Detection of Conflicts, Contradictions and Inconsistencies in Regulatory Documents: A Literature Review

1st Gerrit Schumann
University of Oldenburg
Dept. of Business Informatics
Oldenburg, Germany

2nd Jorge Marx Gómez
University of Oldenburg
Dept. of Business Informatics
Oldenburg, Germany

***Abstract*—Companies create regulatory documents, such as policies, standards, and guidelines, to define their processes and structures. These documents are used by employees to ensure the proper execution of processes and by auditors to check the actual compliance with the respective requirements. Especially in large organizations, these documents are constantly updated, which can lead to conflicts, inconsistencies, or contradictions between these documents, which in turn can contribute to errors and delays, the compromise of assets, non-compliance or the facilitation of fraud. Given the large number of different regulatory documents as well as the different interpretations of conflicts and the technical approaches to detect them, this paper aims to provide an overview of previous research in this field. To this end, we conducted a structured literature review based on 46 publications to analyze various aspects, such as the types, domains, language, and structure of the analyzed regulatory documents as well as the types of conflicts and the technical approaches to detect them. In this way, we have shown the current state-of-the-art, identified potential challenges, and highlighted possible research directions.**

Keywords — Conflicts, Contradictions, Inconsistencies, Regulatory Documents, Literature Review

I. INTRODUCTION

In order to define and describe the target state of their processes and structures, companies create various regulatory documents, such as policies, guidelines, rules of conduct, workflow descriptions, operating procedures, manuals, work and process standards or work and service instructions. These documents are used by employees to plan and carry out a process in accordance with the respective rules, and by auditors to familiarize themselves with the respective target state of an object to be audited. In this way, they are able to identify deviations between the target and actual state of the audit object [1][2][3].

In large companies and corporate groups, there is a particularly high number of regulatory documents, which are also subject to constant change due to different company and market developments as well as external legal reasons [4][5][6]. For example, after the merger of two companies, the data protection guidelines, communication guidelines and process standards must be revised to ensure a uniform level of security across the newly combined networks. However, such regulatory documents are not necessarily written and provided by a central authority, but also by actors from different departments [7], which in turn can lead to certain contents being contradictory. Such conflicts in policies and procedures can create confusion among employees toward the right course of action, which could lead to errors or delays in performing tasks and have a cascading effect on the overall business. Moreover, contradictions can have a negative impact on the effectiveness of internal auditing in two ways: If the contradictions are identified by the auditors in advance, this leads to clarification work, which in turn results in longer audit preparation. If, on the other hand, the contradictions are not identified, an audit may be carried out based on inconsistent target specifications, which can lead to several internal and external risks for the company.

Unlike structured data, such as spreadsheets, csv-files or relational databases, regulatory documents cannot be processed by traditional auditing data analysis methods due to their natural language, textual nature. Ensuring consistency and the absence of relevant changes between these documents is therefore often only possible in traditional compliance processes by manually comparing the textual contents [9]. However, with respect to several hundred regulatory documents, which in turn may consist of a large number of pages, sections, subsections, and paragraphs [10], such manual reconciliation is error-prone, extremely time-consuming, and still not scalable to the total number of all documents. Especially because the number of regulatory documents is constantly increasing [11].

To overcome the manual comparison of natural language regulatory texts, various automated approaches have been developed in the last decade. In view of the large number of different regulatory documents with which employees and auditors are confronted, as well as the different interpretations of conflicts and the technical approaches to detecting them, this paper aims to provide an overview of the research conducted to date. Based on this objective, our study extends existing literature reviews that focused on a specific notation of business rules [6] with a holistic view of all forms of natural language regulatory texts. In this way, our paper generates the following contributions:

- Presentation of the state-of-the-art in the detection of conflicts, contradictions, and inconsistencies in regulatory documents.
- Identification of current research priorities and gaps based on the analysis of 9 categories (including the types, domains, language, and form of the documents, as well as the types of conflicts, the method of content comparison and the technical approaches to conflict detection).

