



ARTICLE



<https://doi.org/10.1057/s41599-024-03322-9>

OPEN

# Identifying stance in legislative discourse: a corpus-driven study of data protection laws

Le Cheng<sup>1</sup>, Xiuli Liu<sup>2</sup>  & Chunlei Si<sup>3</sup> 

Mirroring public ideologies and value systems in legislative discourse, stance not only functions as a powerful instrument for legislators to balance legal values and stakeholders' interests but also acts as a valuable reference for individuals to understand legislative texts. This study conducts a corpus-driven analysis of stance expressions in legislative discourse. Using three self-compiled corpora that incorporate data protection laws from the United States, the European Union, and China, we apply Hyland's stance model to contrastively analyse evidence of hedging, boosting, self-mention, and attitude markers across these jurisdictions and eventually propose a specialised research model of stance in law. This study unveils the nature of modesty and sufficient discursive space of data protection laws, as well as legislative values and public ideologies conveyed by different jurisdictions within the broader socio-legal cultural context. Besides uncovering the legal constructiveness of data protection laws, the results also suggest that the overall representation of stance in data protection legislation aligns with its performance in legislative discourse, which showcases a legislative tendency to achieve an overtly neutral appearance through covert stance expressions.

<sup>1</sup>Guanghua Law School, Zhejiang University, Hangzhou, People's Republic of China. <sup>2</sup>School of International Studies, Zhejiang University, Hangzhou, People's Republic of China. <sup>3</sup>Koguan School of Law, Shanghai Jiao Tong University, Shanghai, People's Republic of China. ✉email: [chunleisi@sjtu.edu.cn](mailto:chunleisi@sjtu.edu.cn)

## Introduction

Data serves as a central element (Sadowski 2019) in the contemporary digital world. The late 1950s witnessed a revolution in computing, resulting in the ongoing computerisation and digitisation of diverse paper-based archives and records (Hu et al. 2021), which led to a heightened concentration of data and information. With the rising tide of data protection concerns (Romanou 2018), people have been increasingly aware of their data and privacy rights (Wagner 2024). In light of these concerns, lawmakers globally have drafted and refined their respective data protection laws to prevent or mitigate associated social structural risks arising from the wrongful processing of data (Tamburri 2020) or data breaches (Zhao and Cheng 2024). Stance on data protection can be seen as one of the sources of numerous debates, suspicions, and fears. The extent to which personal data is restricted in its usage and the degree to which it is protected are closely intertwined with the socioeconomic development of the digital society, as well as human dignity and privacy rights.

The present study focuses on the stance of data protection laws in the United States, the European Union, and China. The reason for narrowing down our choice to these three jurisdictions is due to their salient features and significant impacts on the global landscape of data protection. The EU took the lead in creating the “right to the protection of personal data” (Fuster and Gutwirth 2013). Established in the EU Charter of Fundamental Rights in 2000, it not only plays a pivotal role in personal data protection in the EU but marks a milestone for the protection of personal data around the world. Data protection emerged almost simultaneously in the U.S. and the EU during the 1970s (Bennett 1988). Due to advancements in computer technology and large-scale Internet applications (Cheng and Liu 2024), the U.S. collects and stores almost the largest amount of data, which gives it rich practical experience in data processing and legal regulation. China, like other jurisdictions, enacted data protection laws later than the aforementioned blocks (Pernot-Leplay 2020). However, it is building its legal framework at a rapid pace (Cheng et al. 2023), showing its strong potential and burgeoning power. Given the notable representativeness of the data protection legislation across these three jurisdictions, a comparative investigation into their respective frameworks helps to gain a deep understanding of the stance spectrum of data protection.

The landscape of data protection laws is dynamically broadening, exemplified by data protection challenges (Li et al. 2023) posed by generative artificial intelligence typified by ChatGPT (Cheng and Liu 2023). This trend has infused the discourse on data protection laws and privacy with vitality and complexity. In addition to the collision of diverse legal views, one particular challenge is that, it calls for more interdisciplinary insights and knowledge paradigms (González et al. 2022) to enhance understanding of data protection. Whilst the discussion of data protection expands in scope and depth apparently, linguistic exploration of this theme remains an uncharted domain. Relevant literature has been published so far involving an institutional argumentative exploration of Swiss Federal Data Protection (Greco et al. 2016), and a study on data protection laws from the perspective of terminological variation and ideology by Pei and Cheng (2020). Compared to the latter one, the present study extends its scope from the U.S. and the EU to three jurisdictions with the inclusion of China, and shifts the perspective to “stance” so as to expand the breadth and depth of linguistic exploration in the field of data protection laws.

The research questions of this study are:

- (1) What are the divergences and convergences of stance expressions between data protection laws of the U.S., the EU, and China?
- (2) What is the overall stance orientation of data protection laws across the three jurisdictions?
- (3) What are the underlying rationales for the choices and representations of stance among the three jurisdictions?

This study aims to make contributions from both theoretical and practical perspectives. The theoretical significance of this study is twofold. Firstly, it fills in the gap in the study of data protection laws from a stance perspective, offering contributions to the literature on stance in legislative discourse. Secondly, it aims to put forward a specialised research model of stance which contributes to more informed and comprehensive stance identification and analysis in legal discourse. The practical significance of this study can also be delineated from two perspectives. The first is to provide guidance for individuals and assist them in navigating and seeking improved protection of their personality and privacy rights by uncovering the stance of data protection laws. Second, it helps legislators convey their conceptual views, value balances, and consensus expectations in a more precise way through the presentation of stance features when revising or refining laws.

## Stance and stance in law

**Background review of stance.** Generally, stance expressions play an important role in legislators’ articulation of legal values and strategic intentions within legislative texts. To investigate how legislators in different jurisdictions use stance as a tool to shape the identity of data protectors, thereby meeting the expectations of target audiences in legislative discourse, this study employs Hyland’s (2005, 2018) concept and framework of stance for analysis.

The conceptualisation of stance involves a set of umbrella terms, including attitude (Halliday 2004), appraisal (Martin 2000; Martin and White 2005), evaluation (Bednarek 2006; Hunston and Thompson 2000), interaction (Du Bois 2002), and posture (Grabe 1984). Among the current interpretations of the concept of stance, three of them are particularly influential and widely applied. Biber and Finnegan (1989: 93) define stance as “the lexical and grammatical expression of attitudes, feelings, judgments, or commitment regarding the propositional content of a message”. Biber et al. divide stance devices into three categories: epistemic stance, attitudinal stance, and style-of-speaking stance (Biber et al. 2000; Conrad and Biber 2000). Another conceptualisation of stance underscores its interactive aspect (Du Bois 2007), portraying stance as a form of social public behaviour and putting forward “stance triangle” as a structure for interactive discourse. Hyland (2005) posits that stance “expresses a textual ‘voice’ or community-recognised personality” and introduces the “stance and engagement framework” (Hyland and Jiang 2016), which is chosen as the main theoretical framework for the present study.

In terms of research fields, previous studies have shown that most stance-related research focuses on media discourse (Zhang 2018; Vis et al. 2012; Wu 2018; Myers 2014; Vertommen et al. 2012) and political discourse (Delandshere and Petrosky 2004; Chand 2011; Nir et al. 2014). This is attributed to the ease of identifying explicit stance expressions in these texts (Hyland and Jiang 2016). In recent years, stance research has flourished in academic writing, specifically branching into stance analysis of various academic genres such as abstracts (Hyland and Tse 2005; Pho 2008), undergraduate essays (Aull and Lancaster 2014), doctoral dissertations (Chan 2015), and research articles (Hyland 2012). It can be seen that stance research in legal discourse is far less abundant compared to other linguistic domains, which underscores the necessity for further studies.

**Integration into the profession: stance in law.** While research on stance as a theoretical or methodological framework in legal discourse may not be as extensive as in other discourse domains (Breeze 2011), it still covers various aspects within legal discourse. Evidence can be found in research on the construction of stances using the Noun in analysing strategic behaviour exhibited by Justices (Charles 2007), study on stance expressions in courtroom discourse (Chaemsaitong 2015, 2017), analysis of stances in court judgments (Cheng and Cheng 2014), study of the use of epistemic lexical verbs as hedging devices in different legal genres (Vass 2017), examination of stance adverbs in judicial opinions (Poole 2021), the investigation into stance markers in oral argument about marriage laws (Tracy 2011), corpus studies on the ideological stance of threatening language (Gales 2010), exploration of stance-taking on abortion jurisprudence (McKeown 2022) and evaluative analysis of the stance on reputational harm in defamation law (Izes 2023).

It can be seen that previous research on stance in law mainly focuses on the judicial domain, with comparatively fewer studies in the legislative field. As previously discussed, stance expressions and features are more readily detectable in discourse areas such as media, politics, and the judiciary as well. Compared to legislative discourse, this easily detectable presence of stance is notable in courtroom discourse and court judgments. According to Hyland and Jiang (2016), there is a close correlation between personal ideologies, value systems, and stance. Although the connection between legislation and individuals does not seem as close as in the judiciary, at a more abstract level, legislative texts do reflect deep-seated public ideologies and values.

