ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 12, No. 7, pp. 1337-1348, July 2022
DOI: https://doi.org/10.17507/tpls.1207.13

Epistemic Modality in English-Medium Artificial Intelligence Research Articles: A Systemic–Functional Perspective

Gefei Bo
School of Foreign Languages, Dalian University of Technology, Dalian, China

Jianjun Ma
School of Foreign Languages, Dalian University of Technology, Dalian, China

***Abstract*—Epistemic modality is a crucial tool for describing the speaker's judgment—or request for the judgment of the listener—that has gained broad recognition and been widely investigated in academic writing. This research investigates the use of epistemic modality in 50 English-medium artificial intelligence (AI) research articles (RAs) from a systemic–functional perspective. Our research focuses on the frequency and function of the value, orientation, and polarity of epistemic modality in AI RAs. The results reveal that these AI RA writers tend to use both high- and low-value modality with explicitly objective orientation and positive polarity and that the epistemic modality performs the interpersonal function. This indicates that AI RA writers tend to ensure the objectivity of their results and that these writers adopt varying attitudes and hold differing views toward AI technology and thus have not yet achieved consensus. The study furthers the understanding of scholars' attitudes toward AI statements in recent years.**

***Index Terms*—epistemic modality, value, orientation, polarity, systemic–functional grammar**

I. Introduction

Epistemic modality relates to the modification of statements in terms of expressing confidence, truthfulness, and probability. Simply expressed, it relates to the speaker's level of certainty that what they are stating is correct. Simpson (1993) states that modality is one of the most important means of expressing interpersonal function. The importance of modality has gained wide recognition, and it has been widely investigated in the discourse of many fields, such as politics (Vukovic, & Milica, 2014), legislation (Cheng, 2014), academic discourse (Gao, 2012; Poole et al., 2019; Li, 2019), diplomatic writing (Yang et al., 2015), and translation/interpretation (Fu, 2016). Vukovic and Milica (2014) found that in parliamentary discourse, the most frequent mode of expressing modality is to use a strong epistemic adverb; the runners-up are strong epistemic verbs. Cheng's (2014) findings suggest that both subjective and objective epistemic modality are employed in court judgments. Epistemic modality has also been discussed from the perspective of speaker characteristics. Shakirova (2019) identified gender and age as the key characteristics for determining the linguistic impacts on the choice of markers; the speech of people from various genders and age groups differs. Recent research has also focused on epistemic modality in academic discourse. Gao (2012) explored the use of epistemic modality in academic English writing, while Poole et al. (2019) investigated epistemic modality in scientific and technological writing, such as biochemical research articles (RAs). However, to date, little research has directly investigated epistemic modality in artificial intelligence (AI) RAs.

Most studies of epistemic modality focus on two aspects: frequency and function. Previous research comparing the frequency of epistemic modality from a cross-disciplinary perspective has found that literary critics favor the combined use of deontic and epistemic modality, whereas medical researchers, biologists, and health scientists restrict the use of deontic modality and favor the use of epistemic modality (Piqué-Angordans et al., 2001, 2002). Conversely, Vazquez and Giner (2008) conclude that the "soft" sciences resort more to epistemic modality than the "hard" sciences. Some studies have researched epistemic modality from the perspective of value; Li (2019) found significant differences in the use of low-value modality between humanities academic discourses and natural science academic discourses, and median-value modality is more frequently used in social sciences. Other research about epistemic modality focuses on move; Salager-Meyer (1992) found that in medical RAs, the most heavily hedged moves (use of modality) were the conclusion recommendations, and data synthesis. Gao (2012) analyzed the three major interpersonal functions of epistemic modality: politeness, negotiation, and being constructive. Yang et al. (2015) contend that "the functions of epistemic modality are twofold. One function is propositional or semantic; the use of epistemic modality indicates the degree of certainty of the proposition and the addresser's confidence in the truth of the proposition. The other is interpersonal or pragmatic; by adopting the politeness strategy through the use of epistemic modality, addressers can establish relationships and successfully communicate with addressees" (p. 1-2). Although numerous studies have

attempted to investigate epistemic modality, there are relatively few historical corpus-based studies of it from the perspective of Halliday's systemic–functional grammar (SFG). In SFG, epistemic modal expression is considered to have two variables: "value" and "orientation" (Halliday, 2008). "Value indicates the degree of certainty and the addresser's modal commitment; orientation shows the linguistic forms of expressing modality and the addresser's modal responsibility" (Yang et al., 2015, p.3).

Defined as the ability of technology to model intelligent behavior with minimal human intervention, AI is currently an extremely controversial topic and has gradually become an extremely important field; engineering technology in the AI field is widely used in automation control, computer networking, electronic technology, information engineering, and other fields. In recent years, myriad studies have demonstrated its efficacy in many fields. Many of these studies focus on the ethics of AI in medical procedures (Keskinbora, 2019), news communication and media (Lei,2020), education (Zhao et al., 2020), legislation (Song et al., 2020), and politics (Kane, 2019). However, few studies have examined scholars' attitudes toward AI in RAs. This study discusses the use of epistemic modality in English-medium AI RAs from the perspective of epistemic modality.