This paper is structured as follows: Section II provides a brief overview of different types of regulatory documents and their categorization, as well as an attempted definition of regulatory conflicts. Section III presents the review methodology and the results of the individual evaluation categories. Section IV summarizes the findings of the literature review and discusses them in terms of potential challenges and future research directions.

II. Fundamentals

A. Regulatory Documents

Regulatory documents (also known as *Normative Documents* [27]) can be interpreted as internal or external in terms of their requirements. *External regulatory documents* are official documents created by external authorities, government agencies or international organizations to regulate compliance and standards that companies and organizations must adhere to. These include, for example, laws, regulations, standards and guidelines that define how organizations should operate in certain areas such as environmental protection, finance, data protection and occupational safety. Companies must comply with these external requirements in order to avoid legal sanctions and retain their operating licenses. *Internal regulatory documents*, on the other hand, are documents created by the company itself that aim to standardize internal procedures and behaviors. These can include handbooks, operating instructions, quality management documents and other internal policies that set out specific procedures and rules of conduct for employees. They are critical to day-to-day management by providing clear instructions and helping to ensure organizational efficiency and compliance with internal standards. At the same time, these documents typically also reflect many aspects that have been specified in external regulatory documents [8]. A detailed overview of the different types of regulatory documents identified in this literature review is provided in chapter III.B.2.

B. Conflicts, Inconsistencies and Contradictions

There are various interpretations in the literature according to which two normative statements are regarded as conflicting. Moreover, *"Conflict"* is not a universally valid term in this field of research. Instead, it is often used as a synonym for *Contradiction* or *Inconsistency* and in rare cases also for *Anomaly*. Since the normative statements made in the regulatory documents can often be assigned to deontic logic, a common interpretation approach is that conflicts arise when *obligations*, *prohibitions* or *permissions* are mutually exclusive and their simultaneous compliance is therefore impossible [56]. A simple example of such mutual exclusion is illustrated in the following two sentences by comparing a permission with a prohibition.

Sentence A: *Smoking is permitted on the company premises in the designated areas C and D.*

(Permission)

Sentence B: *Since 01.01.2024, there has been an unrestricted smoking ban on the entire company premises.*

(Prohibition)

A detailed overview of the different interpretations of regulatory conflicts or inconsistencies identified in this literature review can be found in Chapter III.B.7.

III. Literature Review Methodology

A. Literature Selection and Filtering

In order to find potentially relevant papers, we searched the six databases "ScienceDirect", "ACM", "IEEE Xplore", "Web of Science", "Scopus" and "Google Scholar" using the search query shown in Fig. 1. On the one hand, this query covers the common terms for regulatory documents (e.g., *"instructions", "guidelines", "policies", "standards",* etc.) and, on the other hand, the terms typically used for the detection of conflicting statements within these documents (e.g., *"contradiction", "conflict", "inconsistent"* + *"detection", "identification"*, etc.). However, since the latter can also be carried out with the help of the NLP research field "Natural Language Inference" (also known as "Recognizing Textual Entailment"), we have also included these two terms in the search query. For a match, all terms had to appear either in the title, in the abstract or in the keywords of the respective publication.

As a result of this search, a total of 2,318 papers were found, which we reduced by 383 duplicates to 1,935 unique papers. We then removed all papers that either (1) understand the respective characteristic document term in a completely different context, (2) provide approaches that are not suitable for natural or controlled natural language, but e.g. for system rules that are exclusively read and processed by machines, (3) perform the actual detection of the inconsistencies / contradictions manually, (4) describe purely theoretical frameworks or approaches, or (5) are not written in English (Fig. 1).

After applying these filter criteria, 37 relevant papers remained, for which we then carried out a manual forward and backward search. Considering the five filter criteria described above, this resulted in 9 further studies, which increased the final number of relevant papers to 46 (Fig. 1).

