, and difficulty. This contrast was especially evident in judgment resources, in which negative judgment resources occurred less frequently in the AI corpus. Overall, the quantitative results and qualitative examples suggest that human texts in this corpus are more often grounded in personal experience and social relations. In contrast, AI-generated texts are more often characterised by generalised, instructive evaluative formulations.

Keywords: appraisal theory, attitude system, interpersonal function, TED talks, AI-generated text, human-machine comparison.

Introduction

With the rapid development of artificial intelligence, large language models such as ChatGPT and DeepSeek can now generate fluent, well-structured text. These AI-generated texts are increasingly used in writing assistance, content creation, and educational communication. Although they may appear similar to human writing at the surface level, it remains unclear whether they construct stance, attitude, and interpersonal meaning in the same way as human-authored texts. Public speaking, especially TED talks, provides a suitable context for exploring this issue because speakers not only present ideas but also draw on emotion, judgment, and value judgments to connect with audiences. Leadership-themed TED talks are particularly relevant, as they often involve personal experience, moral evaluation, and inspirational communication.

This study adopts the appraisal theory attitude system as its main analytical framework. Based on a corpus of TED talks on leadership, it compares the use of attitudinal resources in human-authored texts and AI-generated texts. Specifically, the study examines how the two types of texts differ in the distribution of affect, judgment, and appreciation, how these resources are used in terms of emotional strength, style, and evaluation targets, and what these differences reveal about the current limitations of AI in constructing interpersonal meaning. The AI corpus was generated by ChatGPT and DeepSeek using the same prompt conditions to reduce the influence of a single model and make the comparison more reliable.

This study has both theoretical and practical significance. Theoretically, it extends the application of appraisal theory to the comparison between human discourse and machine-generated discourse, offering a linguistic perspective on AI-generated texts. Practically, the findings may help AI developers better understand the limitations of large language models in emotional expression and interpersonal interaction. They may also help English teachers and learners

recognise that effective discourse depends not only on grammatical accuracy and logical organisation, but also on evaluative strategies, emotional involvement, and interpersonal connection.

Research Problem

With the rapid development of large language models such as ChatGPT and DeepSeek, AI-generated texts have become increasingly fluent, coherent, and human-like. However, language is not only used to convey information; it also expresses attitudes, constructs values, and builds interpersonal relationships. Therefore, an important question remains: although AI can imitate the surface features of human discourse, can it construct interpersonal meaning in the same way as human speakers?

Existing studies on appraisal theory have examined attitudinal resources in political speeches, news reports, academic discourse, and TED talks, but most of them focus on human-authored texts. Few studies have systematically compared human discourse and AI-generated discourse in terms of affect, judgment, and appreciation. This study, therefore, addresses this gap by exploring how AI-generated texts and human-authored TED talks differ in their use of attitudinal resources. It aims to provide corpus-based evidence of observable discourse patterns and discuss possible explanations consistent with the data.

Research Focus

This study focuses on the use of attitudinal resources in AI-generated and human-authored texts from an interpersonal perspective. Leadership-themed TED talks were selected as the corpus because they contain rich emotional expression, moral evaluation, personal experience, and value construction. These features make them suitable for examining how speakers or text producers use language to influence audiences and construct interpersonal meaning.

Based on the appraisal theory of attitude, this study compares the distribution and polarity of affect, judgment, and appreciation resources in human TED talks and AI-generated TED-style texts. It also examines qualitative differences in emotional strength, stylistic strategies, and evaluation targets to clarify whether AI-generated discourse tends to be more abstract, generalised, and formulaic than human discourse.

Research Aim and Research Questions

This study aims to compare how human TED talks and AI-generated TED-style texts use attitudinal resources to reveal differences in how they construct interpersonal meaning. Accordingly, this study addresses the following three research questions:

1. How do AI-generated texts and human-authored texts differ in the types and frequencies of attitudinal resources they use?
2. How are attitudinal resources used differently in AI-generated texts and human-authored texts in terms of emotional strength, writing style, and evaluation targets?

3. What do these differences suggest about the possible limitations of AI-generated texts in constructing interpersonal meaning?

Literature Review

Appraisal theory is rooted in systemic functional linguistics, which views language as a resource for meaning-making in social interaction. Halliday (2004) and Halliday and Matthiessen (2013) regard interpersonal meaning as one of the major metafunctions of language, while Almurashi (2016) offers a general introduction to Halliday's systemic functional linguistics. Based on this theoretical tradition, Martin and White (2005) developed appraisal theory to examine how speakers and writers express attitudes, negotiate stance, and construct value judgments. Recent studies by Cheng (2024) and Phan (2025) further show that research on interpersonal metafunction and appraisal theory has expanded across discourse analysis, multilingual comparison, and language education.

Within appraisal theory, the attitude system is central to the present study because it focuses on how evaluation is expressed in discourse. It consists of affect, judgment, and appreciation. Affect refers to emotional responses; judgment concerns the evaluation of people's behaviour, abilities, and moral qualities; and appreciation evaluates objects, processes, events, and ideas (Martin & White, 2005). This framework is useful for analysing leadership discourse because leadership talks often involve personal experience, emotional disclosure, moral evaluation, and value construction. Therefore, the attitude system provides a suitable theoretical basis for comparing how human speakers and AI-generated texts construct interpersonal meaning.

Previous studies have applied appraisal theory to various discourse types. Huo and Zhang (2023) analyse the interpersonal function of Joe Biden's inaugural speech, while Gong (2024) examines interpersonal meaning in English climate speeches. Zhang (2024) studies attitude resources in COP28 news reports, and Tang and Shen (2025) investigate attitude resources in Singaporean news reports on "a community with a shared future for mankind". Ming (2025) analyses attitudinal orientation in the translation of the 2025 New Year's address. These studies show that attitudinal resources are closely related to discourse topic, communicative purpose, and speaker-audience relationship. In addition, Duan and Degand (2024) investigate attitudinal resources in academic talks across languages, showing that spoken discourse also contains complex evaluative meanings rather than merely transmitting information.