Stance in legislative discourse involves a series of coherent rhetorical choices, which enables legislators or drafters to negotiate and balance various values associated with the law and legal subjects so as to meet the requirements of the rationality, validity, and fairness of legislation and the expectations of the general public toward the law. Based on the above literature review, it is evident that there is a need for ongoing academic focus on the stance in legislative discourse.

Inspired by the review of stance and stance in law, this study explores stance expressions in legislative discourse with a corpus-driven approach, especially the patterns of stance construction in data protection laws in the U.S., the EU, and China to identify their respective strategic expressions and stance orientation.

**Methodology**

**A corpus-driven approach to stance analysis.** Corpus, as a tool, is increasingly prevalent in quantitative research. Etymologically, the term *corpus* (plural is *corpora*) traces its origins to the Latin word for *body*. Following this, Baker (2010: 6) defines the corpus as “a ‘body’ of language, or more specifically, a (usually) very large collection of naturally occurring language, stored as computer files”. Therefore, it could be said that the corpus is “a finite collection of machine readable text” (McEnery and Wilson 2001: 197). As a form of empirical research, corpus analysis facilitates the exploration of real-life large-scale examples using tools such as *Antconc*, *Wmatrix*, *Lancsbox*, and others, rather than relying solely on a limited number of typical examples or introspection. Put in another way, corpus analysis enables us to gain a more objective understanding of the theme and object under study. For instance, in legal studies, traditional doctrinal research relies on in-depth analysis of one typical case to derive insights into a specific field. In contrast, corpus analysis allows extracting a large number of relevant court judgments (Mattioli and McAuliffe 2021) to obtain more extensive, evidence-based or data-based results for further attribution. Tognini-Bonelli (2001) distinguishes between *corpus-based* and *corpus-driven* research. The

**Table 1 The detailed information of the three corpora.**

Name of the corpus	Texts	Tokens
United States Legislation Corpus (USLC)	7	92,812
European Union Legislation Corpus (EULC)	2	58,902
Chinese Legislation Corpus (CLC)	3	17,453

former employs the corpus to validate researchers’ intuition through a source of examples, while the latter takes an inductive approach by utilising a series of real data and procedural designs. The corpus-driven approach is based on the view of corpus analysis as a method or methodology. Whether it is a corpus-driven or corpus-based study, representativeness is a fundamental and crucial consideration (Biber 2006) for the selection of corpus data.

Stance, as a kind of typical linguistic marker (Aull and Lancaster 2014), needs to be annotated or identified effectively, especially in large-scale texts. The manual annotation or identification of stance markers is a time-consuming process, which makes corpus analysis the most convenient and, consequently, the most prevalent method for researchers in stance studies. Scholars (e.g., Crosthwaite and Jiang 2017; Aull et al. 2017; Rezaei et al. 2021) employ the corpus as a tool to investigate various stance categories or conduct horizontal and longitudinal comparative studies. The present study, focusing on the stance analysis in data protection laws of the U.S., the EU, and China, appropriately adopts the corpus-driven approach, which shows both sound rationale and practical feasibility.

**Materials and corpora.** The corpus-driven approach requires the corpus data to be explored must be well matched to the research questions (McEnery and Hardie 2012). It implies that the selection of corpus data is aimed to best address the proposed research questions. In response to the research questions posed in this study, three comparable sub-corpora are self-compiled, consisting of 12 legislative texts pertaining to data protection laws and comprising a total of 169,167 tokens. In this context, the term ‘token’ refers to the occurrence of a word form in a text or corpus. As for the concept of comparable corpora, Hunston (2002) takes sampling from different varieties of the same language as an exemplary standard. Four aspects of specific standards were considered in this study to retrieve the materials, which include comparable language, representativeness, pertinent search terms, and authoritative database. The composition and size of the corpora are shown in Table 1.

Firstly, the United States Legislation Corpus (hereinafter referred to as USLC) comprises one federal and six state data protection legislative texts. Specifically, it includes the American Data Privacy and Protection Act (ADPPA), the California Privacy Rights Act (CPRA), the California Consumer Privacy Act (CCPA), the Colorado Privacy Act (CPA), the Connecticut Data Privacy Act (CTDPA), the Virginia Consumer Data Protection Act (VCDPA), and the Utah Consumer Privacy Act (UCPA). For an extended duration, the U.S. lacked a unified data protection law at the federal level. However, the enactment of the ADPPA on June 3, 2022, by the U.S. House of Representatives and Senate marked a groundbreaking moment in federal legislation, which establishes a robust national framework for personal data protection. In order to show a more comprehensive view of the U.S. data privacy laws, this study also includes the enacted laws of data privacy and protection in five states.

Secondly, the European Union Legislation Corpus (hereinafter referred to as EULC) includes two key legislative texts concerning personal data protection: the General Data Protection Regulation

(GDPR) and the Data Protection Law Enforcement Directive (Directive). Enacted in 2016, GDPR is not only the world’s earliest but the most comprehensive data protection law, serving as a pivotal benchmark in global legislative efforts (Hoofnagle et al. 2019; De Hert and Papakonstantinou 2016). Since its implementation in 2018, GDPR has produced numerous administrative enforcement cases. The leading role of the EU in data protection legislation and enforcement is tied up with its long-standing tradition of respecting and safeguarding privacy rights as fundamental human rights.

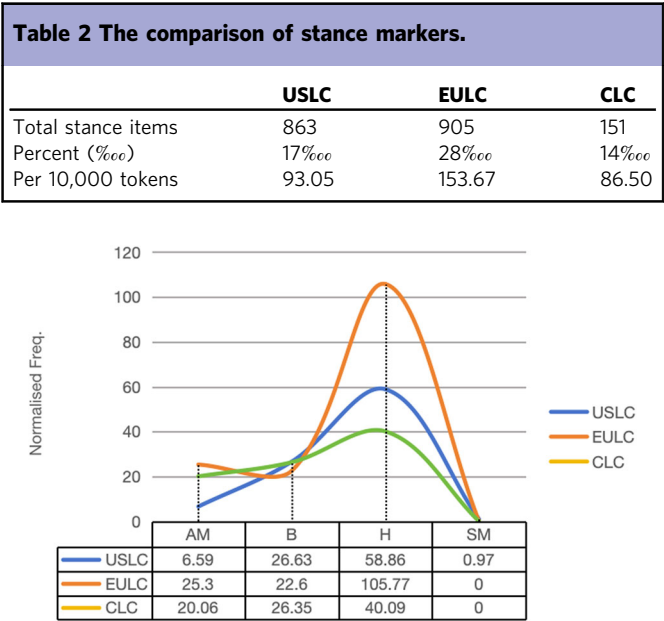
Thirdly, the Chinese Legislation Corpus (hereinafter referred to as CLC) is extracted from the PKU-law database (English version) and comprises three highly relevant data protection laws, namely, the Personal Information Protection Law of the People’s Republic of China (PIPL), the Data Security Law of the People’s Republic of China (DSL), and the Cybersecurity Law of the People’s Republic of China (CL). The CL, enacted on November 7, 2016, lays the groundwork for China’s data protection legal framework (Cheng and Liu 2022; Wang et al. 2020), explicitly emphasising the protection of personal information. Passed on June 10, 2021, the DSL stands as a foundational legal instrument in China’s data security domain, with a primary focus on national security. Enacted on August 20, 2021, the PIPL is China’s first comprehensive legislation on personal information protection (Zhao and Feng 2021), which shares notable similarities with GDPR.

Method

The following corpus analysis was carried out in three steps. First, the data for the three corpora were converted to plain text files, then into a corpus analysis tool *Lancsbox* (Brezina et al. 2015). In what follows, the corpora were searched for stance features following the stance markers list of Hyland (2018). This list of stance devices is widely applicable and characterised by its non-disciplinary specific nature (Crosthwaite et al. 2017), which makes it suitable for examining the stance features of data protection laws in the present study. The full list of stance items can be found in the Appendix. After automated searching in the corpora, each stance item was manually examined and counted to ensure its fidelity in performing the stance function. For instance, in the present corpora to be explored, *agree* in “Pre-Dispute Arbitration Agree” does not function as a stance marker but rather as part of a specific legal term. In this case, it requires manual exclusion. The statistical results of stance items were double-checked by the co-author and if necessary, further consultation with peers were sought to ensure the reliability of the data analysis.

Second, the “stance and engagement framework” (Hyland and Jiang 2016) was borrowed to analyse different dimensions of stance. While this framework includes both the writer-oriented features of “stance” and the reader-oriented features of “engagement”, we only examined the former. Regarding the latter category, which includes reader pronouns, directives, questions, shared-knowledge devices, and asides, they were observed to be rare in legislative texts. Hyland’s framework (Hyland 2005; Hyland and Jiang 2016) on stance comprises three main components: evidentiality, affect and presence, along with four types of stance devices: attitude markers, hedges, boosters, and self-mentions, which we will explain and analyse in detail below.