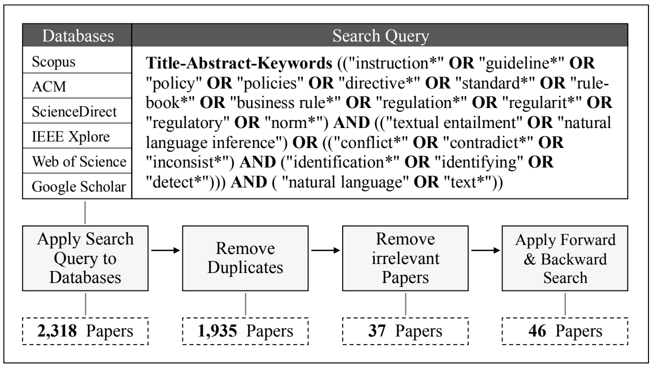

Fig. 1. Literature Selection and Filtering

B. Literature Analysis

As part of the actual literature analysis, we categorized the 46 relevant papers according to both methodological and data-specific aspects. The data-specific aspects include (B.2) the types of the respective regulatory documents (e.g., contracts, standards, etc. + domain), (B.3) the language and form of the textual content (controlled or plain natural language) as well as (B.9) the number and origin of the contained inconsistencies or contradictions (synthetic or real). The methodological aspects include (B.4) the procedure for selecting the contents to be analyzed (e.g. selection of sentences containing modal verbs), (B.5) the procedure for sentence pairing (e.g. based on matching subjects or actions), (B.6) the scope of the sentence comparisons (e.g. within the same document or across different document types), (B.7) the interpretation of the respective conflicts, inconsistencies or contradictions (e.g.

by mutually exclusive concepts of deontic logic), (B.8) the size of the context window (e.g. comparison of individual sentences or additional consideration of surrounding sentences) as well as (B.10) the technical approach for detection (e.g. rule-based, machine-learning-based, etc.). Please note that a paper can appear in several categories at the same time and that we reference a maximum of three papers per category due to the often high number of papers in certain categories.

1) Publications per Year

The first papers dedicated to the automated detection of conflicts, inconsistencies or contradictions in regulatory documents were published in 2009. From this point onwards, the number of publications mainly fluctuated between two and five per year, reaching a maximum of seven publications in 2018 and 2021 (Fig. 2).

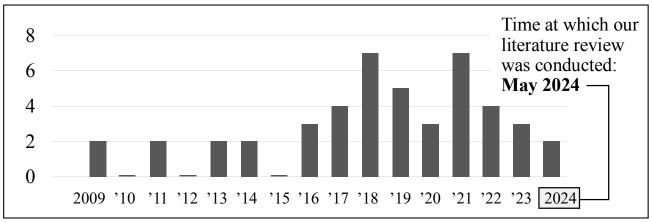

Fig. 2. Publications per Year

2) Types and Domains of the Regulatory Documents

The studies essentially focused on seven different types of regulatory documents (Fig. 3). In this context, it is interesting to note that although contracts were not explicitly mentioned in the search query, they represent the most common type of regulatory document (e.g., [12]-[30]). This is due to the fact that contracts contain normative expressions or "norms", which in turn was covered by our search query (see Fig. 1: "norm*"). Norms typically describe obligations, permissions, or prohibitions, which makes them relevant for this literature review if they are in conflict with each other. The second most frequently used documents were policies / guidelines (e.g., [11][42]-[47]), followed by business rules (e.g., [5][36]-[41]), legal provisions / legislations (e.g., [31]-[35]), standards (e.g. [50][51]) and operating procedures [47][49]. Other normative requirement specifications [35] were only used once.

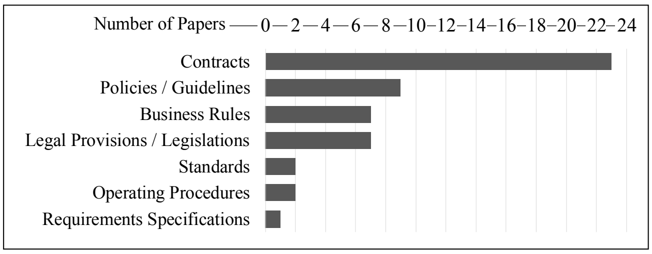

Fig. 3. Types of Regulatory Documents

The regulatory documents shown in Figure 3 cover a total of 17 domains (Fig. 4). In addition to a frequently used workflow description from the car rental industry (e.g., [5][38][40]) and a check-in example from an airline (e.g., [13]-[16]), the most frequently used regulatory documents were either from the medical field (e.g., [44]-[47][49]) or covered several contractual topics (e.g., [21][23][25]).