TED talks have also become an important object of discourse research because they combine public speaking, knowledge sharing, storytelling, and persuasion. Xia and Hafner (2021) show that TED speakers engage online audiences through multimodal resources, while Dong and Zhang (2026) further demonstrate that TED speakers construct stance through both verbal and multimodal semiotic resources. Studies by Gao (2024), Mao (2025), and Wang (2025) also indicate that TED speeches are rich in interpersonal meaning and suitable for analysis from the perspective of appraisal theory and interpersonal metafunction. However, most existing TED studies focus on education, technology, ecology, health, or general speech features, while leadership-themed TED talks remain relatively underexplored.

Leadership-themed TED talks are particularly suitable for appraisal analysis because leadership discourse often involves self-presentation, emotional appeal, moral judgment, and social responsibility. Research on authentic leadership emphasises self-awareness, relational transparency, internalised moral perspective, and balanced processing as important dimensions of leadership construction (Walumbwa et al., 2008). These dimensions are closely connected with the use of first-person narration, affective expression, and judgment resources in leadership speeches. When speakers discuss failure, courage, uncertainty, responsibility, or trust, they are not only presenting leadership ideas but also constructing interpersonal relationships with the audience.

With the rapid development of large language models, recent research has begun comparing human- and AI-generated texts. Mindner et al. (2023) examine features for classifying human- and AI-generated texts, while Kumar et al. (2024) compare human and AI-generated texts from a computational perspective. Tengler and Brandhofer (2025) explore the quality differences between AI-generated and human-written texts. Georgiou (2025) finds that human-written and AI-generated texts differ in multiple automatically extracted linguistic features, and Zanotto and Aroyehun (2024) report that human texts show greater linguistic variability and richer emotional content than large language model outputs. These studies suggest that although AI-generated texts may appear fluent and coherent, they may still differ from human texts in emotional richness, linguistic variability, and interpersonal expression.

Some studies have further examined AI-generated discourse from linguistic and social perspectives. Zhang et al. (2024) analyse attitude resources in ChatGPT news discourse from the perspective of appraisal theory, while Križan and Barbič (2025) discuss appraisal analysis and AI chatbots. Flaih (2026) compares traditional and AI-generated literary texts, and Zhang et al. (2026) examine lexical features in ChatGPT-generated and human translations. Gillings et al. (2025) argue that large language models bring new challenges to critical discourse studies because AI-generated texts can construct stance, authority, and value in social communication. These studies indicate that AI-generated texts should be evaluated not only for fluency or accuracy but also for interpersonal meaning and evaluative positioning.

Overall, previous research has provided useful insights into appraisal theory, TED discourse, leadership communication, and human-AI textual comparison. Despite these contributions, there are still areas that require further investigation. First, most appraisal studies focus on human-authored texts, whereas AI-generated discourse has received less attention from an attitude-system perspective. Second, existing TED studies rarely examine leadership-themed speeches as a distinct persuasive genre characterised by personal experience, moral evaluation, and emotional appeal. Third, studies on AI-generated texts have mainly focused on detection, readability, lexical features, or overall quality. At the same time, fewer have systematically compared how human and AI-generated texts use affect, judgment, and appreciation. Therefore, this study uses the appraisal theory attitude system to compare human TED talks and AI-generated TED-style leadership texts and to examine their differences in the construction of interpersonal and evaluative meaning.

Materials and Methods

This study adopts a corpus-based comparative research design. Both quantitative and qualitative methods are used to examine the differences between human-authored TED talks and AI-generated TED-style texts in the use of attitudinal resources. The quantitative analysis focuses on the frequency, percentage, and polarity distribution of affect, judgment, and appreciation resources. The qualitative analysis further explains how these resources are used in relation to emotional strength, stylistic strategies, and evaluation targets.

The appraisal theory attitude system is used as the main analytical framework. In this study, affect refers to emotional expressions; judgment refers to evaluations of people's behaviour, abilities, and character; and appreciation refers to evaluations of things, ideas, processes, and phenomena. This framework enables comparison of how human speakers and AI-generated texts construct interpersonal meaning through evaluative language.

Corpus and Data Sources

The corpus consists of two parts: human TED talk transcripts and AI-generated TED-style texts. The human corpus comprises 20 TED Talks on leadership, selected from the official TED website. To ensure comparability, the selected talks are all original English monologic speeches, of similar length and with a shared thematic focus on leadership. Non-linguistic elements such as "laughter," "applause," and other stage directions were removed before annotation.

The AI corpus includes 20 TED-style texts generated by two large language models, ChatGPT and DeepSeek. For each human TED talk title, the same standardised prompt was used to generate a corresponding TED-style speech. To reduce single-model bias, 10 texts were selected from ChatGPT and 10 from DeepSeek. The final human corpus contains 25,796 words, and the AI corpus contains 25,694 words, which makes the two corpora comparable in size.

Instruments and Procedures

Two main research tools were used in this study. UAM CorpusTool was used to manually annotate attitudinal resources. Based on the appraisal theory attitude system, all texts were annotated sentence by sentence for affect, judgment, and appreciation. Each attitudinal resource was also coded as positive, negative, or neutral.