Third, based on the results of step 2, a concordance analysis was carried out to delve deeper into the findings obtained from the comparison of the three corpora. It involves extracting salient stance items and performing a concordance search to locate their contextual information, as outlined by Scott and Tribble (2006). According to Sinclair (1991), the meaning and function of a word



**Fig. 1 The distribution of stance devices in three corpora.** AM attitude markers, B boosters, H hedges, SM self-mentions.

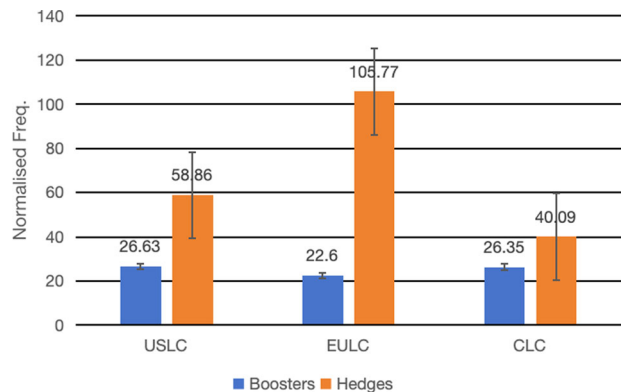
in a specific discourse are revealed through its co-textual features. Given that the analysis of concordances is a qualitative form of analysis and the analysis of frequency data is a quantitative form, this study adopts a mixed-methods approach combining both qualitative and quantitative analysis.

Corpus findings and analysis

**Cross-corpora comparison of stance markers.** The present study makes a holistic comparison of stance markers across the three corpora. As shown in Table 2, the proportion of stance items is highest in EULC, at approximately 28‰, while CLC has the lowest proportion at around 14‰, with USLC falling in between. It can be seen that in terms of data protection legislation, the stance of the EU appears more distinct and prominent compared to that of the U.S. and that of China. The findings also align with the view proposed by Pei and Cheng (2020) regarding the stringent standards set by the EU in data protection. However, the lower proportion of stance items in CLC does not directly imply a lack of clarity in China’s stance on data protection or even lower standards for data protection. Instead, it might indicate an intentional legislative strategy, which will be further explored and explained in the discussion section.

Figure 1 shows the distribution of stance devices across the three corpora. As we will present a comparative analysis of four distinct dimensions in detail below, here we will mainly focus on summarising their commonalities. The commonality analysis offers insights into the performance of various stance dimensions in data protection laws. It is evident that all three corpora exhibit the highest proportion of hedging devices, which indicates that hedging devices are most often used to express stance in data protection laws. Furthermore, it reflects the prudent and rational but intricate nature (Cheng and Cheng 2014) of data protection legislation and the discursive space constructed by legislators in this domain. Additionally, it is obvious that all three corpora show a remarkably low presence of self-mentions. The data reveals that both EULC and CLC demonstrate a complete absence of self-mention devices. This finding appears to contrast sharply with trends observed in stance studies of other genres. In certain contexts such as speeches or dialogues, self-mentions are relatively more frequent, serving to express the discursal self





**Fig. 2 The frequency of boosters and hedges used in data protection laws.** “Boosters” is represented by blue blocks and “Hedges” by orange blocks.

(Hyland 2005) and foster interaction between the author and potential audiences. However, it comes as no surprise that self-mentions are less common in expository texts, particularly in legislative texts, as the lower self-presentation aims to underscore the objectivity of the text. Therefore, it can be inferred that not only in the legislative texts of data protection in this study, but in legislative texts in general, self-mentions display a low prevalence to maintain and safeguard the objectivity of legislation. We then turn to the use of these stance features by section.

**Cross-corpus comparison by section**

*Evidentiality.* According to Hyland (2005), evidentiality refers to “the writer’s expressed commitment to the reliability of the propositions he or she presents and their potential impact on the reader”. Within the legislative context, the assessments of certainty enable legislators to convey legal statements from a particular perspective. Evidentiality is manifested through hedges and boosters (Hyland and Jiang 2016). Both elements indicate legislators’ explicit intrusion into a text to convey a legislative stance, whether to ensure reliability or convey legal reasoning. Therefore, by assessing the weight legislators impose on the propositions, we can infer their commitment to legislative reliability and legal reasoning. Figure 2 reports the distribution frequency of boosters and hedges in data protection laws. Next, we will conduct a comparative corpus analysis of these two elements.

Hedges convey a prudent stance of the writer by indicating a lack of complete commitment to a proposition, which allows information to be presented as an opinion rather than a fact (Hyland 2005). According to Fig. 2, we can observe that hedges are most prevalent in EULC, reaching 105.77 per 10,000 tokens, significantly exceeding those in USLC and CLC. It suggests the EU places greater emphasis on legal reasoning and rigour regarding data protection.

Further examination of the wordings for hedges reveals differences and similarities among the three jurisdictions regarding permissions and obligations in personal data protection laws. Table 3 reveals that the most frequently occurring hedging devices across all three corpora are primarily “modality”, including epistemic modality like *likely* and *possible*, as well as deontic modality such as *may* and *should*. Modality has been investigated in legal discourse in different contexts of situation (e.g., Cheng and Cheng 2014; Cheng and Sin 2011; Li et al. 2016), where it is used to reflect the ways in which legislators or jurists construct legal facts and indicate legal possibilities, or to examine the distribution of legal permissions and obligations. Table 3 shows that *should* has the highest frequency in EULC, whereas *may* is the most frequent hedging device in USLC and CLC. Each

**Table 3 The top 5 hedging devices ranked by frequency.**

Rank	Hedges		
	USLC	EULC	CLC
1	may (30.92)	should (41.93)	may (30.94)
2	about (7.54)	may (35.82)	relatively (2.86)
3	would (5.71)	possible (6.69)	assume (1.72)
4	should (5.28)	could (4.41)	possible (1.15)
5	possible (1.94)	likely (4.24)	

concordance line containing *should* or *may* was examined in context to validate the aforementioned claim.

- (1) [GDPR.txt] Results of such verification *should* be communicated to the person or entity referred to in point (h) and to the board of the controlling undertaking of a group of undertakings, or of the group of enterprises engaged in a joint economic activity, and *should* be available upon request to the competent supervisory authority.
- (2) [directive.txt] The principles of, and rules on the protection of natural persons with regard to the processing of their personal data *should*, whatever their nationality or residence, respect their fundamental rights and freedoms, in particular their right to the protection of personal data.
- (3) [ADPPA.txt] any time beyond the initial 2 times described in subparagraph (A), *may* allow the individual to exercise such right for a reasonable fee for each request.
- (4) [ADPPA.txt] A covered entity *may not* infer that an individual has provided affirmative express consent to an act or practice from the inaction of the individual or the individual’s continued use of a service or product provided by the covered entity.
- (5) [CL.txt] No individual or organisation *may* conduct any activity endangering cybersecurity,

The data in Table 3 initially appears to show that the EU personal data protection legal system places a strong emphasis on obligations, particularly those imposed on data processors. However, a further concordance analysis reveals that this preliminary observation does not seem to be accurate. While *may* typically conveys permission, its meaning changes to obligation when used in conjunction with *no* or *not* (Li et al. 2016). For instance, in the case of the U.S., Fig. 3 reports the occurrence of *may not* in USLC. However, despite this, obligations in the EU personal data protection laws still appear to be generally higher compared to those in the U.S. and China.

In terms of epistemic modality, *possible* appears as one of the top five hedging devices in all three corpora. As can be seen from the data in Table 3, the frequency of *possible* in EULC is 6.69 per 10,000 tokens, significantly higher than the frequencies in USLC (1.94) and CLC (1.15). It underscores the prudent approach of the EU personal data protection laws in indicating legal possibilities, which allows for discretion in legal interpretation (Cheng and Cheng 2012).

Boosters, on the other hand, express conviction and assert claims (Hyland 1999). As shown in Fig. 2, the frequency of boosters appears relatively average across the three corpora, with slightly higher occurrences observed in USLC and CLC compared to EULC. Noteworthy, this contrasts with the analysis of hedges earlier as indicated in Table 3. The relatively limited use of certainty-based stance markers in the EU personal data protection laws reflects a preference for objectivity and a reduced reliance on subjective beliefs (Hyland and Jiang 2016). This reaffirms the prudent and modest approach of the EU data protection legislation, as indicated in Fig. 2 and Table 3.