Other recurring domains were industrial insurance applications (e.g., [38]-[40]), information technology [19][30][48], manufacturing [21][22][26] as well as finance [9][11][53]. Regulatory documents on building requirements [33][41], data privacy / protection / security [8][34] and sale / purchase [29][54] were used less frequently and only one study each addressed, for example, the areas of safety [50] and tax [35].

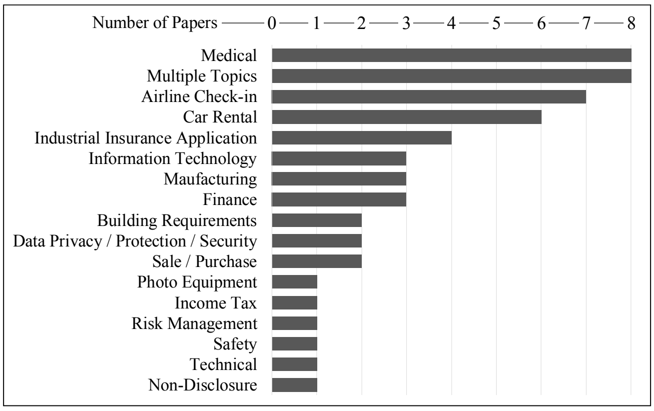

Fig. 4. Domains of the Regulatory Documents

3) Language and Form of the Textual Content

A large proportion of the approaches developed to detect inconsistencies or contradictions are only suitable for rules in (non-natural) languages that never have to be processed by humans, but only by machines. These approaches are not the focus of our literature review. Normative statements formulated in natural language, on the other hand, can be roughly divided into two categories: Those that are formulated in "Plain Natural Language" and those that are formulated using a "Controlled Natural Language". The latter uses natural language but has a declarative character and a standardized representation that must be adhered to (e.g., "Condition-Subject-Object-Action" or "Subject-Verb-Object": *"The ground crew* (Subject) *is required to open* (Verb) *the desk* (Object)*"*). Such controlled natural languages include, for example, CL (Contract Language) or SBVR (Semantics of Business Vocabulary and Rules), which have been used in studies such as [36]-[40] and [30]. Overall, the controlled natural languages were used in 43% of all studies, 65% of which were manually translated into this form (e.g., [15][19][48]) and 15% automatically (e.g., [37]) (Fig. 5, pie chart). Regardless of the form (controlled or not), the language of the regulatory texts was English, with the exception of three studies in which the documents were written in French [45][46] and Dutch [47].