AntConc was used as a supplementary tool for lexical search and frequency checking. It helped identify high-frequency attitudinal words, examine the context of specific evaluative expressions, and check the use of emotional intensifiers and first-person pronouns. Before annotation, all texts were cleaned and divided into sentences. To improve annotation reliability, a subset of the corpus was re-annotated, yielding an inter-rater agreement of 94.2%.

Data Analysis

After annotation, the data were first analysed quantitatively. The frequency and percentage of affect, judgment, and appreciation resources were calculated separately for the human and AI corpora. Their polarity distributions were also compared to determine whether the two types of texts differed in positive, negative, and neutral evaluations.

The second stage involved qualitative analysis. Representative examples were selected to explain the main differences between human and AI texts. Particular attention was paid to emotional intensity, use of first-person perspective, stylistic features, and evaluation targets. On this basis, the qualitative analysis examined whether the selected examples differed in their degree of personal involvement, emotional intensity, stylistic realisation, and targets of evaluation. Through the combination of quantitative and qualitative analysis, this study aims to reveal how human and AI discourse differ in constructing interpersonal meaning.

Results

Based on the appraisal theory of attitude, this section presents a comparative analysis of attitudinal resources in 20 human TED talks on leadership and 20 AI-generated TED-style texts. The analysis first examines the overall distribution and polarity of attitudinal resources in the human texts and the AI texts, respectively. It then statistically compares the two corpora to determine whether the observed differences are significant. Finally, the results are further interpreted in terms of affect, judgment, and appreciation resources.

Analysis of Attitudinal Resources in Human TED Talks on Leadership

Through manual annotation and statistical analysis, a total of 862 attitudinal resources were identified in the 20 human TED talks on leadership. As shown in Table 1, affect resources accounted for the largest proportion, reaching 39.6%, followed by judgment resources at 34.5% and appreciation resources at 25.9%.

This distribution suggests that human speakers tend to rely heavily on emotional expression when discussing leadership. Rather than presenting leadership only as an abstract concept, they often construct evaluation through personal feelings, lived experience, and direct emotional involvement. Judgment resources also occupy a considerable proportion, indicating that leadership discourse frequently involves evaluations of people's behaviour, ability, and moral qualities. By contrast, appreciation resources appear less frequently, which suggests that human speakers are relatively less inclined to focus primarily on abstract concepts, strategies, or phenomena. Their evaluative language is more often directed toward people, actions, and personal experience.

Table 1

Overall Distribution of Attitudinal Resources in Human Texts

Attitude Category	Frequency	Percentage
Affect	341	39.6%
Judgment	298	34.5%

Appreciation	223	25.9%
Total	862	100%

The following examples illustrate the use of affect resources in the human texts:

(1) I was proud to be their daughter.

(Human corpus: How vulnerability makes you a better leader)

(2) I was so afraid of what people might think of me as a new mother and CEO.

(Human corpus: How vulnerability makes you a better leader)

(3) I'm starting to lose it.

(Human corpus: How to claim your leadership power)

In these examples, the speakers express emotions such as pride, fear, and frustration. These affective expressions are closely connected with specific personal experiences. Example (1) presents pride through a family relationship, while Example (2) expresses fear in relation to gender identity and professional role. Example (3) conveys a more intense emotional state through direct, informal wording. These examples indicate that human speakers often use emotional disclosure to reduce distance from the audience and create interpersonal connections.

In terms of polarity, Table 2 shows that positive attitudinal resources were the most frequent in human texts, accounting for 60.2% of the total. Negative resources accounted for 33.4%, while neutral resources accounted for 6.4%. This pattern is consistent with the persuasive and motivational nature of leadership speeches, which generally tend to promote positive values. However, the relatively high proportion of negative resources is also important. It shows that human speakers do not avoid discussing failure, fear, difficulty, or doubt. These negative evaluations help construct credibility and make the positive message more convincing.

Table 2

Polarity Distribution of Attitudes in Human Texts

Polarity	**Frequency**	**Percentage**
Positive	519	60.2%
Negative	288	33.4%
Neutral	55	6.4%
Total	862	100%

To further examine the internal structure of attitudinal resources in human texts, the polarity distributions of affect, judgment, and appreciation were calculated separately. As shown in Table 3, positive affect accounted for 57.1%, while negative affect accounted for 37.0%. This indicates that

human speakers express both positive emotions, such as pride and hope, and negative emotions, such as fear and frustration. In judgment resources, positive judgment accounted for 61.4%, while negative judgment accounted for 33.1%, suggesting that speakers both praise desirable qualities and criticise problematic actions or values. In appreciation resources, positive appreciation accounted for 63.2%, while negative appreciation accounted for 28.3%, indicating that evaluations of ideas and phenomena are mainly positive but still include critical assessment.

Table 3

Polarity Distribution of Three Attitudinal Resources in Human Texts

Attitude Category	Positive Frequency	Positive Percentage	Negative Frequency	Negative Percentage	Neutral Frequency	Neutral Percentage
Affect	195	57.1%	126	37.0%	20	5.9%
Judgment	183	61.4%	99	33.1%	16	5.5%
Appreciation	141	63.2%	63	28.3%	19	8.5%

Analysis of Attitudinal Resources in AI-Generated TED-Style Texts

A total of 858 attitudinal resources were identified in the 20 AI-generated TED-style texts. As shown in Table 4, judgment resources accounted for the highest proportion at 35.5%, followed by affect resources at 33.1% and appreciation resources at 31.4%.

Compared with human texts, AI texts show a more balanced distribution across the three types of attitudinal resources. The differences among judgment, affect, and appreciation are relatively small. This suggests that AI-generated texts tend to distribute evaluative resources more evenly, rather than relying heavily on a single dominant resource type. The slightly higher proportion of judgment resources indicates that AI texts tend to evaluate behaviour, ability, and personal qualities more directly. In comparison, the lower proportion of affect resources suggests relatively weaker emotional involvement.