Search may not	Occurrences 26 (2.80)	Texts 3/7	▼ Corpus	USLC	▼ Context	7	▼ Display Text
Index	File	Left	Node	Right			
1	ADPPA.txt	EXPRESS CONSENT REQUIRED.—A cov-	16	ered entity	may not	infer that an individual has	17 provided
2	ADPPA.txt	PRETEXTUAL CONSENT PROHIBITED.—	22	A covered entity	may not	obtain or attempt to	23 obtain the
3	ADPPA.txt	NON-APPLICATION TO SERVICE PRO-	19	VIDERS.—An entity	may not	be considered to be	20 a third-party
4	ADPPA.txt	MINIMIZATION. 12 (a) IN GENERAL.—	A covered entity	may not	collect, 13 process, or transfer	covered data	
5	ADPPA.txt	15 ICE.—A covered entity or service	provider	may not	engage 16 in deceptive advertising	or marketing	
6	ADPPA.txt	a covered entity or 24 service	provider	may not—	kjohnson on DSK79L0C42PROD	with BILLS •HR 8152	
7	ADPPA.txt	SERVICE OR PRICING 16 PROHIBITED.—	A covered entity	may not	retaliate against 17 an individual	for exercising	
8	ADPPA.txt	18 (1) REQUIRED EXCEPTIONS.—A covered	entity 19	may not	permit an individual to exercise	a right	
9	ADPPA.txt	of the individual; 3 and 4 (B)		may not	process or transfer such addi-	5 tional	
10	ADPPA.txt	number of verified requests, on its own,		may not	4 be considered to render	compliance with	
11	ADPPA.txt	(1) IN GENERAL.—A covered entity—	14 (A)	may not	transfer or direct the transfer	15 of	
12	ADPPA.txt	TO 16 CHILDREN AND MINORS.—A covered	entity	may not	engage 17 in targeted advertising	to any	
13	ADPPA.txt	MINORS.— 21 (1) IN GENERAL.—A covered	entity	may not	22 transfer or direct the transfer of		
14	ADPPA.txt	covered entity or a service 24	provider	may not	collect, process, or transfer covered	kjohnson on	
15	ADPPA.txt	would not otherwise do so; 20 (2)		may not	collect, process, or transfer service	21 provider	
16	ADPPA.txt	in the guidelines, then the 14	Commission	may not	withdraw approval of such	guidelines 15 on	
17	ADPPA.txt	under this Act, the Com-	11 mission	may not	seek a cease and desist order	against	
18	ADPPA.txt	State 5 (B) RETAINED AUTHORITY.—	Subparagraph 6 (A)	may not	be construed to limit the	authority 7	
19	ADPPA.txt	9 (2) STATE LAW PRESERVATION.—	Paragraph (1) 10	may not	be construed to preempt, displace,	or sup-	
20	ADPPA.txt	a regulation 17 promulgated under	this Act	may not	be pleaded 18 as an element of		
21	ADPPA.txt	performing the purposes authorized	by subdivision (a)	may not	use the sensitive personal	information, after it has received	
22	CPRA.txt	an investigation or civil action. The	Agency	may not	limit the authority of the	Attorney General	
23	UCPA.txt	taking action. 408 (4) (a) A	controller	may not	charge a fee for information	in response	
24	UCPA.txt	otherwise provided in this chapter,	a controller	may not	process sensitive 477 data	collected from a	
25	UCPA.txt	and exemptions. 483 (4) (a) A	controller	may not	discriminate against a consumer	for exercising a	
26	UCPA.txt	each allegation. 622 (b) The	attorney general	may not	initiate an action if the	controller or	

Fig. 3 The occurrences of *may not* in USLC.

Table 4 The top 5 boosting devices ranked by frequency.			
Rank	Boosters		
	USLC	EULC	CLC
1	establish (3.77)	established (9)	establish (10.89)
2	established (3.77)	demonstrate (3.4)	finds (4.58)
3	clear (3.45)	clear (2.21)	TRUE (2.29)
4	known (3.02)	establish (1.7)	find (1.72)
5	clearly (2.37)	must, certain (1.19)	truly, certain, clearly (1.15)

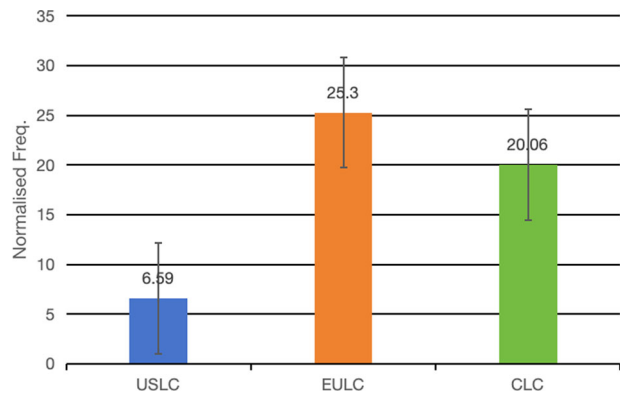
When we turn our attention to wordings for boosters, a notable observation stands out: personal data protection legislation demonstrates a strong sense of legal constructiveness. It is closely related to data protection as an emerging field, where rules and principles necessitate establishment and reinforcement through stance markers. It can be seen from Table 4 that *establish(ed)* appears with high frequency across all three corpora, with values of 7.54 in USLC, 10.7 in EULC, and 10.89 in CLC. In USLC, *establish* follows with *meaningful enforcement*, which indicates the emphasis placed on enforcement of the U.S. legal rules of data protection laws. In EULC, the occurrence of *established* also reflects the adaptational requirements of the principles and rules of the GDPR. The high frequency of *establish* in CLC aligns with the characteristic of strong legal constructiveness in Chinese legislation (Paler 2005). In Chinese data protection legislative texts, *establish* consistently co-occurs with *improve*, further reflecting this attribute. Based on the analysis of the boosters, it can be noted that legislative texts on data protection consolidate and emphasise the legal rules and principles constructed by stance markers such as *establish(ed)*.

- (6) [ADPPA.txt] To provide consumers with foundational data privacy rights, create strong oversight mechanisms, and *establish* meaningful enforcement.
- (7) [directive.txt] Regulation (EC)No 45/2001 and other Union legal acts applicable to such processing of personal data should be adapted to the principles and rules *established* in Regulation (EU)2016/679.

- (8) [CL.txt] *establish* and improve the cybersecurity guarantee system, and enhance the capability to protect cybersecurity.

**Affect.** Attitude markers, as one of the dimensions of stance, convey the writer’s affect, involving emotions, perspectives, and beliefs towards propositions (Hyland 2005). Examples include linguistic markers such as *agree*, *interestingly*, *appropriate*, and others that directly express the writer’s affective stance. Affect typically conveys the writer’s sentiments or evaluations, carrying a strong sense of subjectivity. Consequently, affective expressions are relatively rare in legislative texts. However, it is precisely because of this rarity that the attitude markers appearing in the corpus to be explored hold greater research value, as they denote a strong and distinct stance and value orientation (Yu 2023). Legislators seek to convey stance on specific legal topics through shared attitudes, values, and reactions in an implicit way. Attitude markers are typically categorised into negative and positive forms (Hyland and Jiang 2016), reflecting the attitudinal inclinations of the writer.

Figure 4 presents a comparison of the frequency of attitude markers across the three corpora. It is evident that attitudinal stance is more frequent in EULC, while appearing less frequently in USLC, at only 6.59 per 10,000 tokens. It suggests a concerted effort by the EU legislators to convey and underscore a strong sense of stance and value orientation in the realm of personal data protection. This point will be further elaborated in the discussion section in conjunction with the socio-legal context.



**Fig. 4** The frequency of attitude markers used in data protection laws. A comparison of the USLC, EULC, and CLC.

More valuable evidence comes to light when we look at the specific use of attitude markers. As depicted in Table 5, *appropriate*, as an attitude marker, appears relatively frequent in the three corpora. It peaks in EULC at 21.73 occurrences per 10,000 tokens, followed by 7.45 in CLC, and though lower at 4.53 in USLC, it still ranks first within this sub-corpus. *Appropriate* seems to convey a leaning-towards-negative meaning (Appel et al. 2016), which indicates a higher level of prudence compared to positive-oriented attitude markers. However, negative-oriented attitude markers can also convey a clear stance, that is, a prudent and impartial legislative attitude. For instance, expressions like “take reasonable and appropriate steps” in the CPRA text and “at appropriate time” in the CL text serve this function effectively.

The findings also unveil another function of negative-oriented attitude markers, that is, the neutralisation and weakening effect to disguise overly strong and direct stance expressions by legislators. In the following examples, GDPR imposes requirements for technical and organisational measures on data processing activities for scientific, historical research, or statistical purposes. This requirement itself represents an obligation and a direct and strong expression of stance. However, the inclusion of *appropriate* in this expression seems to mitigate the compulsory nature of the requirement. Nevertheless, the demand in this statement remains the same even without *appropriate*. It also illustrates that *appropriate* in this context serves as a legislative discursive strategy to attenuate the stance.

- (9) [GDPR.txt] accordance with Article 89(1) subject to implementation of the *appropriate* technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
- (10) [GDPR.txt] processed in a manner that ensures *appropriate* security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using *appropriate* technical or organisational measures ('integrity and confidentiality').
- (11) [CPRA.txt] grants the business rights to take reasonable and *appropriate* steps to help to ensure that the third party, service provider, or contractor uses the personal information transferred in a manner consistent with the business's obligations under this title.
- (12) [CL.txt] The standardisation administrative department of the State Council and other relevant departments of the State Council shall, according to their respective functions, organise the formulation of and revise at *appropriate* time national and industry standards relating to cybersecurity administration and the security of network products, services and operations.