The primary objective of this study is to investigate the use of epistemic modality in English-medium AI RAs from a systemic–functional perspective, focusing on the distribution of the value, orientation, and polarity of epistemic modality and their functions in RAs. Based on a corpus of 50 AI RAs, a small-scale exploratory study was conducted to examine how the authors of published AI RAs employ epistemic modality. The present study addresses the following two questions: 1. What are the frequency and distribution of the various values, orientations, and polarities of epistemic modality used in AI RAs? 2. What are the possible functions of the values, orientations, and polarities used in AI RAs? The findings of the study may help non-native AI RA writers to produce more acceptable AI RAs and help identify scholars' attitudes toward AI in recent years.

II. Theoretical Framework

"Modality refers to the area of meaning that lies between yes and no—the intermediate ground between positive and negative polarity. What this implies more specifically depends on the underlying speech function of the clause. If the clause is an 'information' clause, this type is referred to as modalization; if the clause is a 'goods & services' clause, this type is referred to as modulation" (Halliday, 2008, p.618). The system is outlined in Figure 1, below. In philosophical semantics, probability is referred to as "epistemic" modality, and obligation as "deontic" modality.

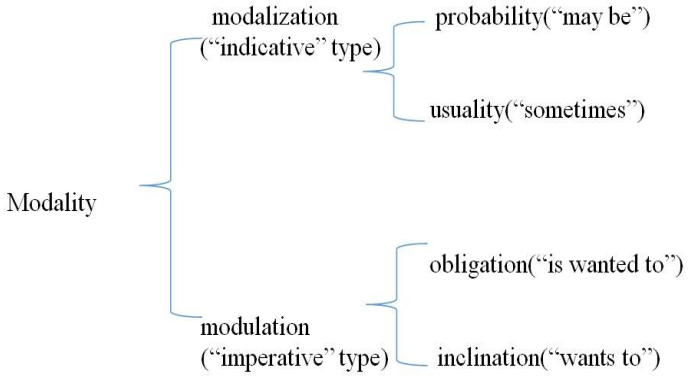

Figure 1 System of Modality Types (Halliday, 2008, p.618)

A. Value

The first variable in modality is value, which indicates the degree of certainty and the addresser's modal commitment. It is attached to the modal judgment: low, median, and high. These values are summarized in Table 1.

Table 1
Three Values Of Modality (Halliday, 2008, p.619)

Value	Examples
High	certain, always, required, determined
Median	probable, usually, supposed, keen
Low	possible, sometimes, allowed, willing

B. Orientation

"Orientation is the basic distinction that determines how each type of modality is realized: that is, the distinction between subjective and objective modality, and between the explicit and implicit variants" (Halliday, 2008, p.619). It

indicates the linguistic forms of expressing modality and the addresser's modal responsibility (Thompson, 1996). The system is shown in Figure 2.

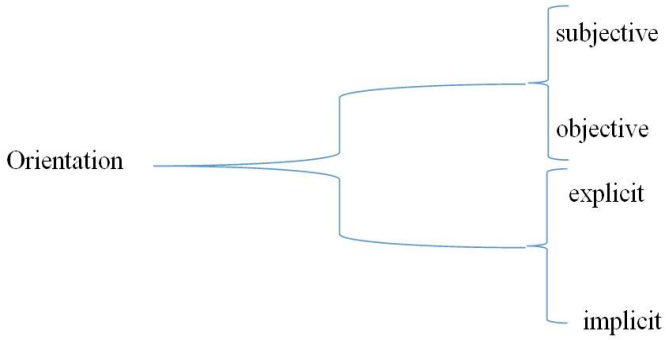

Figure 2 System of Orientation Types in Modality (Halliday, 2008, p.619)

"These combine with all the four types of modality, but with gaps; for example, there is no systematic form of making the subjective orientation explicit in the case of usuality or inclination (no coded expressions for "I recognize it as usual that…" or "I undertake for…to…"). This is a systematic gap; these particular combinations would represent semantic domains where the speaker cannot readily pose as an authority" (Halliday, 2008, p.619). Examples of the combination of orientation and type are given in Table 2.

TABLE 2
MODALITY: EXAMPLES OF "TYPE" AND ORIENTATION COMBINED (HALLIDAY, 2008, P.620)

	Subjective:	Subjective:	Objective:	Objective:
	explicit	implicit	implicit	explicit
Modalization: probability	I think Mary knows	Mary'll know	Mary probably knows	It's likely that Mary knows
Modalization: usuality		Fred'll sit quite quiet	Fred usually sits quite quiet	It's usual for Fred to sit quite quiet
Modulation: obligation	I want John to go	John should go	John's supposed to go	It's expected that John goes
Modulation: inclination		Jane'll help	Jane'll be keen to help	

C. Polarity

Polarity is "the opposition between positive ("It is. Do that!") and negative ("It isn't. Don't do that!"). The positive-negative opposition is fairly likely to be grammaticalized in every language, in association with the clause, as proposition or proposal. Typically, positive clauses are formally unmarked, while negative clauses are realized by some additional elements—for example, in English, by adding the word *not* located in the neighborhood of the verb" (Halliday, 2008, p.143).

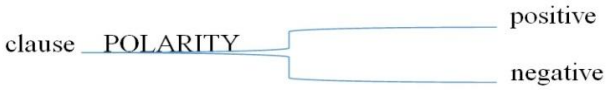

Figure 3 System of Polarity

Positivity and negativity are the contrast features of the clause, which may occur in a wide variety of ways. They represent an aspect of the meaning potential of the language. They can be defined mutually: "not positive" means "negative", and "not negative" means "positive" (Halliday, 2008).