Table 4

Overall Distribution of Attitudinal Resources in AI Texts

Attitude Category	Frequency	Percentage
Affect	284	33.1%
Judgment	305	35.5%
Appreciation	269	31.4%
Total	858	100%

The following examples illustrate typical evaluative expressions in the AI-generated texts:

(4) Leadership is not a title. Leadership is communication.

(AI-ChatGPT corpus: What's your leadership language?)

(5) Great leaders don't rely on just one language. They learn to speak several languages.

(AI-DeepSeek corpus: The hot shot rule)

(6) Leadership power is not something you wait for. It is something you claim.

(AI-ChatGPT corpus: How to claim your leadership power)

These examples show that AI-generated texts often evaluate abstract concepts such as leadership, communication, and power. The expressions are coherent and motivational, but they are less closely connected with specific personal experience. Unlike the human examples, which are grounded in lived experience and emotional disclosure, the AI examples tend to present generalised statements. This suggests that AI-generated discourse is more concept-oriented and explanatory in its evaluative style.

The polarity distribution of AI-generated texts is shown in Table 5. Positive resources accounted for 68.5%, negative resources for 25.2%, and neutral resources for 6.3%. Compared with human texts, AI texts contain a higher proportion of positive evaluations and a lower proportion of negative evaluations. In this corpus, AI-generated texts showed a more encouraging and affirmative evaluative pattern. Although negative evaluation occurred in both corpora, negative resources were less frequent in the AI corpus than in the human corpus.

Table 5

Polarity Distribution of Attitudes in AI Texts

Polarity	**Frequency**	**Percentage**
Positive	588	68.5%
Negative	216	25.2%
Neutral	54	6.3%
Total	858	100%

Table 6 presents the polarity distribution of affect, judgment, and appreciation in the AI texts. Positive affect accounted for 64.3%, while negative affect accounted for 31.7%. This shows that AI-generated texts can express emotion, but their emotional orientation is more positive than that of human texts. In judgment resources, positive judgment accounted for 71.8%, whereas negative judgment accounted for only 23.5%. This is the most notable pattern in the AI texts, suggesting that AI tends to praise actions, abilities, and qualities rather than criticise them. In appreciation resources, positive appreciation accounted for 69.1%, while negative appreciation accounted for 20.1%, indicating that AI evaluations of ideas and concepts are also predominantly affirmative.

Table 6

Polarity Distribution of Three Attitudinal Resources in AI Texts

Attitude Category	Positive Frequency	Positive Percentage	Negative Frequency	Negative Percentage	Neutral Frequency	Neutral Percentage
Affect	183	64.3%	90	31.7%	11	4.0%
Judgment	219	71.8%	72	23.5%	14	4.7%
Appreciation	186	69.1%	54	20.1%	29	10.8%

Comparative Analysis of Human and AI Texts

Table 7 presents a crosstabulation of attitude resource types by text source. Human texts contain more affect resources than AI texts, while AI texts contain more judgment and appreciation resources. Specifically, affect resources appear more frequently in human texts, accounting for 54.6% of all affect resources across the two corpora. By contrast, appreciation resources appear more frequently in AI texts, accounting for 54.7% of all appreciation resources. Judgment resources are relatively balanced, though AI texts show a slightly higher count.

This comparison suggests that human speakers are more likely to construct interpersonal meaning through emotion and personal involvement, while AI-generated texts are more likely to construct evaluation through judgments of behaviour and appreciation of concepts. However, the differences are not extremely large, which indicates that AI texts can imitate the general structure of human evaluative discourse to some extent.

Table 7

Crosstabulation of Attitude Resource Types by Text Source

Attitude Resource Type	Measure	Human Texts	AI Texts	Total
Affect	Count	341	284	625
	Expected Count	313.2	311.8	625.0
	% within Attitude Resource Type	54.6%	45.4%	100.0%
	Standardized Residual	1.57	-1.57	
Judgment	Count	298	305	603
	Expected Count	302.2	300.8	603.0
	% within Attitude Resource Type	49.4%	50.6%	100.0%
	Standardized Residual	-0.24	0.24	
Appreciation	Count	223	269	492
	Expected Count	246.6	245.4	492.0
	% within Attitude Resource Type	45.3%	54.7%	100.0%
	Standardized Residual	-1.50	1.51	
Total	Count	862	858	1720
	Expected Count	862.0	858.0	1720.0
	% within Attitude Resource Type	50.1%	49.9%	100.0%

To determine whether the observed differences in attitude resource distribution were statistically significant, a chi-square test was conducted. As shown in Table 8, the Pearson chi-square value was 9.60, with a significance level of 0.008. Since the p-value is below 0.05, the difference in the distribution of attitude resource types between human and AI texts is statistically significant.

Table 8

Chi-Square Test Results for Attitude Resource Types

Test Item	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9.60	2	0.008
Likelihood Ratio	9.63	2	0.008
Linear-by-Linear Association	0.005	1	0.941
N of Valid Cases	1720		

Note: All expected cell counts are greater than 5. The minimum expected count is 245.46.

However, statistical significance does not necessarily indicate a strong association. Therefore, measures of association were also examined. As shown in Table 9, Cramér's V was 0.075. This indicates that although the distributional difference is statistically significant, the strength of association is small. In other words, the two types of texts differ in their use of attitudinal resources, but the difference is better understood as a tendency rather than a strong categorical divide.