**Table 5** Top 3 The attitude markers ranked by frequency.

Rank	Attitude markers		
	USLC	EULC	CLC
1	appropriate (4.53)	appropriate (21.73)	important (9.17)
2	essential (0.65)	important (1.02)	appropriate (7.45)
3	expected (0.32)	agree (0.68)	agreed (2.29)

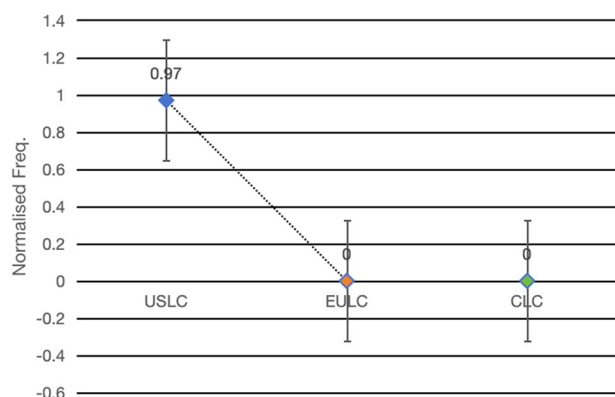
Additionally, it is important to highlight certain attitude markers conveying positive meanings. For instance, *important*, which ranks first in CLC and second in EULC, as well as *essential*, which ranks second in USLC. Analysis of collocations and concordances reveals that *important* in CLC is always collocated with *data*. It underscores a significant feature of China's data protection legislative system, which emphasises graded classification for data protection. The security and protection of important data are central to China's data classification system. Balancing data security with developmental needs is a pressing task, highlighting the urgency for streamlined data management. Both the Cybersecurity Law and Data Security Law in China outline specific provisions for managing important data (Cheng et al. 2023), including cataloguing, designated personnel, and risk evaluation. It underscores China's prioritisation of “security”, “national interests”, and “public interests” within its legislative endeavours in data protection.

In EULC, *important* typically collocates with *public interest*. It reflects the EU's exception principle in data processing, whereby data processing restrictions can be overridden based on important public interests of the member states. It also mirrors the EU's emphasis on public interest in the domain of personal data protection. In USLC, *essential* usually co-occurs with *goods* and *services*, shedding light on a key aspect of the U.S. data protection legislation. Prior to the enactment of comprehensive national data privacy laws, the most comprehensive and typical personal data protection is the CCPA focusing on consumer privacy protection. The primary concern in the U.S. regarding personal data and privacy rights stemmed from the balance of interests between businesses and consumers in the market economy (Acquisti et al. 2016).

- (13) [CL.txt] Taking measures such as data categorisation, and back-up and encryption of *important* data.
- (14) [DSL.txt] The national data security work coordination mechanism shall coordinate the relevant departments in the formulation of a catalogue of *important* data to strengthen the protection of *important* data.
- (15) [GDPR.txt] exercise or defence of legal claims or for the protection of the rights of another natural or legal person or for reasons of *important* public interest of the Union or of a Member State.
- (16) [CPA.txt] “DECISIONS THAT PRODUCE LEGAL OR SIMILARLY SIGNIFICANT EFFECTS CONCERNING A CONSUMER” MEANS.....OR ACCESS TO *ESSENTIAL* GOODS OR SERVICES.

**Presence.** Presence typically refers to the author's projection into the text (Hyland 2005), often achieved through self-mention devices such as first-person pronouns and possessive determiners. Self-mention devices commonly include “I, we, me, my, our, mine, us, the author, the author's, the writer, and the writer's” (Hyland 2018: 269–270). It can be inferred this kind of stance-expressing device tends to appear less frequently in legal texts, especially legislative texts. This is because legislative texts typically





**Fig. 5 The frequency of self-mention used in data protection laws. A comparison of the USLC, EULC, and CLC.**

strive to avoid subjective statements and expressions as much as possible. Furthermore, the legislation represents not the will or thoughts of individual legislators, but rather serves as legal instruments designed by the entire nation to achieve specific purposes. As illustrated in Fig. 5, self-mentions have very low prevalence across all three corpora, with their absence in EULC and CLC being particularly notable. However, it is worthy of exploration how self-mentions appear and function within USLC to express stance.

The corpus search results indicate that *my* appeared 5 times and *our* appeared 4 times in USLC. Further concordance analysis reveals that *my* in USLC does not specifically refer to legislators or legislatures but rather refers to data subjects in the form of headline-like quotations. While this way facilitates interaction between the legislature and the audience targeted by legislative texts, it does not strictly function as a self-mention marker. In addition, *our* in USLC effectively performs stance function by linking legislators with readers through phrases such as “our society”, “our personal information”, and “our desire”, enhancing the readers’ identification with legislative values.

- (17) [CCPA.txt] Provide a clear and conspicuous link on the business’ Internet homepage, titled “Do Not Sell My Personal Information”, to an Internet Web page that enables a consumer, or a person authorised by the consumer, to opt out of the sale of the consumer’s personal information.
- (18) [CPRA.txt] California is the world leader in many new technologies that have reshaped *our* society.
- (19) [CCPA.txt] A series of congressional hearings highlighted that *our* personal information may be vulnerable to misuse when shared on the Internet. As a result, *our* desire for privacy controls and transparency in data practices is heightened.

However, it is noteworthy that the self-mention device *our* appears in the legislative text of the US data protection laws not within statutes but rather in the declaration section of the legislature. Hence, it remains evident that self-mention devices are rarely present in formal legislative texts, in order to ensure the authority and objectivity of the law to the greatest extent.

## Discussion

**Stance practice in data protection laws.** In the preceding section, we conducted a thorough analysis of the divergences and convergences in stance expression between the data protection laws of the U.S., the EU, and China, thus addressing RQ1. Building upon this foundation, we aim to answer RQ2 and RQ3 through

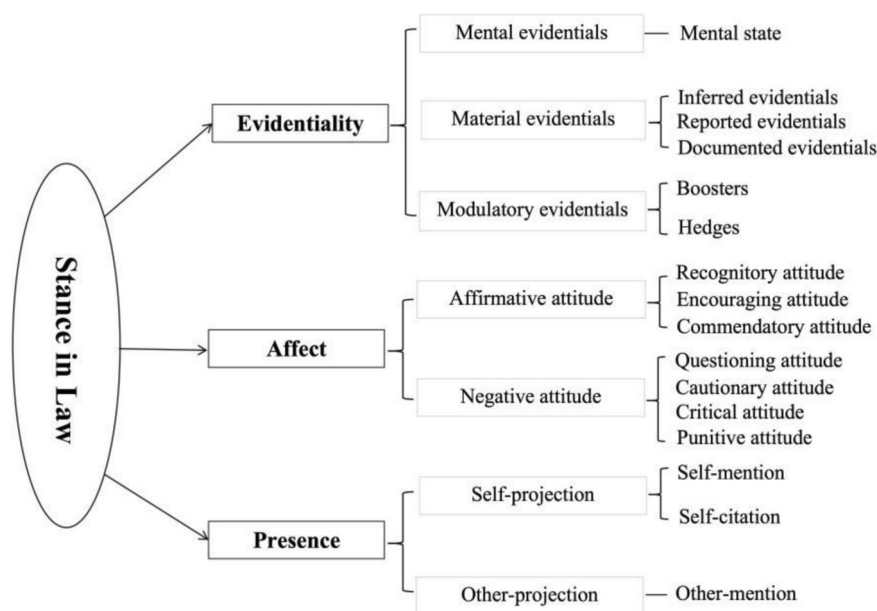
in-depth discussion in this section, thereby exploring and refining the theoretical contributions of this comparative study.

Regarding RQ2, the data protection laws of the U.S., the EU, and China demonstrate two overarching stance orientations. One is the modesty of legislation. Across all three jurisdictions, legislative texts present a prevalence of prudent hedging devices and a scarcity of subjective self-mention devices. This stance orientation seems to be found in various genres of legislative texts. Current exploration of data protection laws reflects the prudence and objectivity that legislators have shown when navigating the complexities of legislating in an emerging domain. The other is the sufficient discursive space left for the data protection judiciary. It is dictated by the distinctive nature of data protection laws compared to other genres of law. In the digital world, various unprecedented legal cases concerning data continuously emerge, such as those related to data scraping (Liu and Chen 2021), which leads to a unique feature that judicial precedents often precede legislative action in data protection regulation. This characteristic is evident in both civil law and common law systems. Essentially, in domains heavily influenced by technology like data protection and intellectual property rights, statutory laws frequently fail to keep pace with societal advancements. Consequently, judicial decisions can provide normative guidance for social behaviour (Kordzadeh and Ghasemaghaei 2022; Pei and Cheng 2022). Given this dynamic, it is imperative for data protection legislation to afford sufficient discursive space for judicial discretion.