TABLE 3
FINITE VERBAL OPERATORS: POSITIVE AND NEGATIVE (HALLIDAY, 2008, P.116)

Temporal operators			
	Past	Present	Future
Positive	did, was, had, used to	does, is, have	will, shall, would, should
Negative	didn't, wasn't, hadn't, didn't+used to	doesn't, isn't, hasn't	won't, shan't, wouldn't, shouldn't
Modal operators			
	Low	Median	High
Positive	can, may, could, might, (dare)	will, would, should, is/was to	must, ought to, need, has/had to
Negative	needn't, doesn't/didn't+need to, have to	won't, wouldn't, shouldn't, (isn't/was not to)	mustn't, oughtn't to, can't, couldn't, (may not, mightn't, hasn't/hadn't to)

D. Introduction-Methodology-Result-Conclusion Framework

The sample RAs selected for the corpus all followed the Introduction-Methodology-Result-Conclusion (IMRD) framework. Scholars in the English for specific purposes (ESP) School propose many move models. Bhatia (1993) found that four obligatory moves should be contained in an RA in his influential work *Language Use in Professional Settings:* introducing the purpose, describing the methodology, summarizing the results, and presenting the conclusions. These four moves comprise the IMRD model. The introduction move introduces the topic of the study and shows readers what progress has occurred, the previous research or success in related areas, and the purpose of this study; the methodology move lists the general processes of research, as well as the materials, tools, personnel, and objects of study involved in the research; in the results move, all findings, results, and conclusions are presented according to the methodology part; at finally, in the discussion move, the author makes a brief conclusion about the study and then clearly explains the results obtained during the process and gives suggestions and possible applications of the results. The IMRD model is widely employed in research articles in China and abroad.

III. Methodology

A. Corpus Compilation

As a part of the data gathering process, all data were searched and extracted from the introductions and discussions of 50 RAs chosen from the journal *Artificial Intelligence*, which commenced publication in 1970. It is now the generally accepted international forum for the publication of results of current research in this field. In our corpus, all chosen RAs were published between 2010–2019, five RAs were chosen from each year.

According to the purpose of each move, the attitude of the author is reflected primarily in the introduction and discussion moves, as such, this study analyzes the epistemic modality in only these two moves of the identified RAs.

This study used a two-round random stratified sampling method to choose the 50 sample RAs for inclusion in the corpus. In the first round, to reduce the influence of differences in corpus length, the length of the RAs was limited to 10000–15000 words. All RAs meeting the criteria for each year were numbered, and 10 samples were selected from each year using a table of random numbers. One hundred total RAs were selected. In the second round, five from each of the 10 years were randomly selected; these comprise the corpus. As is shown in Table 4, the data contained 59,645 words in the introduction moves, and 20,981 words in the discussion moves. In total, 80,626 words were gathered.

Table 4
Basic Information On The Corpora

	Corpora
Genre	AI RAs
Journal	*Artificial Intelligence*
Years	2010–2019
Number of papers	50
Words	80,626
Researched sections of the papers	Introduction and discussion

B. Labeling

In SFG, orientations are classified as explicitly subjective (ES), implicitly subjective (IS), implicitly objective (IO), or explicitly objective (EO). Each of the orientations has a specific linguistic realization. Values are classified as low, median, or high. Polarities are classified as positive or negative. Tables 5–7 show the linguistic realization and some examples of the labeling results of orientation, value, and polarity which are highlighted with underlines. These epistemic modalities were manually labeled and then manually classified according to their orientation types.

Table 5
Orientations Of Epistemic Modality In Our Corpus

Orientation	Linguistic realization	Examples from our corpus
Explicitly subjective (ES)	An epistemic lexical verb with a first-person subject in a separate clause	"Therefore, we believe that our framework is effective and general to both RNN Search and Transformer."
Implicitly subjective (IS)	Modal auxiliary	"The resulting model may be represented as a CP-net, but the task here is to derive the preferences rather than the structure of the network."
Implicitly objective (IO)	Modal adverb	"For the same reasons, perhaps in algorithm portfolios, one should not just pick the best n algorithms."
	Expansion of the predicator	"…since complex algorithms are unlikely to be used in such trivial cases."
	Adjective	"However, yet more liberal definitions are possible."
Explicitly objective (EO)	Projecting clause of relational processes	"In this light, it seems likely that larger arities again yield further improvements…"
	Nominalization of a modal adjective	"This possibility is retained in the present scenario, which leads to a contradiction unless an additional snub literal is assumed."
	Modal lexical verb with an inanimate subject referring to the data, results, findings, etc	"Our experimental comparisons show that Sense Uniform Paths dominates the other two by significant margins."

TABLE 6
VALUES OF EPISTEMIC MODALITY IN OUR CORPUS

Value	Examples from our corpus
Low	"As well, it might be the case that we learn that two existing arguments are in an attack relation…"
	"It may be that one can define a learning process where these values are adjusted continuously based on experience…"
	"That additional expressivity could be matched by the extension of our approach in the direction of description logic that was outlined in the previous section."
Median	"This second objective function is probably the more natural one but note that the key arguments for our upper bound are essentially the same."
	"Data included 602 patients having C-reactive protein values above 60 mg/l upon admission to the hospital, which indicated a bacterial infection."
	"One potential problem with the proposed approach is that there is likely to be a need for several abnormality predicates for each class."
High	"Our results show that strong equivalence w.r.t. admissible, preferred, ideal, semi-stable, and respectively, eager semantics coincides…"
	"These variation operators must be unbiased, that is, treat the bit positions and the bit entries 0 and 1 in an unbiased way."
	"We conclude that neither b nor c is attacked by an in G ∪ H, (b, c) ∈ (G ∪ H)∗ a, and (b, c) ∈ ((G ∪ H)∗ a)gκ."