Table 9

Symmetric Measures of Association for Attitude Resource Types

Measure Type	Value	Approximate Significance
Cramér's V (Nominal by Nominal)	0.075	0.008
N of Valid Cases	1720	

Table 10 compares the polarity distribution of attitudinal resources in human and AI texts. The results show a statistically significant difference in distribution, although the effect size is small. AI texts contain more positive evaluations, while human texts contain more negative evaluations. Among all positive resources, 53.1% appear in AI texts and 46.9% in human texts. Among all negative resources, 57.1% appear in human texts and 42.9% in AI texts. This indicates that AI-generated texts are more strongly oriented toward positive evaluation, whereas human texts are more willing to include negative evaluation.

Table 10

Crosstabulation of Attitude Polarity by Text Source

Attitude Polarity	Measure	Human Texts	AI Texts	Total

Positive	Count	519	588	1107
	Expected Count	554.7	552.3	1107.0
	% within Attitude Polarity	46.9%	53.1%	100.0%
	Standardized Residual	-1.91	1.91	
Negative	Count	288	216	504
	Expected Count	252.6	251.4	504.0
	% within Attitude Polarity	57.1%	42.9%	100.0%
	Standardized Residual	2.30	-2.30	
Neutral	Count	55	54	109
	Expected Count	54.6	54.4	109.0
	% within Attitude Polarity	50.5%	49.5%	100.0%
	Standardized Residual	0.08	-0.08	
Total	Count	862	858	1720
	Expected Count	862.0	858.0	1720.0
	% within Attitude Polarity	50.1%	49.9%	100.0%

The chi-square test results in Table 11 show that the difference in polarity distribution is statistically significant. The Pearson chi-square value was 14.560, with a significance level of 0.001. This confirms that human and AI texts differ significantly in their use of positive, negative, and neutral evaluations.

Table 11

Chi-Square Test Results for Attitude Polarity

Test Item	**Value**	**df**	**Asymptotic Significance (2-sided)**
Pearson Chi-Square	14.560	2	0.001
Likelihood Ratio	14.610	2	0.001
Linear-by-Linear Association	5.210	1	0.022
N of Valid Cases	1720		

Note: All expected cell counts are greater than 5. The minimum expected count is 54.37.

As shown in Table 12, Cramér's V was 0.092. This indicates that the association between text source and attitude polarity is statistically significant but relatively weak. Therefore, the polarity difference should be interpreted as a consistent tendency: AI texts tend to be more positive, while human texts contain more negative evaluation.

Table 12

Symmetric Measures of Association for Attitude Polarity

Measure Type	Value	Approximate Significance
Cramér's V (Nominal by Nominal)	0.092	0.001
N of Valid Cases	1720	

The following examples illustrate the polarity difference between the two corpora:

(7) That's the story of how I lost my mind. The end.

(Human corpus: How to claim your leadership power)

(8) The only thing that makes you a leader is your ability to unlock the potential in others.

(AI-DeepSeek corpus: The one question every aspiring leader needs to ask)

Example (7) contains a strongly negative self-evaluation. The phrase "lost my mind" expresses frustration and vulnerability in an exaggerated but rhetorically effective way. It helps the speaker construct authenticity and emotional closeness with the audience. By contrast, Example (8) presents a positive and instructive statement about leadership. It is clear and motivational, but it avoids difficulty, failure, or emotional conflict. This contrast shows that negative evaluation in human texts often functions as a means of constructing credibility, while AI-generated texts tend to avoid such expressions and maintain a more affirmative tone.

To further determine whether the polarity differences were consistent across all attitudinal resource types, affect, judgment, and appreciation were tested separately.

Table 13 shows the polarity distribution of affect resources. Human texts contain more negative affect resources than AI texts, while AI texts contain slightly more positive affect resources than expected. However, the difference is not very large.

Table 13

Affect Polarity by Text Source

Affect Polarity	Measure	Human Texts	AI Texts	Total
Positive	Count	195	183	378
	Expected Count	206.2	171.8	378.0
	% within Affect Polarity	51.6%	48.4%	100.0%
	Standardized Residual	-0.78	0.85	
Negative	Count	126	90	216
	Expected Count	117.8	98.2	216.0
	% within Affect Polarity	58.3%	41.7%	100.0%
	Standardized Residual	0.76	-0.83	
Neutral	Count	20	11	31
	Expected Count	16.9	14.1	31.0

	% within Affect Polarity	64.5%	35.5%	100.0%
	Standardized Residual	0.75	-0.82	
Total	Count	341	284	625
	Expected Count	341.0	284.0	625.0

The chi-square test in Table 14 shows that the difference in affect polarity is not statistically significant, with a Pearson chi-square value of 3.83 and a p-value of 0.147. This suggests that although human texts contain somewhat more negative affect, the difference is not strong enough to be considered statistically meaningful.

Table 14

Chi-Square Test for Affect Polarity

Test Item	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	3.83	2	0.147
N of Valid Cases	625		

Note: All expected cell counts are greater than 5. The minimum expected count is 14.09.

Judgment resources show a clearer pattern. As shown in Table 15, AI texts contain more positive judgment resources, while human texts contain more negative judgment resources. Among positive judgment resources, 54.5% appear in AI texts and 45.5% in human texts. Among negative judgment resources, 57.9% appear in human texts and 42.1% in AI texts. This suggests that, in this corpus, positive judgment resources occurred more frequently in the AI-generated texts, whereas negative judgment resources occurred more frequently in the human texts.