Regarding RQ3, the underlying reasons for the distinctions in stance practices among the U.S., the EU, and China stem from their respective legislative values and public ideologies. The voice of the text transmitted by stance expressions mirrors the different national voices in data governance. Specifically, the prevalence of hedges in EULC underscores the EU’s commitment to precision and reasoning in data protection legislation. This inclination can be traced back to the EU’s pioneering role in enacting personal data protection legislation on a global scale, without any existing framework to draw from. Consequently, the EU draws upon its rich historical traditions and legislative experiences concerning personal privacy protection (Li and Kit 2021), shaping the right to the protection of personal data as a fundamental right within its legislative framework. Additionally, the examination of boosters reveals a strong inclination of the EU legislature to convey legislative values. This aligns with the EU’s ambition to promulgate its values and principles of personal data protection globally. Despite its significance, GDPR still exhibits inherent limitations, which China has taken note of and made legislative improvements accordingly. The relatively low proportion of stance items in CLC may suggest a strategic legislative approach. This is because China, in subsequent legislative efforts, recognised the challenges posed by the EU’s stringent standards of personal data protection, which hindered data flow and transactions within the EU. Therefore, China has embraced a comparatively neutral stance in its approach to data protection legislation, seeking to strike a balance between protecting individual data rights and facilitating the utilisation and flow of data (Cheng and Liu 2023). This stance is manifested in the use of covert stance markers aimed at achieving objectivity within the law. Finally, it can be inferred that the legislative stance in the U.S. places importance on national security while also emphasising the protection of personal data and privacy rights, alongside facilitating data utilisation to foster the growth of data-driven economy.

**Stance in law.** This part takes a specific view of the stance in law. Based on the influential work of Hyland (2005) and the close examination of data protection laws, a more detailed and law-





**Fig. 6** The model of stance in law (adapted from Hyland's (2005) stance model).

targeted research model of stance is proposed, as shown in Fig. 6. Through a comprehensive and detailed analysis of data protection legislative corpora, the judgments of the landmark *Schrems* case (Flórez Rojas 2016), which mirrors a principled stance on data protection (Zalnieriute 2022), and supplementary relevant judgments, this study sub-categorises each main type of stance into different sub-types depending on the semantic meanings of particular markers used. These sub-types of stance devices are particularly suitable for stance identification and analysis in legal texts.

Evidentiality in law is important because “fact is not found but reconstructed based on the admissible evidence” (Cheng and Cheng 2014) especially in court judgments (Wu and Cheng 2020). As a kind of writer's commitment to the reliability of the propositions, evidentiality involves not only the degree of certainty but also the source of the propositional information (Aikhenvald 2004). Hence, on one hand, the balance of boosters and hedges in the text indicates a modulatory strategy of the writer to convey a commitment to the text content; on the other hand, the division of mental and material evidentials is based on the source of the propositional information. Elucidating further, mental or mental-state evidentials refer to the evidence or information acquired through one's personal opinion or beliefs. For example, *I think* is the most typical and subjective way to state one's opinion. Material evidentials, contrary to mental ones, are used when the speaker or writer has not personally perceived the information but has acquired it from others or other things. Legally, mental evidentials, pertaining to the cognitive domain, are exempt from evidential burden as thoughts are inherently unverifiable, which reflect a subjective orientation of stance. Conversely, material evidentials often need linguistic evidence as support, which inclines towards a more objective stance. There are three main sub-types of material evidentials: inferred, reported and documented. Inferred evidentials refer to the reasoning process and denote its reliability in view of the quality of the available evidence (Aikhenvald 2004). Therefore, the use of an inferred evidential allows the writer to explicitly qualify their statement and convey to the audience the extent to which the proposition can be considered well-founded or tentative (Kwon 2012). Reported evidentials refer to statements that are based on information reported by others rather than direct observation or

first-hand experience by the speaker or the writer (Clift 2006). The use of reported evidentials allows the recipient to properly evaluate the credibility of the statement based on the original source of information. Documented evidentials denote the information based on an official documented form, such as legislation, legal contracts or any other form of documentation that can be examined and referenced within the legal proceeding (Mithun 2020). In legal judgments, documented evidentials are generally given greater weight and credibility than reported evidentials, as they are considered more reliable and less susceptible to potential biases that can arise from the reporting process. Examples of sub-types of evidentiality extracted from the judgments are shown in Fig. 7.

Affect in law, especially in judicial practice, has been widely explored in the area of law and psychology or sociology of law (e.g., Rossmanith 2015; Maroney 2011; Amaya and Del Mar 2020). For instance, the act of drafting a legal judgment is an exercise of attitudinal persuasion (Anleu and Mack 2021), either advocating for or against a particular stance. By expressing an affirmative or negative attitude, it aims to convince the parties involved, as well as other readers, of the validity and correctness of the court's decision (Baum 2017; Barak 2009). Affirmative attitudes, including recognitory, encouraging, and commendatory attitudes, generally denote positive feelings towards legal actions, which highlights their validity, effectiveness, or compliance with legal standards. Negative attitudes, including questioning, cautionary, critical, and punitive attitudes, indicate negative feelings towards legal actions, which identify shortcomings, violations, or areas where the actions fail to meet legal standards. Therefore, investigating the affect in law is vital for a comprehensive understanding of the attitudinal factors that influence lawmakers and judges. It helps in identifying the motivations, orientations and values of the people who made these affective evaluations. Figure 8 shows examples of sub-types of affect extracted from the judgments.

“All writing carries information about the author” (Hyland 2018: 62), and legal texts are no exception. In legal texts or documents, writers make a conscious choice about the presence or absence of the author to adopt a particular stance and establish a contextually situated authorial identity. When engaging in the act of writing, writers inevitably project their own subjectivity and positionality within the work. Furthermore, the writer's stance

Evidentiality	Mental evidentials	Mental state	"The defendant claimed, ' <i>I think</i> I had the legal right to take that action'."
	Material evidentials	Inferred evidentials	"the referring court asks whether the SCC Decision is capable of ensuring an adequate level of protection of the personal data transferred to third countries <i>given that</i> the standard data protection clauses provided for in that decision do not bind the supervisory authorities of those third countries."
		Reported evidentials	"As <i>reported</i> by the eyewitness, the altercation began shortly after the two individuals exchanged heated words."
		Documented evidentials	"According to Article 46(2)(c) of the GDPR,"
	Modulatory evidentials	Boosters	"as is <i>clear</i> from paragraph 125 above,"
		Hedges	"there are others in which the content of those standard clauses <i>might</i> not constitute a sufficient means of ensuring,"

Fig. 7 Examples of different types of "Evidentiality".

Affect	Affirmative attitude	Recognitory attitude	"The court <i>acknowledges</i> that the defendant has shown genuine remorse for his actions"
		Encouraging attitude	"The court <i>encourages</i> the parties to consider mediation as a means to resolve their disputes amicably."
		Commendatory attitude	"The court is <i>impressed</i> by the innovative measures taken by the company to ensure environmental sustainability."
	Negative attitude	Questioning attitude	"the Commission observed that ' <i>concerns</i> about the level of protection of personal data of [Union] citizens transferred to the [United States] under the Safe Harbour scheme have grown"
		Cautionary attitude	"The court is <i>mindful</i> of the potential for bias in the witness's testimony."
		Critical attitude	"a number of <i>weaknesses</i> in the application of Decision 2000/520... structural <i>shortcomings</i> related to transparency and enforcement"
		Punitive attitude	"effective powers of intervention such as that of imposing a temporary or definitive <i>ban</i> on processing of data"

Fig. 8 Examples of different types of "Affect".

Presence	Self-projection	Self-mention	"On those grounds, <i>the Court</i> (Grand Chamber) hereby rules:"
		Self-citation	" <i>Consistent with our ruling</i> in Adams v. Williams, 556 U.S. 312 (2011), this court maintains that the exclusionary rule is essential for safeguarding Fourth Amendment rights."
	Other-projection	Other-mention	"On 25 June 2013 <i>Mr Schrems</i> made a complaint to the Commissioner by which he in essence asked the latter to exercise his statutory powers by prohibiting Facebook Ireland from transferring his personal data to the United States."

Fig. 9 Examples of different types of "Presence".

and orientation towards others will also be woven into the fabric of the text. Hence, presence in legal discourse can be divided into two types according to the distinction of writers and other actors: self-projection and other-projection. Self-projection indicates the personal projection of the writers as a powerful means of self-representation. It usually has two ways of expression, namely, self-mention and self-citation. Other-projection, on the other hand, indicates referencing actions or statements made by other parties involved in the legal process to show the writer's stance. For instance, a judge might refer to a plaintiff's claims to evaluate their credibility, or to a witness's testimony to support or refute a point of law. Distinguishing between self-projection and other-projection helps to analyse how legal writers position themselves and others within legal discourse. Figure 9 shows examples of sub-types of presence extracted from the judgments and the corpora in this study.