TABLE 7
POLARITIES OF EPISTEMIC MODALITY IN OUR CORPUS

Polarity	Examples from our corpus
Positive	"Many applications are possible for our models."
	"At lower precision, we can increase the size of…"
	"…different line of future work could be the use of the unifying…"
Negative	"…still hold but we do not know if the positive results carry…"
	"…more unusual appliances, this might not be the case…"
	"…and some cities may not even have legal…"

Due to the journal's specific structure, the subtitles of the discussion move in each RA in our corpus vary. All subtitles from the discussion moves in our corpus are shown in Table 8. In this step, all the labeled epistemic modalities are manually classified according to the move: introduction or discussion. Table 9 shows examples of epistemic modality belonging to the introduction and discussion moves.

TABLE 8
SUBTITLES OF INTRODUCTION AND DISCUSSION IN OUR CORPUS

Move	Subtitles
Introduction	Introduction
Discussion	Discussion Discussion and conclusions Conclusion Conclusions Summary, discussion, and future research Conclusions and future work Concluding remarks

TABLE 9
MOVES IN OUR CORPUS

Move	Examples from our corpus
Introduction	"The resulting model may be represented as a CP-net, but the task here is to derive the preferences rather than the structure of the network."
Discussion	"In the Rat domain, we have shown that Qube correctly captures the qualitative relations even when the underlying concept is relatively complex."

C. Methods of Analysis

The unit of analysis for the study was the clause. In SFG, a clause is any stretch of language centering around a verbal group, as shown in Ex.1. In this example, one verb (highlighted with an underline) was identified, and accordingly, one clause was identified.

Ex.1:|| ABD Transformer significantly outperforms the standard Transformer.||

In the study, the frequencies of the values, orientations, and polarities of all epistemic modalities in the RAs were counted using Microsoft Office Excel at the sectional level, and the epistemic modality in the corpus was labeled using Antconc 3.4.4, a freeware for corpus analysis, and checked manually. The total number of clauses was counted manually. The lexical density of the values, the orientations, and the polarities were calculated. To test for differences between the introduction move and the discussion move, a t-test was employed to calculate the P-value by SPSS, version 28.0. The formula is as follows:

$$D \text{ (Lexical density)} = n/N \times 100\%$$

Where n= The number of epistemic modalities

N = The total number of clauses

In this section, a sample epistemic modality from the corpus is labeled, classified, and analyzed, and the RA writer's

attitudes toward the statement (certain or uncertain, subjective or objective, positive or negative) are explained.

Ex.2: "The above-mentioned results show the (unary) unbiased black-box model for several classic test problems.....

The above clause comes from a sample RA from the corpus. Firstly, the epistemic modality is identified. In this clause, the epistemic modal expression is "show". Secondly, the orientation of the epistemic modality is labeled. It is the linguistic realization of EO because its clause structure is a modal lexical verb with an inanimate subject and the subject refers to the data, results, or findings. Thirdly, according to the degree of certainty, the sample is labeled as high value. Fourthly, the sample is labeled as positive polarity. Finally, this clause comes from the introduction move of the sample RA, the sample is labeled into introduction move. As a result, this sample is labeled as follows:

Ex. 3: "The above-mentioned results show that the (unary) unbiased black-box model for several classic test problems…" (explicitly objective, high, positive, introduction)

The sample epistemic modality with an EO orientation shows that the writer makes their viewpoints objective by presenting their views as facts. This orientation conveys the objectivity of the proposition and blurs the relationship between the writer and the proposition by concealing the writer's role in the assessment of the proposition. The use of objective orientations may reflect the AI RA writer's intention to shift readers' attention to the process and results of the research by distancing themselves from the text. By using an EO orientation, the writer avoids referring to themselves and foregrounds the relevance of their proposals for the benefit of the academic findings. The sample epistemic modality with high value shows that the resulting data is precise and reliable. The high degree of certainty may increase the objectivity of the results. The high value in presenting evidence in support of the assertions made can be a good way to make the statements more convincing and acceptable.

D. Rating Reliability

To minimize the risk of arbitrariness and guarantee the reliability of the results, two raters engaged in the manual labeling to label five randomly selected RAs from our corpus independently. The invited raters were both well acquainted with SFG. After the labeling of each RA, both the researcher and the first rater checked the text for labeling divergence. Through discussion and negotiation of label differences, the standard of labeling assignments is clarified. To achieve a high level of reliability, the same 50 RAs were re-labeled by the researcher and the second rater two months after the initial classification. The reliability index for both inter- and intra-rater agreement was over 90%.

IV. Results and Discussion

A. Frequencies of Epistemic Modality

1. Frequencies in Terms of Value

As shown in Table 10, despite the differences in the percentages of the values used in different moves, high- and low-value modality comprise similar percentages, with a total frequency of 264 (40.80%) and 286 (44.20%), respectively. Median-value modality comprises a small proportion, with a total frequency of 97 (15.00%). However, in his research on epistemic modality in medical RAs in 2015, Yang et al. (2015) found that medical RA writers prefer to express uncertainty. The research states that low-value epistemic modality was most frequently used in all four IMRD sections, with the least frequently used being high value (Yang et al, 2015). Similarly, both Vihla's (1999) and Hyland's (1996) studies about medical and biological research writing reported that low- and median-value modality are the most frequently used modal expressions. Li (2019) also found that low value is the most frequently used type and that high value is the least frequently used type in RAs on the humanities, social sciences, and natural sciences.