Table 15

Judgment Polarity by Text Source

Judgment Polarity	**Measure**	**Human Texts**	**AI Texts**	**Total**
Positive	Count	183	219	402
	Expected Count	198.7	203.3	402.0
	% within Judgment Polarity	45.5%	54.5%	100.0%
	Standardized Residual	-1.11	1.10	
Negative	Count	99	72	171
	Expected Count	84.5	86.5	171.0
	% within Judgment Polarity	57.9%	42.1%	100.0%
	Standardized Residual	1.58	-1.56	
Neutral	Count	16	14	30
	Expected Count	14.8	15.2	30.0
	% within Judgment Polarity	53.3%	46.7%	100.0%
	Standardized Residual	0.31	-0.31	

Total	Count	298	305	603
	Expected Count	298.0	305.0	603.0

The chi-square test in Table 16 confirms that the difference in judgment polarity is statistically significant. The Pearson chi-square value was 7.57, with a significance level of 0.023. This indicates that judgment resources are the main source of polarity difference between human and AI texts.

Table 16

Chi-Square Test for Judgment Polarity

Test Item	**Value**	**df**	**Asymptotic Significance (2-sided)**
Pearson Chi-Square	7.57	2	0.023
N of Valid Cases	603		

Note: All expected cell counts are greater than 5. The minimum expected count is 14.82.

The final category is appreciation, which concerns the evaluation of objects, ideas, processes, and phenomena. As shown in Table 17, AI texts contain more positive appreciation resources, while human texts contain more negative appreciation resources. However, the difference is smaller than that found in judgment resources.

Table 17

Appreciation Polarity by Text Source

Appreciation Polarity	**Measure**	**Human Texts**	**AI Texts**	**Total**
Positive	Count	141	186	327
	Expected Count	148.2	178.8	327.0
	% within Appreciation Polarity	43.1%	56.9%	100.0%
	Standardized Residual	-0.59	0.54	
Negative	Count	63	54	117
	Expected Count	53.0	64.0	117.0
	% within Appreciation Polarity	53.8%	46.2%	100.0%
	Standardized Residual	1.37	-1.25	
Neutral	Count	19	29	48
	Expected Count	21.8	26.2	48.0
	% within Appreciation Polarity	39.6%	60.4%	100.0%
	Standardized Residual	-0.59	0.54	
Total	Count	223	269	492
	Expected Count	223.0	269.0	492.0

As shown in Table 18, the chi-square test for appreciation polarity was not statistically significant. The Pearson chi-square value was 4.72, with a p-value of 0.094. Although the result approaches significance, it does not meet the conventional threshold of 0.05. Therefore, appreciation polarity does not constitute a major difference between the two corpora.

Table 18

Chi-Square Test for Appreciation Polarity

Test Item	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	4.72	2	0.094
N of Valid Cases	492		

Note: All expected cell counts are greater than 5. The minimum expected count is 21.76.

Taken together, the results show that human and AI texts differ significantly in the overall distribution of attitudinal resources and in the polarity of evaluation. However, among the resource-specific comparisons, judgment polarity was the only category that showed a statistically significant difference under the unadjusted criterion. AI-generated texts show a clear tendency toward positive judgment and avoid negative moral or behavioral evaluation more than human texts. By contrast, human texts are more willing to include negative judgments, especially when discussing failure, difficulty, personal weakness, or problematic social values.

The following examples illustrate this difference in judgment resources:

(9) My parents worked two full-time jobs without ever complaining.

(Human corpus: How vulnerability makes you a better leader)

(10) She demonstrated exceptional courage and resilience throughout the crisis.

(AI-ChatGPT corpus: The hot shot rule)

(11) I was so constricted by my belief that businesses value maleness more.

(Human corpus: How vulnerability makes you a better leader)

Example (9) presents a positive judgment embedded in a specific personal narrative. The evaluation of the parents' perseverance is connected with concrete actions and family experience. Example (10), although also positive, is more generalized and less contextually grounded. The praise of "exceptional courage and resilience" could apply to many situations and does not depend on a specific interpersonal relationship. Example (11) illustrates negative self-evaluation and social critique, showing how human speakers use judgment resources to express vulnerability and reflect on social values. These examples further support the quantitative finding that human texts tend to construct judgment through lived experience and interpersonal context, while AI texts tend to produce more generalized and positive evaluations.

Discussion

Based on the quantitative analysis presented in the previous section, this section further explores the possible explanations for the differences between human and AI texts in interpersonal meaning construction, and discusses what these differences may reveal about the current limitations of AI-generated discourse.

The first main difference is emotional intensity. Human speakers are more willing to express strong emotions, such as anger, fear, pride, and frustration, and these expressions are often direct and relatively unfiltered. As shown in (3) and (7), human speakers use phrases like "I'm starting to lose it" and "that's the story of how I lost my mind" to show frustration and create a sense of shared experience with the audience. Example (2) also shows a deep fear related to gender identity in the workplace. These expressions not only describe feelings, but also show a kind of personal openness.

This may be related to the fact that human communication is based on real-life experience. When speakers talk about their own difficulties, the emotions are connected to real events, which makes them more convincing. The purpose is not simply to share personal stories, but to build a connection with the audience. By showing vulnerability, speakers create a shared emotional space that helps build trust. In addition, from a qualitative perspective, human texts tend to use more explicit intensifying language to show emotional intensity, such as words like "deeply" and "truly," which further strengthen the sense of emotional involvement. AI texts show a different tendency. Examples (4), (5), (6), and (8) are phrased in a relatively measured, abstract, and instructive register. Expressions such as "Leadership is not a title" are coherent and motivational, but they seem to create a lower degree of emotional involvement than the human examples. One possible explanation is that AI-generated texts tend to reproduce common rhetorical patterns found in public-speaking or motivational discourse, where clarity, balance, and positivity are often foregrounded. Another possible explanation is that the prompt-based generation process encourages the production of broadly acceptable and audience-friendly statements.