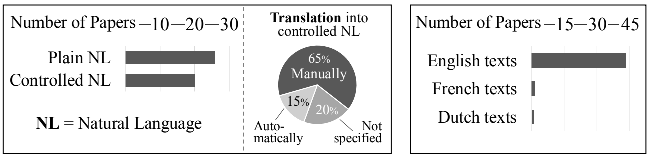

Fig. 5. Form and Language of the Textual Content

4) Procedure for Selecting the Contents to be Analyzed

In order to describe or evaluate their detection approaches, most studies used prefabricated data sets in which all included sentences were relevant for the analysis (e.g., [5][38][40]). In practice, however, inconsistency / contradiction detection is typically not applied to the entire text content of the regulatory documents, but to a specific preselection of sentences that are considered relevant. In the case of the papers examined, this selection was made manually in most cases (e.g., [27][34][46]), for example when the respective sentences refer to a specific topic, such as the 7th Article of the General Data Protection Regulation [34]). Another common approach was the use of a classification model that was specifically trained to identify the relevant sentences (e.g., for the identification of norms within contracts [22][23][26] or rule-sentences in business rules [37]). Just as often as machine learning, rule-based approaches were used which, for example, selected sentences if they contained certain modal verbs, such as: *may, can, must, should, shall* (e.g., [11][21]). On the other hand, there are also studies that did not select sentences, but initially considered all the textual information contained in the respective document for their analysis. This was done, for example, in [47], where all text passages from standard operating procedures (SOPs) were initially considered and then compared with recommendations from medical guidelines if they had a high semantic similarity. In [8], the authors evaluated both an approach in which a signal-word-based selection of sentences from a privacy policy is performed (using modal verbs) and an approach in which all sentences are considered (Fig. 6).

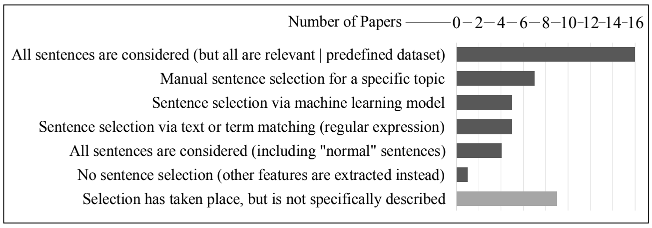

Fig. 6. Procedure for Selecting the Contents to be Analyzed

5) Procedure for Sentence Pairing / Content Comparison

Once the textual contents that are of interest for the analysis have been identified (Section B.4), the next step is typically to determine which of these contents should be compared for the purpose of consistency checking. However, this step strongly depends on whether or not the original natural language texts have previously been converted into a controlled form. If this is the case, tools are typically used for the consistency check that do not use sentence pairing, but a holistic approach in which all extracted rules are taken into account. This approach turned out to be the most common type of comparison and was used, for example, in [37][38] [40], all of which translated the respective rule sentences into corresponding formulas in order to identify inconsistencies in them using an SMT-solver (Satisfiability Modulo Theories, see Section B.10 for more details). The second most frequently used approach was to link sentences into pairs when the similarity of these sentences exceeds a certain threshold (e.g., [8][42][47]). For example, to link sentences from hospital standard operating procedures (SOPs) with recommendations from medical guidelines into candidate pairs, a two-step pairing approach was used in [47], where first the n-gram-based cosine similarity and then the similarity calculated based on sentence embeddings had to reach a certain threshold. Another approach to sentence pairing was to combine sentences that share the same terms or entities. In addition to [32] and [23], this approach was used, for example, in [21], where only those contract norms were compared with each other if they concerned the same parties. Besides such term-based join criteria, there were also studies that created candidate pairs when longer character strings matched. For example, in [53] corresponding pairs were created from two given lists of provisions if a simple *TextDiff* function detects that the provisions differed from each other, or a provision did not occur at all in the other list.

A few other papers created semi-synthetic evaluation data sets by inserting contradictions into conflict-free texts. For example, in [20][24] and [26] a semi-automatic procedure was developed that selected random norm sentences from random contracts, made copies of them and asked a user to change them so that they conflicted with the original norm sentences. In this way, they created annotated sentence pairs, which they then used to train a conflict detection model. In addition to such semi-synthetic candidate pairs, some other studies also created a Cartesian product of all relevant text components and then analyzed it [12][27][34]. For example, in [12] seventeen hypothesis sentences were created and then compared with all sentences of 607 contracts to classify whether the hypotheses are entailed by the contract, contradict it or are not mentioned. Only two studies used other approaches to create candidate pairs, namely with the help of subsuming relationships in ontologies [46] and using clustering [11] (Fig. 7).