The reason is associated with the characteristic of medical and biological science; the resulting data in medical research is often less precise and reliable than that in the hard sciences. Vazquez and Giner (2008) explained that researchers in the soft sciences may not be able to show the same confidence as researchers in the hard sciences.

Table 10
Values Of Epistemic Modality In Research Articles From The Corpus

Value	Frequency	Percentage
Low	264	40.80%
Median	97	15.00 %
High	286	44.20 %
Total	647	100.00%

Table 11 outlines the typical high-value epistemic modality and its frequency. According to the statistics, the most frequently used high-value modalities are "show" and "believe". Both of them are verbs used epistemically. In particular, the verb "show" has a clear higher frequency, at 61 times. As Exs. 4 and 5 demonstrate, almost all occurrences of the verbs "show" and "believe" were used with first-person subjects (I and we) to express a strong commitment to the truth of the utterance.

Ex. 4: "We show that these two questions are…"

Ex. 5: "We believe that the approach proposed in…"

The results show that the usual form of high-value modality is a verb. However, in a study of parliamentary discourse Vukovic and Milica (2014) found that the most usual form of expressing strong (high-value) modality is through

emphasizers (i.e., strong epistemic adverbs); epistemic verbs are the runners-up. The difference might be caused by specific features of parliamentary language. Furthermore, a finding of Vukovic and Milica's study that coincides with ours is that high-value epistemic modality is much more frequently expressed through verbs than nouns and epistemic modal verbs. They believed that this might be a consequence of the characteristics of the English language.

TABLE 11
FREQUENCY OF HIGH-VALUE MODALITY

Modality	Frequency
"show"	61
"believe"	12
"uncertainty"	5
"certainly"	3
"conclude"	3
"uncertain"	2
"not possible"	2
"know"	2
"must"	1

Table 12 presents the typical low-value epistemic modality and its frequency. According to the statistics, the most frequently used low-value modalities are "possible" and "able," at 23 and 21 times, respectively. Both are modal adjectives. In Ex. 6, "[be] possible that" allows a more dialogic interpretation between writer and readers. In Ex. 7 the use of "be able to" demonstrates the author's apparent intention of showing uncertainty. The author is being careful not to convey a rigid description. This brings politeness into the discourse.

Ex. 6: " While it is possible that the coding machinery…"

Ex. 7: "On these methods, we can derive a large-scale taxon…"

The results show that the usual form of low-value modality is realized by modal adjectives.

TABLE 12
FREQUENCY OF LOW-VALUE MODALITY

Modality	Frequency	Modality	Frequency
"possible"	23	"perhaps"	3
"able"	21	"possibly"	3
"suggest"	7	"potentially"	3
"possibility"	6	"could"	2
"may"	4	"might not"	2
"can"	4	"do not know"	1
"might"	4	"probability"	1

The form of median-value modality is diverse, including the modal auxiliary, expansion of the predicator, and modal lexical verb. The modal auxiliary "will" had the highest frequency, at seven times. "Indicate," the modal lexical verb, appeared five times, as did the expansion of the predicator, "be likely to." The modal lexical verb "imply" appeared four times. "Probable" and "think" appeared only once. In Ex. 10, the author is sure about their findings. The function of "indicate" is not to convey vagueness but to bring a certain degree of politeness to the discourse.

Ex. 8: "…to more promising solutions will also be key."

Ex. 9: "…however HMMs are likely to fail for appliances with a…"

Ex. 10: "The experimental results indicate that the algorithms proposed…

Ex. 11: "…and explanatory discourse implies that these are important…"

TABLE 13
FREQUENCY OF MEDIAN VALUE MODALITY

Modality	Frequency
"will"	7
"indicate"	5
"be likely to"	5
"imply"	4
"probable"	1
"think"	1

In short, AI RA writers most frequently use verbs "show" and "believe" as high-value modality; modal adjectives are most frequently used low-value modality.

2. Frequencies in Terms of Orientation

As shown in Table 14, owed by ES 153 (23.65%), IO 124 (19.17%). The IS orientation was the least frequently used, with 78 instances (12.06%). However, research in the epistemic modality in medical RAs claims that medical RA writers prefer to be subjective. Li (2019) states that in humanity, social sciences, and natural sciences RAs, the IS orientation is the most prominent type of epistemic modality among all three disciplines. In the same vein, Yang et al.

(2015) illustrated that the IS orientation was the most frequently used by medical RA writers, while the ES orientation was the least frequently used.

TABLE 14
ORIENTATIONS OF EPISTEMIC MODALITY IN RESEARCH ARTICLES FROM THE CORPUS

Orientation	Frequency	Percentage
ES	153	23.65%
IS	78	12.06%
IO	124	19.17%
EO	292	45.13%
Total	647	100%

Table 15 presents the EO modality in the corpus and its frequency. As shown in the table, the verb "show" came first by a clear majority, with a frequency of 30, followed by "possible," at 23 times. In Ex. 12, the verb "show" is found in copulative structures with impersonal subjects and agentless passive clauses.