The second difference concerns tone and self-positioning. Human speakers frequently place themselves within the discourse. Examples (1), (2), (3), (7), and (9) all feature first-person narration, such as "I was proud to be their daughter" and "I was so afraid." These are not only stylistic choices, but also interpersonal strategies. The first-person perspective allows the speaker to present leadership as something experienced, struggled with, and personally negotiated. By contrast, AI-generated texts in this corpus often use generalized statements or only formulaic first-person openings, such as "Let me start with...". This pattern suggests that AI-generated discourse may simulate the structure of public speaking without necessarily reproducing the same depth of personal self-disclosure.

The third difference concerns what gets evaluated. Human speakers often evaluate specific people, relationships, and socially situated experiences. Examples (1) and (9) are related to the speaker's parents, while examples (2), (3), and (11) focus on the speaker's own fears, beliefs, and social constraints. These evaluations are embedded in interpersonal contexts. By contrast, AI-

generated texts in examples (4), (5), (6), (8), and (10) tend to evaluate abstract concepts such as leadership, power, strategy, and innovation. This contrast is especially visible in judgment resources. Human judgment in examples (9) and (11) is anchored in narrative and personal experience, whereas AI judgment in example (10) is more generalized and less dependent on a specific social relationship. The data therefore suggest that AI-generated texts in this corpus tend to construct evaluation at a higher level of abstraction, while human texts tend to connect evaluation with concrete social experience.

Taken together, these differences suggest that human discourse in this corpus is more strongly grounded in lived experience, interpersonal relations, and emotionally charged self-positioning. AI-generated discourse, by contrast, appears more strongly oriented toward conceptual clarity, generalized evaluation, and a broadly positive tone. These findings do not prove that AI lacks emotional understanding or interpersonal capacity in any absolute sense. Rather, they show that, under the conditions of this study, AI-generated TED-style leadership texts display a different pattern of attitudinal resource use from human TED talks. This pattern is consistent with the present interpretation that current AI-generated public-speaking texts may imitate the surface organization of human discourse more successfully than the experiential and relational grounding of human evaluation.

Conclusions and Implications

The comparison between AI-generated texts and human-authored texts shows that the two types of discourse differ in the way they construct interpersonal meaning through attitudinal resources. Human texts in this corpus tend to use more affect resources, which suggests that human speakers are more likely to express evaluation through personal emotions, lived experience, and direct involvement with the audience. By contrast, AI-generated texts show a more balanced distribution of affect, judgment, and appreciation, with judgment resources taking a slightly higher proportion. This suggests that AI-generated discourse in this study tends to organize evaluation in a relatively stable, explanatory, and concept-oriented way.

The findings also show clear differences in polarity. AI-generated texts tend to be more positive, while human texts in this study contain more negative evaluations, especially when speakers discuss failure, uncertainty, difficulty, or moral problems. This difference is particularly evident in judgment resources. Human speakers are more willing to make negative evaluations of actions, choices, or social realities, whereas AI-generated texts in this corpus are less likely to include strong negative moral judgment. Qualitatively, the selected human examples appeared more frequently grounded in first-person narration, emotionally charged experience, and socially situated evaluation. AI-generated texts, however, are more likely to evaluate abstract concepts such as leadership, success, growth, and responsibility, and their evaluative style is generally more generalized, measured, and instructive.

These results have both theoretical and practical implications. Theoretically, this study extends the application of appraisal theory to the comparison between human discourse and machine-generated discourse. It shows that the attitude system can be used not only to analyze traditional

human-authored texts, but also to examine how AI-generated texts construct stance, value, and interpersonal meaning. This provides a useful linguistic perspective for understanding the discourse features of large language models.

Practically, the findings may be useful for English teaching, AI-assisted writing, and public speaking training. For English learners, the comparison helps them understand that effective discourse does not only depend on grammatical correctness or logical organization, but also on emotional involvement, evaluative precision, and interpersonal connection. For teachers, the findings can be used to guide students in identifying possible limitations of AI-generated writing and in developing more context-sensitive evaluative expression. For AI developers, the study suggests that future language models may need to produce more varied and context-sensitive evaluative language.

Limitations and Suggestions for Future Research

Several limitations should be acknowledged. First, the corpus is relatively small, consisting of only 20 human TED talks and 20 AI-generated TED-style texts. Although the two sub-corpora are comparable in word count, the findings should be interpreted as tendencies within this dataset rather than as general conclusions about all human and AI-generated discourse. Second, this study focuses only on leadership-themed TED-style talks. Therefore, the findings may not be directly applicable to other discourse types, such as political speeches, news reports, academic writing, classroom interaction, or social media communication. Third, the AI corpus was generated by ChatGPT and DeepSeek. Although this design helps reduce single-model bias, the findings cannot represent all large language models or all prompting conditions.

In addition, the annotation was completed by two researchers. Although re-annotation was conducted and the inter-rater agreement rate was relatively high, manual annotation of appraisal resources still involves a degree of interpretive judgment. Finally, this study is based on corpus evidence and cannot directly verify the internal mechanisms of large language models. Therefore, explanations related to training data, safety constraints, prompt effects, or model design should be understood as possible hypotheses that are consistent with the observed textual patterns, rather than as mechanisms established by the corpus itself. Future research could expand the corpus, include more genres and AI models, involve more annotators, and combine corpus-based discourse analysis with experimental or prompt-controlled methods to better understand how AI-generated texts construct interpersonal meaning.

Acknowledgements

None.

Conflict of Interest

None.

Funding

This work was supported by Education and Teaching Reform Research Project of Southwest University "A Study on Pathways for Improving the Quality of English Education through Digital Empowerment" (Grant No. 2024JY076); Graduate Education and Teaching Reform Research Project of Chongqing Municipality: "Research on the AI-Based Paradigm Transformation of Graduate English Education" (Grant No. yjg250048).