On balance, this study makes three main contributions at the intersection of stance and legal discourse. Firstly, compared to other genres of law, stance expressions of data protection laws show a unique sense of legal constructiveness. In this context, legal constructiveness is reflected in the use of stance markers and other linguistic markers in order to achieve the establishment and reinforcement of rules and principles for emerging laws such as the data protection legislation being explored in our study. As for the traditional or long-standing legislation, this type of legal constructiveness through stance expression may not be as pronounced but has developed a relatively stable form. The study of stance markers in marriage law (Tracy 2011) and defamation law (Izes 2023) serves as evidence that researchers usually view legislative stance a consistent benchmark for studying stance in judicial opinions or oral arguments. The second contribution lies in the alignment of stance construction patterns in data protection legislation with

broad legislative discourse. Contrary to the anticipated prevalence of overt stance markers in legal discourse, such as the extensive use of boosters in judicial dissenting (Boginskaya 2022), the analysis reveals a higher frequency of hedging devices in legislative discourse on data protection. It reflects that compared to the explicit and overt stance expressions, the legislative discourse on data protection prioritises covert ones, such as negative-oriented attitude markers, with the aim of upholding prudence, authority and objectivity in legislation. Lastly, the present study formulates a specialised research model of stance in law based on Hyland's foundational work on stance, which provides a more targeted way to understand how stance is conveyed in legal texts. The enriched sub-categories shed light on how legal professionals express the subjective, objective, positive, negative, explicit, implicit and other more nuanced orientations of stance in their writing.

### Concluding remarks

To sum up, this study set out to explore the patterns of stance representation and construction in data protection laws, while also re-contextualising the underlying strategic ideological expressions of lawmakers within the socio-legal framework. In this study, we adopted a corpus-driven approach to investigate stance expressions in data protection legislation across the U.S., the EU, and China. Drawing upon Hyland's framework on stance, we derived valuable insights into various dimensions of stance. At a micro-level, the findings of this study reveal both commonalities and distinctions of stance expression within data protection legislation across three jurisdictions, which provide a relatively comprehensive understanding of their stance construction patterns. From a macro-level perspective, these different stance choices also mirror distinct text-external factors like legal cultures (Wagner et al. 2020), values, and ideologies. Notably, the research has also shown that stance tends to be conveyed implicitly rather than through explicit grammatical devices in legislative discourse on data protection.

The insights gained from this study highlight the important role of stance in legal discourse. Stance, as a series of linguistic rhetorical choices and a type of legal value orientation, enables legislators to intrude into texts explicitly or implicitly to convey legal reasoning and ideologies, reconcile conflicting interests and demands among stakeholders, and ensure the objectivity and fairness of the law. Simultaneously, for the target audience of legislative texts, unravelling the stance embedded in the law facilitates a more precise grasp of legislative intent, thereby aiding in a more effective understanding and application of legal protections for their rights and interests.

Overall, this study serves as a stepping stone for future explorations in identifying stance in legislative discourse and pushing the boundaries of our knowledge in stance theory. Future works are encouraged to build upon our findings and explore the rich territories of stance representation within the data protection judicial domain, especially within court judgments or courtroom discourse.

### Data availability

The datasets analysed during the current study are available from the corresponding author upon reasonable request.

Received: 17 February 2024; Accepted: 10 June 2024;

Published online: 21 June 2024

### References

- Acquisti A, Taylor C, Wagman L (2016) The economics of privacy. *J Econ Lit* 54(2):442–492
- Aikhenvald A Y (2004) *Evidentiality*. Oxford University Press, Oxford

- Amaya A, Del Mar M (2020) *Virtue, emotion and imagination in law and legal reasoning*. Hart Publishing, Oxford, New York
- Anleu SR, Mack K (2021) *Judging and emotion: a socio-legal analysis*. Routledge, London, New York
- Appel O, Chiclana F, Carter J, Fujita H (2016) A hybrid approach to the sentiment analysis problem at the sentence level. *Knowl Based Syst* 108:110–124
- Aull LL, Bandarage D, Miller MR (2017) Generality in student and expert epistemic stance: a corpus analysis of first-year, upper-level, and published academic writing. *J Engl Acad Purp* 26:29–41
- Aull LL, Lancaster Z (2014) Linguistic markers of stance in early and advanced academic writing: a corpus-based comparison. *Writ Commun* 31(2):151–183
- Baker P (2010) *Sociolinguistics and corpus linguistics*. Edinburgh University Press, Edinburgh
- Barak A (2009) *The Judge in a Democracy*. Princeton University Press, Princeton
- Baum L (2017) Judges and Their Audiences. In: Epstein L, Lindquist SA (eds) *The Oxford Handbook of US Judicial Behavior*, Oxford University Press, Oxford, New York, pp 343–360
- Bednarek M (2006) *Evaluation in media discourse: analysis of a newspaper corpus*. Continuum, London
- Bennett CJ (1988) Different processes, one result: the convergence of data protection policy in Europe and the United States. *Governance* 1(4):415–441
- Biber D (2006) *University language: a corpus-based study of spoken and written registers*. John Benjamins, The Netherlands
- Biber D, Finnegan R (1989) Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text* 9(1):93–124
- Biber D, Johansson S, Leech G, Conrad S, Finegan E (2000) *Longman grammar of spoken and written English*. Longman, New York
- Boginskaya O (2022) Dissenting with conviction: boosting in challenging the majority opinion. *Int J Leg Discours* 7(2):257–279
- Breeze R (2011) Disciplinary values in legal discourse. *Ibérica* 21:93–115
- Brezina V, McEnery T, Wattam S (2015) Collocations in context: a new perspective on collocation networks. *Int J Corpus Linguist* 20(2):139–173
- Chaemsaitong K (2015) Stance expressions in the courtroom. *Engl Lang Linguist* 21(2):41–59
- Chaemsaitong K (2017) Evaluative stancetaking in courtroom opening statements. *Folia Linguist* 51(1):103–132
- Chan THT (2015) A corpus-based study of the expression of stance in dissertation acknowledgements. *J Engl Acad Purp* 20:176–191
- Chand V (2011) Elite positionings towards Hindi: Language policies, political stances and language competence in India. *J Socioling* 15(1):6–35
- Charles M (2007) Argument or evidence? Disciplinary variation in the use of the Noun that pattern in stance construction. *Engl Specif Purp* 26(2):203–218
- Cheng L, Cheng W (2012) Legal interpretation: meaning as social construction. *Semiotica* 192:427–448
- Cheng L, Liu X (2022) Exploring Chinese cyber discourse: integrating political and legal perspectives. *Int J Leg Discours* 7(1):33–52
- Cheng L, Liu X (2023) From principles to practices: the intertextual interaction between AI ethical and legal discourses. *Int J Leg Discours* 8(1):31–52
- Cheng L, Liu X (2024) Unravelling power of the unseen: towards an interdisciplinary synthesis of generative AI regulation. *Int J Digit Law Gov* 1(1):29–51
- Cheng L, Qiu J, Yang Y (2023) Constructing cybersecurity discourse via deconstructing legislation. *Int J Leg Discours* 8(2):273–297
- Cheng L, Sin KK (2011) A sociosemiotic interpretation of linguistic modality in legal settings. *Semiotica* 185:123–146
- Cheng L, Xu M, Chang CY (2023) Exploring network content ecosystem evaluation model based on Chinese judicial discourse of digital platform. *Int J Leg Discours* 8(2):199–224
- Cheng W, Cheng L (2014) Epistemic modality in court judgments: a corpus-driven comparison of civil cases in Hong Kong and Scotland. *Engl Specif Purp* 33:15–26
- Clift R (2006) Indexing stance: reported speech as an interactional evidential. *J Socioling* 10(5):569–595
- Conrad S, Biber D (2000) Adverbial marking of stance in speech and writing. In: Hunston S, Thompson G (eds) *Evaluation in text*. Oxford University Press, Oxford, pp 56–73
- Crosthwaite P, Cheung L, Jiang FK (2017) Writing with attitude: stance expression in learner and professional dentistry research reports. *Engl Specif Purp* 46:107–123
- Crosthwaite P, Jiang K (2017) Does EAP affect written L2 academic stance? A longitudinal learner corpus study. *System* 69:92–107
- De Hert P, Papakonstantinou V (2016) The new General Data Protection Regulation: still a sound system for the protection of individuals? *Comput Law Secur Rev* 32(2):179–194
- Delandshere G, Petrosky A (2004) Political rationales and ideological stances of the standards-based reform of teacher education in the US. *Teach Teach Educ* 20(1):1–15
- Du Bois JW (2002) Stance and intersubjectivity in dialogic interaction. In: *Linguistics colloquium*, University of California at Santa Barbara