Ex. 12: "…and it is shown that this method is feasible…"

This indicates that the usual forms of EO modality are verbs and modal adjectives. The results are also confirmed by other researchers. In his research, Yang et al. (2015) reported, "We also found that more than 75 percent of modal expressions with an EO orientation were realized by modal lexical verbs with an inanimate subject referring to the data, results or findings of the current research" (Yang et al., 2015, p.7).

TABLE 15
FREQUENCY OF EXPLICITLY OBJECTIVE MODALITY

Modality	Frequency	Modality	Frequency
"show"	30	"imply"	4
"possible"	23	"certainly"	3
"suggest"	7	"possibly"	3
"possibility"	6	"uncertain"	2
"indicate"	5	"not possible"	2
"likely"	5	"probable"	1
"uncertainty"	5	"believe"	1

The corpus contains a very small variety of IO modalities (Table 16): "able," "potentially," "perhaps," and "probability." The modal adjective "able" is prominent, with a high frequency of 21. The others appear only a small number of times. Ex. 13 demonstrates the politeness function in the author's use of "be able to." In these cases, the author is trying to transmit an implicit recognition of alternative voices in their approach. Politeness is, of course, also implied in this action.

Ex. 13: "Our approach is generally able to make accurate predictions…"

TABLE 16
FREQUENCY OF IMPLICITLY OBJECTIVE MODALITY

Modality	Frequency
"able"	21
"potentially"	3
"perhaps"	2
"probability"	1

As shown in Table 17, all IS modalities in the corpus are modal verbs, except for the modal adverb "perhaps." The most frequently used IS modality is "will."

Ex. 14: "…text genres and domains, perhaps including micro-blog and…"

Ex. 15: "…to more promising solutions will also be key."

TABLE 17
FREQUENCY OF IMPLICITLY SUBJECTIVE MODALITY

Modality	Frequency
"will"	7
"can"	4
"might"	4
"may"	4
"could"	2
"might not"	2
"may not"	1
"perhaps"	1
"must"	1

Table 18 presents the ES modalities and their frequencies. The modalities are not diverse but have a significant regularity; that is, all ES modalities are verbs. Among them, "show" has the highest frequency, of 31 times. Following it is "believe," at 11.

Ex. 16: "We showed that appliance models built…"
Ex. 17: "…we believe that a robot should base its…"

TABLE 18
FREQUENCY OF EXPLICITLY SUBJECTIVE MODALITY

Modality	Frequency
"show"	31
"believe"	11
"conclude"	3
"know"	2
"think"	1
"not know"	1

In short, verbs and modal adjectives are the most usual forms of EO modality, the most usual form of ES modality is the verb, and the most usual form of IO modality is the modal adjective. For IS modality, the most usual is a modal verb.

3. Frequencies in Terms of Polarity

Table 19 shows that the total frequency of positive epistemic modality is 595 (91.96%), while the total frequency of negative epistemic modality is 52 (8.04%). The frequency of positive modality was approximately nine times higher than that of negative modality. Coincidentally, the result is the same as Halliday's statement in his book *An Introduction to Functional Grammar*. Halliday (2008) asserts that all clauses selected in the system of Polarity are positive terms and negative terms, if he takes account of a wide range of discourse types, positive probably works out around ten times as frequently as negative. Thus, he attaches probabilities to the two terms: 'positive, 0.9; negative, 0.1.'

TABLE 19
POLARITIES OF EPISTEMIC MODALITY IN RESEARCH ARTICLES FROM THE CORPUS

Polarity	Frequency	Percentage
Positive	595	91.96%
Negative	52	8.04%
Total	647	100%

As shown in Table 20, the positive modal words comprise more than 20 examples; most of the modalities are positive polarity. The verb "show" came first by a clear margin. These statistics demonstrate that the usual form of positive modality is the verb and the modal adjective.

TABLE 20
FREQUENCY AND VALUE OF POSITIVE MODALITY

Modality	Frequency	Modality	Frequency
"show"	61	"imply"	4
"possible"	23	"can"	4
"able"	21	"certainly"	3
"believe"	12	"conclude"	3
"suggest"	7	"perhaps"	3
"will"	7	"possibly"	3
"possibility"	6	"potentially"	3
"likely"	5	"could"	2
"indicate"	5	"know"	2
"might"	4	"must"	1
"may"	4	"probability"	1
		"think"	1

Table 21 presents the negative modalities in the corpus. Both the kind and frequency are few. All of the negative modalities are modal expressions in the negative form. "Uncertainty" has the highest frequency, at five times.
Ex. 18: "…to take into account the uncertainty in the composition of…"

TABLE 21
FREQUENCY OF NEGATIVE MODALITY

Modality	Frequency
"uncertainty"	5
"might not"	2
"not possible"	2
"uncertain"	2
"may not"	1
"no more probable"	1
"do not know"	1

In short, the usual form of positive modality is the verb and the modal adjective. All of the negative modalities are modal expressions in the negative form.

4. Comparison between Introduction and Discussion

To test for differences between the introduction and discussion moves, an independent t-test was employed to calculate the P-value by SPSS, version 28.0. This section calculates and discusses the density of modality in terms of value, orientation, and polarity. According to the resulting SPSS data, if the P-value is less than 0.05, the density has a difference between the introduction and discussion moves, and the differences are further discussed in detail. On the contrary, if the P-value is greater than 0.05, the density has no difference between the two moves.