References

Almurashi, W. A. (2016). An introduction to Halliday's systemic functional linguistics. *Journal for the Study of English Linguistics, 4*(1), 70–80.

Cheng, S. (2024). A review of interpersonal metafunction studies in systemic functional linguistics (2012–2022). *Journal of World Languages, 10*(3), 623–667.

Dong, J., & Zhang, M. (2026). Stance beyond words: How TED speakers construct stance through multimodal semiotic resources. *English for Specific Purposes, 81*, 150–168. https://doi.org/10.1016/j.esp.2025.10.001

Duan, Y., & Degand, L. (2024). Attitudinal resources in academic talks: A corpus-based analysis across languages. *Corpus Pragmatics, 8*(4), 335–358. https://doi.org/10.1007/s41701-024-00171-4

Flaih, M. A. A. K. N. (2026). Human vs. machine: A comparative study of traditional and AI-generated literary texts. *Tasnim International Journal for Human, Social and Legal Sciences, 5*(1), 152–163. https://doi.org/10.56924/tasnim.s1.2026/10

Gao, X. (2024). Positive discourse analysis of TED artificial intelligence education speech under attitude system. *Journal of Higher Education Research, 5*(5), 442–444. https://doi.org/10.32629/jher.v5i5.3057

Georgiou, G. P. (2025). Differentiating between human-written and AI-generated texts using automatically extracted linguistic features. *Information, 16*(11), Article 979. https://doi.org/10.3390/info16110979

Gillings, M., Kohn, T., & Mautner, G. (2025). The rise of large language models: Challenges for critical discourse studies. *Critical Discourse Studies, 22*(6), 625–641. https://doi.org/10.1080/17405904.2024.2373733

Gong, X. (2024). An interpersonal function analysis of English climate speeches. *Modern English*, (18), 96–98.

Halliday, M. A. K. (2004). *An introduction to functional grammar* (3rd ed.). Hodder Education.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2013). *Halliday's introduction to functional grammar* (4th ed.). Routledge.

Huo, H., & Zhang, Y. (2023). An analysis of the interpersonal function of Joe Biden's inaugural speech. *International Journal of Languages, Literature and Linguistics, 9*(3), 204–207.

Križan, A., & Barbič, A. (2025). Appraisal analysis and AI chatbots: Do we even need humans? ELOPE: English Language Overseas Perspectives and Enquiries, 22(1), 35–52. https://doi.org/10.4312/elope.22.1.35-52

Kumar, S., Tiwari, S., Prasad, R., Rana, A., & Arti, M. K. (2024). Comparative analysis of human and AI generated text. In *2024 11th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 168–173). IEEE. https://doi.org/10.1109/SPIN60856.2024.10511301

Mao, Y. (2025). An analysis of a TED speech from the perspective of interpersonal meta-function. International Journal of Linguistics, Literature and Translation, 8(3), 16–21. https://doi.org/10.32996/ijllt.2025.8.3.4

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation*: *Appraisal in English*. Palgrave Macmillan.

Mindner, L., Schlippe, T., & Schaaff, K. (2023). Classification of human- and AI-generated texts: Investigating features for ChatGPT. In *International Conference on Artificial Intelligence in Education Technology* (pp. 152–170). Springer Nature Singapore.

Ming, Z. (2025). Attitudinal orientation in the translation of the 2025 New Year address from the perspective of appraisal theory: A corpus-based sentiment analysis of the attitude system. *Language and Culture Studies, 33*(3), 199–202.

Phan, L. T. (2025). Tracing the development of research on appraisal theory within systemic functional linguistics (2003–2025): A bibliometric analysis. *Language Teaching Research Quarterly, 52*, 87–109. https://doi.org/10.32038/ltrq.2025.52.05

Tang, Z., & Shen, W. (2025). Attitude resources of "a community with a shared future for mankind" in Singaporean news reports: A two-level sentiment analysis based on appraisal theory. *Journalism & Communication Research, 32*(12), 103–120, 123.

Tengler, K., & Brandhofer, G. (2025). Exploring the difference and quality of AI-generated versus human-written texts. *Discover Education, 4*(1), 113. https://doi.org/10.1007/s44217-025-00529-z

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. (2008). Authentic leadership: Development and validation of a theory-based measure. *Journal of Management, 34*(1), 89–126. https://doi.org/10.1177/0149206307308913

Wang, J. (2025). An AntConc-based study of linguistic features in TED speeches from the perspective of interpersonal function. *Journal of Zunyi Normal University, 27*(5), 73–77.

Xia, S. A., & Hafner, C. A. (2021). Engaging the online audience in the digital era: A multimodal analysis of engagement strategies in TED talk videos. *Ibérica, 42*, 33–58. https://doi.org/10.17398/2340-2784.42.33

Zanotto, S. E., & Aroyehun, S. (2024). Human variability vs. machine consistency: A linguistic analysis of texts generated by humans and large language models. *arXiv*. https://arxiv.org/abs/2412.03025

Zhang, H., Zhang, L., Nian, H., & Zhang, L. (2024). Positive discourse analysis of attitude resources in Chat GPT news discourse from the perspective of appraisal theory. *Lecture Notes on Language and Literature, 7*(3), 61–66.

Zhang, N., Liu, X., & Li, D. (2026). A comparative study of lexical features in ChatGPT translation and human translation: A case study of the literary work *Big Breasts and Wide Hips*. *Foreign Language Research, 43*(2), 18–25, 113.

Zhang, Y. (2024). Discourse analysis of COP28 news reports from the perspective of attitude in appraisal theory. *Studies in Linguistics and Literature, 8*(2), 143-152. https://doi.org/10.22158/sll.v8n2p143