- Du Bois J (2007) The stance triangle. In: Englebretson R (ed) *Stancetaking in discourse: subjectivity, evaluation, interaction*. John Benjamins Publishing, Amsterdam, pp 138–182
- Flórez Rojas ML (2016) Legal implications after Schrems case: are we trading fundamental rights? *Inf Commun Technol Law* 25(3):292–309
- Fuster GG, Gutwirth S (2013) Opening up personal data protection: a conceptual controversy. *Comput Law Secur Rev* 29(5):531–539
- Gales T (2010) Ideologies of violence: a corpus and discourse analytic approach to stance in threatening communications. *Int J Speech Lang Law* 17(2):299–302
- González G, Van Brakel R, De Hert P (2022) *Research handbook on privacy and data protection law: values, norms and global politics*. Edward Elgar Publishing, Cheltenham, Northampton
- Grabe W (1984) *Towards defining expository prose within a theory of text construction*. Unpublished doctoral dissertation, University of Southern California
- Greco S, Palmieri R, Rigotti E (2016) Institutional argumentation and conflict prevention: the case of the Swiss Federal Data Protection and Information Commissioner. *J Pragmat* 105:39–53
- Halliday MAK (2004) The language of science. J Webster (ed). 5th volume of a of the Collected Works of M.A.K. Halliday. Continuum, London, New York
- Hoofnagle CJ, Van Der Sloot B, Borgesius FZ (2019) The European Union general data protection regulation: what it is and what it means. *Inf Commun Technol Law* 28(1):65–98
- Hu M, Hu X, Cheng L (2021) Exploring digital economy: a sociosemiotic perspective. *Int J Leg Discours* 6(2):181–202
- Hunston S (2002) *Corpora in applied linguistics*. Cambridge University Press, Cambridge
- Hunston S, Thompson G (2000) *Evaluation in text*. Oxford University Press, Oxford
- Hyland K (1999) Disciplinary discourses: writer stance in research articles. In: Candlin C, Hyland K (eds) *Writing: texts, processes and practices*. Longman, London, pp 99–121
- Hyland K (2005) Stance and engagement: a model of interaction in academic discourse. *Discourse Stud* 7(2):173–192
- Hyland K (2012) *Disciplinary identities: individuality and community in academic discourse*. Cambridge University Press, Cambridge
- Hyland K (2018) *Metadiscourse: exploring interaction in writing*. Bloomsbury Publishing, London, New York
- Hyland K, Jiang F (2016) Change of attitude? A diachronic study of stance. *Writ Commun* 33(3):251–274
- Hyland K, Tse P (2005) Hooking the reader: a corpus study of evaluative that in abstracts. *Engl Specif Purp* 24(2):123–139
- Izes A (2023) Fact versus opinion in US Defamation Law: a corpus and appraisal analysis of speaker stance toward reputational harm. *Int J Semiot Law* 36(3):1–32
- Kordzadeh N, Ghasemaghaei M (2022) Algorithmic bias: review, synthesis, and future research directions. *Eur J Inf Syst* 31(3):388–409
- Kwon I (2012) Please confirm what I inferred: on the Korean inferential-evidential marker-napo. *J Pragmat* 44(8):958–969
- Li J, Cai X, Cheng L (2023) Legal regulation of generative AI: a multidimensional construction. *Int J Leg Discours* 8(2):365–388
- Li J, Cheng L, Cheng W (2016) Deontic meaning making in legislative discourse. *Semiotica* 209:323–340
- Li S, Kit C (2021) Legislative discourse of digital governance: a corpus-driven comparative study of laws in the European Union and China. *Int J Leg Discours* 6(2):349–379
- Liu W, Chen P (2021) Justification of the behavior regulatory pattern on data scraping. *Comput Law Secur Rev* 43:1–15
- Maroney TA (2011) Emotional regulation and judicial behavior. *Calif Law Rev* 99:1485–1556
- Martin JR (2000) Beyond exchange: APPRAISAL systems in English. In: Hunston S, Thompson G (eds) *Evaluation in text*. Oxford University Press, Oxford, pp 142–175
- Martin J R, White P R (2005) *Language of evaluation: appraisal in English*. Palgrave Macmillan, London
- Mattioli V, McAuliffe K (2021) A corpus-based study on opinions of advocates general of the court of justice of the European Union: changes in language and style. *Int J Leg Discours* 6(1):87–111
- McEnery T, Hardie A (2012) *Corpus linguistics: method, theory and practice*. Cambridge University Press, Cambridge
- McEnery T, Wilson A (2001) *Corpus linguistics*, 2nd edn. Edinburgh University Press, Edinburgh
- McKeown J (2022) Stancetaking in the US Supreme Court's abortion jurisprudence (1973-present): epistemic (im) probability and evidential (dis) belief. *Int J Leg Discours* 7(2):323–343
- Mithun M (2020) Context and consciousness: documenting evidentials. *Folia Linguist* 54(2):317–342
- Myers C (2014) The new water cooler: implications for practitioners concerning the NLRB's stance on social media and workers' rights. *Public Relat Rev* 40(3):547–555
- Nir B, Dori-Hacohen G, Maschler Y (2014) Formulations on Israeli political talk radio: from actions and sequences to stance via dialogic resonance. *Discourse Stud* 16(4):534–571
- Paler L (2005) China's legislation law and the making of a more orderly and representative legislative system. *China Q* 182:301–318
- Pei J, Cheng L (2020) Mapping terminological variation and ideology in data protection laws. *Terminology* 26(2):159–183
- Pei J, Cheng L (2022) Deciphering emoji variation in courts: a social semiotic perspective. *Hum Soc Sci Commun* 9(1):1–8
- Pernot-Leplay E (2020) China's approach on data privacy law: a third way between the US and the EU? *Penn St J Int'l Aff* 8(1):49–117
- Pho PD (2008) Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Stud* 10(2):231–250
- Poole R (2021) A corpus-aided study of stance adverbs in judicial opinions and the implications for English for Legal Purposes instruction. *Engl Specif Purp* 62:117–127
- Rezaei S, Kuhi D, Saeidi M (2021) Diachronic corpus analysis of stance markers in research articles: the field of applied linguistics. *Cogent Art Humanit* 8(1):1–12
- Romanou A (2018) The necessity of the implementation of Privacy by Design in sectors where data protection concerns arise. *Comput Law Secur Rev* 34(1):99–110
- Rossmanith K (2015) Affect and the judicial assessment of offenders: feeling and judging remorse. *Body Soc* 21(2):167–193
- Sadowski J (2019) When data is capital: datafication, accumulation, and extraction. *Big Data Soc* 6(1):1–12
- Scott M, Tribble C (2006) *Textual patterns: key words and corpus analysis in language education*, vol 22. John Benjamins Publishing, Amsterdam, Philadelphia
- Sinclair J (1991) *Corpus, concordance, collocation*. Oxford University Press, Oxford
- Tamburri DA (2020) Design principles for the General Data Protection Regulation (GDPR): a formal concept analysis and its evaluation. *Inf Syst* 91:1–14
- Tognini-Bonelli E (2001) *Corpus linguistics at work*. John Benjamins, Amsterdam
- Tracy K (2011) What's in a name? Stance markers in oral argument about marriage laws. *Discourse Commun* 5(1):65–88
- Vass H (2017) Lexical verb hedging in legal discourse: the case of law journal articles and Supreme Court majority and dissenting opinions. *Engl Specif Purp* 48:17–31
- Vertommen B, Vandendaele A, Van Praet E (2012) Towards a multidimensional approach to journalistic stance. Analyzing foreign media coverage of Belgium. *Discourse Context Media* 1(2-3):123–134
- Vis K, Sanders J, Spooren W (2012) Diachronic changes in subjectivity and stance—a corpus linguistic study of Dutch news texts. *Discourse Context Media* 1(2-3):95–102
- Wagner A (2024) Rhizomatic healthcare: the digital transformation of France's medical landscape. *Int J Digit Law Gov* 1(1):79–89
- Wagner A, Matulewska A, Cheng L (2020) Law as a culturally constituted sign-system—a space for interpretation. *Int J Leg Discours* 5(2):239–267
- Wang C, Cheng L, Pei J (2020) Exploring the cyber governance discourse: a perspective from China. *Int J Leg Discours* 5(1):1–15
- Wu J, Cheng L (2020) Evidentiality of court judgments in the People's Republic of China: a semiotic perspective. *Semiotica* 236–237:477–500
- Wu X (2018) Framing, reframing and the transformation of stance in news translation: a case study of the translation of news on the China–Japan dispute. *Lang Intercult Commun* 18(2):257–274
- Yu W (2023) Negotiation of justice: the discursive construction of attitudinal positioning in bilingual legal judgments of HKSAR v KWAN WAN KI. *Int J Leg Discours* 8(2):299–333
- Zalnierute M (2022) Data transfers after schrems II: the EU-US disagreements over data privacy and national security. *Vand J Transnatl Law* 55(1):1–48
- Zhang M (2018) Stance and mediation in transediting news headlines as paratexts. In: Valdeón RA (ed) *Chinese translation studies in the 21st century*. Routledge, London, pp 96–111
- Zhao B, Feng Y (2021) Mapping the development of China's data protection law: major actors, core values, and shifting power relations. *Comput Law Secur Rev* 40:1–16
- Zhao Y, Cheng L (2024) A bibliometric study of research trends in cross-border cybercrime. *Int J Leg Discours* 9(1):1–31

## Acknowledgements

This work was supported by the major project of the National Social Science Foundation of China under Grant 20ZDA062 and the Chinese University Foundation.

## Author contributions

XL designed the work, conducted data analysis, and wrote the initial draft of this article. LC provided valuable feedback and critically revised the manuscript. CS offered insightful suggestions during the revision process. All authors contributed to the article.

## Competing interests

The authors declare no competing interests.

## Ethical approval

This article does not contain any studies requiring ethical approval.

## Informed consent

This article does not contain any studies with human participants performed by any of the authors.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-024-03322-9>.

**Correspondence** and requests for materials should be addressed to Chunlei Si.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024