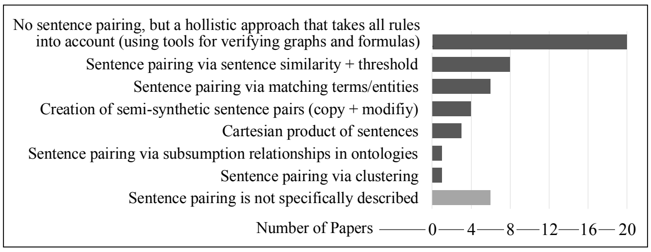

Fig. 7. Procedure for Sentence Pairing / Comparison

6) Scope of the Sentence Comparisons

The comparisons described in Section B.5 were mostly made between information from the same document or between information from different documents of the same type (e.g., [30][37][38]). In contrast, textual information between different document types was compared much less frequently (Fig. 8). These included, for example, comparisons between "Standard Operating Procedures" (SOPs) and "Medical Guidelines" [47], between the "European Union Directive" and "Member State law" [33] or between "Data Protection Guidelines" and the "General Data Protection Regulation" (GDPR) [34].

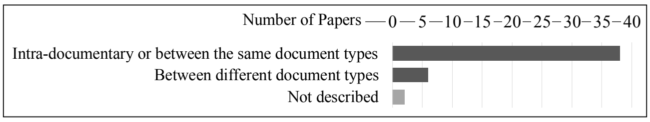

Fig. 8. Scope of Information Comparison

7) Interpretation of the Conflicts

Whether two or more normative statements are in conflict with each other depends on how such a conflict is defined. The majority of studies are based on deontic logic and the associated possibilities for the emergence of a conflict, according to which either obligations and permissions, permissions and prohibitions and obligations and prohibitions are mutually exclusive (e.g., [21]-[23][54]). Other studies extended these three cases with a broader definition, according to which a conflict also arises when two statements within the same deontic conceptual category are mutually exclusive (e.g., contradictions between two obligations | e.g., [17][20][24]) or, more generally, when two actions are mutually exclusive (e.g., [18][28][49]). In contrast, studies that focused on regulatory texts in the controlled natural language SBVR often covered specific types of logical anomalies (e.g., [5][37][40]). These anomalies included, for example, cases of subsuming rules, where the condition of one rule implies or includes the condition of another rule, so that whenever one rule applies, the other rule also applies (redundant rules), cases where rules refer to each other in a way that creates an endless loop without providing a concrete solution (circular rules), and cases where rules logically contradict each other, so that no possible situation exists in which all rules could be fulfilled at the same time (inconsistent rules).

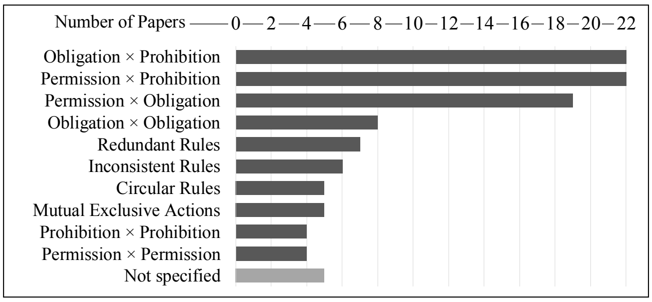

Fig. 9. Types of Conflicts

8) Context Window

When detecting conflicts in regulatory texts, different context windows can be taken into account for each information comparison. According to our literature analysis, a very frequently occurring context window is one sentence per "side", i.e. the comparison of two sentences (e.g., [20]-[27] [17]). This means that the information contained in one sentence is used to decide whether there is a contradiction or inconsistency with another sentence. In contrast, there are only two studies in which several coherent sentences form the context of a textual statement, namely in [47], where entire passages from Standard Operating Procedures (SOPs) were compared with recommendations from medical guidelines, and in [12], where inferences were drawn between coherent texts from Non-Disclosure Agreements (NDAs) and predefined hypothesis sentences. However, the most common context window is not based on a pairwise comparison, where *n* sentences represent the respective context, but on a holistic approach that considers several or all rules simultaneously. However, every study that carried out such a holistic verification first converted the corresponding textual information into a controlled form, such as CL (e.g., [30][48][54]), ontologies (e.g., [33][46]), or SBVR (e.g., [5][37][40]). The latter is usually further translated into the standardized SMT-LIBv2 format, which can then be processed by corresponding SMT-solvers for the purpose of consistency checks (see B.10 for more details).