Table 22 presents the results of lexical density in terms of value. All of the values in the discussion move seem to have higher lexical density than the introduction moves, except for median value. However, the t-test results show that the density has no difference between the two moves (P=0.657>0.05). Our results are in accord with the results of Yang et al. (2015): low-value epistemic modality is most frequently used in all four IMRD sections, which means all four IMRD sections have the same low-value modality density.

TABLE 22
RESULTS OF LEXICAL DENSITY IN TERMS OF VALUE

	Low (%)	Median (%)	High (%)	Total (%)
Introduction	4.66%	1.93%	5.08%	11.67%
Discussion	6.34%	1.72%	6.79%	14.90%
Overall	5.10%	1.87%	5.52%	12.49%

Table 23 presents the results of lexical density in terms of orientation. Similarly, all orientations in the discussion move appear to have higher lexical density than the introduction moves, except for IO orientation, but the results of the independent t-test show that the density has no difference between the two moves (P=0.684>0.05).

TABLE 23
RESULTS OF LEXICAL DENSITY IN TERMS OF ORIENTATION

	ES (%)	IS (%)	IO (%)	EO (%)	Total (%)
Introduction	2.71%	1.35%	2.53%	5.08%	11.67%
Discussion	3.65%	1.94%	2.01%	7.30%	14.90%
Overall	3.00%	1.51%	2.40%	5.64%	12.49%

Table 24 presents the result of lexical density in terms of polarity. Again, all of the polarities in the discussion move appear to have higher lexical density than the introduction move, but the results show that the density has no difference between the two moves (P=0.720>0.05).

TABLE 24
RESULTS OF LEXICAL DENSITY IN TERMS OF POLARITY

	Positive (%)	Negative (%)	Total (%)
Introduction	10.68%	0.99%	11.67%
Discussion	13.80%	1.10%	14.90%
Overall	11.48%	1.01%	12.49%

B. Functions of Epistemic Modality

1. Interpersonal Function from the Perspective of Semantics.

In our research, the high- and low-value modalities comprise similar percentages, while the median-value modality comprises a small proportion. Median value expresses a lack of commitment; tentative, low-value weak commitment, and high-value strong commitment (Simon-Vandenbergen, 1997). Low value helps writers to be more tentative and cautious in the presentation of their propositions. Using a high value means that writers are positive about their conclusions. The data indicates that writers take different attitudes and hold different views toward AI technology and have not achieved consensus. The possible reasons for this are as follows. The reason for the use of high-value modality may be associated with the nature of AI. AI technology belongs to the hard sciences including the natural and physical sciences, such as physics, biology, and chemistry, in which research is done by using hypotheses and experiments. Since research in hard science is usually based on data and very systematic, the research data and conclusions of hard science are accurate and reliable. Thus, writers tend to use high-value modality to express certainty. Nevertheless, AI is an emerging subject and has not fully developed. Although the recent breakthroughs of AI technology are very impressive and based on methods with parallels in the human brain, the question of whether such systems display real intelligence remains relevant. On this, opinions are divided. As previously mentioned, AI has already surpassed humans in some areas. In other areas, more computing power, a combination of existing technologies, and new technologies being developed may lead to similar levels of human performance. As such, many writers tend to use low-value modalities to express uncertainty. As a consequence, the data obtained from RAs taking place in the AI area results in a mixed situation. The finding obtained from the data, that the high- and low-value modalities comprise nearly half percentage, respectively, greatly reflects scholars' attitudes toward the statements of AI nowadays. As such, value expresses the certainty or uncertainty of the clause, realizing the interpersonal function of epistemic modality from the perspective of semantics.

2. Interpersonal Function from the Perspective of Pragmatics.

EO was the most frequent orientation in the corpus, followed by ES and IO. The IS orientation was the least frequently used. The frequent use of EO means AI RA writers prefer to be objective and tend to keep with the objective and impersonal features of academic writing. The low percentage of IS orientation indicates that AI RA writers tend to avoid presenting the subjective source of their claims. The total frequency of ES together with IS also comprises a relatively large proportion, which shows that AI RA writers appropriately express subjectivity. Modal expressions with subjective orientation convey the idea that what the writers have expressed is no more than their personal opinions, and that the readers' points of view are visibly invited. It promotes interaction between writers and readers, gains readers' acceptance of their claims, and contributes to the establishment of a good writer—reader relationship. In other words, epistemic modalities perform the negotiating function there. As such, the orientation expresses the subjectivity or objectivity of the clause, realizing the interpersonal function of epistemic modality from the perspective of pragmatism.

3. Interpersonal Function from the Perspective of Mood.

The frequency of positive modality was approximately nine times higher than that of negative modality. The results indicate that AI RA writers prefer to be positive and express certainty toward their propositions. The reason is that an association still exists with the characteristic of the AI subject, as a hard science, whose research results are usually based on data, systematic, accurate, and reliable. Authors use the positive modality to ensure the objectivity of their results. As such, the polarity expresses the positive or negative features of the clause, realizing the interpersonal function of epistemic modality from the perspective of mood.

V. Conclusion

Our research about scholars' attitudes (certain or uncertain, subjective or objective, positive or negative) toward the statement of AI reveals that these AI RA writers tend to use high- and low-value modality in similar percentages, with EO orientation and positive polarity. These results indicate that AI RA writers tend to ensure the objectivity of their results and that these writers adopt differing attitudes and hold varying views regarding AI technology; as such, they have not achieved consensus. This research further finds that epistemic modality performs an interpersonal function.

In terms of theory, this study enriches the study of epistemic modality and English academic discourse analysis from the perspective of SFG. In this study, the variables of the modality system, including value, orientation, and polarity, have been systematically analyzed in detail. By comparing the use of epistemic modality in different moves, studies of scholars' attitudes realized through epistemic modality in RAs are enriched. In practice, by analyzing the functions of the values, orientations, and polarities of epistemic modality, the findings of this study may help non-native AI RA writers to produce more acceptable AI RAs. A larger data collection of AI RAs and a further analysis of the functions of epistemic modality in AI RAs are expected in future work.

References

[1] Bhatia, V. (1993). *Analysing Genre: Language Use in Professional Settings.* London: Taylor & Francis Group.
[2] Cheng, W., & Cheng, L. (2014). Epistemic modality in court judgments: a corpus-driven comparison of civil cases in Hong Kong and Scotland. *English for Specific Purposes, 33*(1), 15-26.
[3] Fu, R. (2016). Comparing Modal Patterns in Chinese-English Interpreted and Translated Discourses in Diplomatic Setting: A Systemic Functional Approach. *Babel (Frankfurt), 62*(1), 104-121.
[4] Gao, Q, P. (2012). Xue shu ying yu xie zuo zhong ren zhi qing tai biao da de ren ji gong neng. *Chinese Journal of Applied Linguistics, 35*(3), 352-364.
[5] Halliday, M. A. K. (2008). *An Introduction to Functional Grammar* (-3rd ed.). Foreign Language Teaching and Research Press.
[6] Hyland, K. (1996a). Talking to the academy: Forms of hedging in science research articles. *Written Communication, 13*(2), 251-281.
[7] Hyland, K. (1996b). Writing without conviction? Hedging in science research articles. *Applied Linguistics, 17*(4), 433-454.
[8] Kane, T. (2019). Artificial Intelligence in Politics: Establishing Ethics. *IEEE Technology & Society Magazine, 38*(1), 72-80.
[9] Keskinbora, K. (2019). Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience, 64*, 277-282.
[10] Lei, X. (2020). Ji yu yu kun jing: ren gong zhi neng shi yu xia de chuan mei xin ye tai he lun li si kao. *Dong nan chuan bo,* (6), 32-34.
[11] Li, X, M. (2019). *Xi tong gong neng yu fa shi jiao xia ying yu xue shu yu pian zhong de qing tai dui bi yan jiu* [Unpublished doctoral dissertation]. Shandong University.
[12] Piqué-Angordans, J., Posteguillo, S., & Andreu-Besó, J. V. (2001). A pragmatic analysis framework for the description of modality usage in academic English contexts. *ELIA, 2*, 213-224.
[13] Piqué-Angordans, J., Posteguillo, S., & Andreu-Besó, J. V. (2002). Epistemic and deontic modality: A linguistic indicator of disciplinary variation in academic English. *LSP & Professional Communication, 2*(2), 49-65.
[14] Poole, R., Gnann, A., & Hahn-Powell, G. (2019). Epistemic stance and the construction of knowledge in science writing: A diachronic corpus study. *Journal of English for Academic Purposes, 42*, 100784.
[15] Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes (New York, N.Y.), 11*(2), 93-113.
[16] Shakirova, R., & Safina, A. (2019). Gender and Age Aspects within the Pragmatic Potential of the Epistemic Modality Markers. *Tarih Kültür Ve Sanat Araştırmalari Dergisi, 8*(1), 254-262.

[17] Simon-Vandenbergen, A. (1997). Modal (un)certainty in political discourse: A functional account. *Language Sciences (Oxford), 19*(4), 341-356.
[18] Simpson, P. (1993). *Language, Ideology and Point of View.* Florence: Routledge.
[19] Song, Z, J., & Hu, G, S. (2020). Ren gong zhi neng de lun li yu fa lv wen ti yan jiu. *Zhong guo Ke Ji Xin Xi,* (*16*), 98-99.
[20] Thompson, G. (1996). *Introducing functional grammar* (2nd ed.). London: Arnold.
[21] Vázquez, I., & Giner, D. (2008). Beyond Mood and Modality: Epistemic Modality Markers as Hedges in Research Articles: A Cross-Disciplinary Study. *Revista Alicantina De Estudios Ingleses, 21*(21), 171.
[22] Vihla, M. (1999). *Medical writing: Modality in corpus.* Amsterdam-Atlanta, GA: Rodopi B.V.
[23] Vukovic, & Milica. (2014). Strong epistemic modality in parliamentary discourse. *Open Linguistics, 1*(1), 37-52.
[24] Yang, A., Zheng, S., & Ge, G. (2015). Epistemic modality in English-medium medical research articles: A systemic functional perspective. *English for Specific Purposes, 38,* 1-10.
[25] Zhao, L, L., Jiang, B, J., & Li, K. (2020). Jiao yu ren gong zhi neng lun li de kun jing ji zhi li lu jing. *Dang dai jiao yu ke xue,* (*5*), 3-7.

Gefei Bo was born in Dalian, China in 1997.

She is currently an MA in School of Foreign Languages, Dalian University of Technology, Dalian, China. Her researches focus on systemic functional linguistics.

Jianjun Ma received her PH.D. degree in computer applications technology from Dalian University of Technology, Dalian, China in 2012.

She is currently a full professor in School of Foreign Languages, Dalian University of Technology, Dalian, China. Her research interests include systemic functional linguistics, academic discourse analysis, and corpus-based translation.