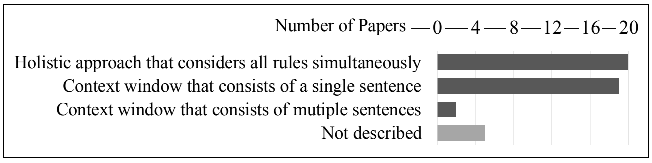

Fig. 10. Context Window

9) Number of Conflicts / Inconsistencies / Contradictions

Most of the studies assessed the accuracy of the approaches they developed using a quantitative evaluation by comparing the contradictions or inconsistencies detected by the respective algorithm / model with the actual ground truth cases. These cases were annotated either by the authors themselves, by independent volunteers (e.g., [20]-[26]) or by domain experts [9]. The number of resulting annotated candidate pairs and the number of actual contradiction cases they contained varied greatly between the studies. Most studies used between 1 and 299 (e.g. [9][23][26]) and between 11,000 and 11,999 candidate pairs (e.g. [21][24][25]). Only one study [27] used more, with a total of 17,230 candidate pairs from two data sets. Although the counting intervals of the contained contradiction or inconsistency cases are closer together, they are significantly lower compared to the total of all annotated candidate pairs. Specifically, this means that, with the exception of two studies [12][27], no more than 249 contradiction or inconsistency cases were used in the studies. However, in this context it should be noted that not all of the 46 studies used quantitative evaluations to assess their approaches. Instead, 23 studies only presented the functionality and effecttiveness of their approaches using selected individual examples (Fig. 11, grey bars).

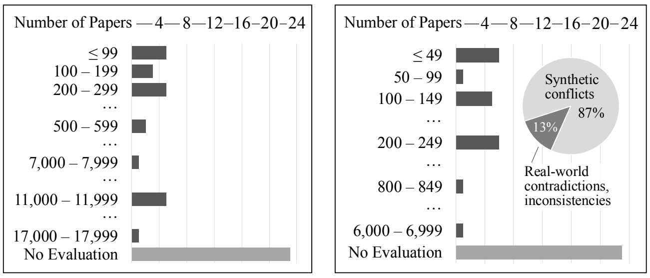

Fig. 11. Number of Candidate Pairs (left), Number of Conflicts (right)

Another important aspect is the origin of the contradictions or inconsistencies contained in the datasets. In this context, our analysis has shown that 87% of all studies (that carried out a quantitative evaluation) created semi-synthetic conflicting sentence pairs and inserted them into their datasets (Fig. 11, pie chart). For example, in [20][24] and [26], two volunteers were asked to modify randomly selected contract norms in such a way that they contradicted the original norms. In this way, they created up to 228 norm conflicts resulting from different modal verbs or deontic conflicts within the respective norm actions. In [27] norm pairs were also modified so that they contained corresponding conflicts. However, the authors additionally increased the number of resulting sentence pairs by performing various methods of text expansion [55], such as random insertion, random swapping and random deletion. Using this approach, they increased the number of their conflicting norm pairs from 523 to 6,230.

10) Technical Approach for Detection

The actual detection of conflicts, contradictions and inconsistencies was carried out in most studies using formal verification methods (Fig. 12). For this form of detection, the rules contained in the normative sentences were first converted into formal logic expressions (e.g., from SBVR notation to SMT-LIBv2) in order to verify their fulfillment by different solvers. For example, in the case of [37]-[40] and [5], an SMT-solver was used that summarizes all transferred rule conditions in specific formulas in order to check the compatibility of the rules. In this process, the solver searches for a value assignment that fulfills all requests. If such an assignment is missing, this indicates conflicts between the rules (unsatisfiability).

The second most common detection approach is based on a classifier that was previously trained with the help of machine learning. Different model architectures were used for this training, such as convolutional neural networks / CNN (e.g., [22][25][26]), long short-term memory / LSTM (e.g., [27][34]), support vector machines / SVM (e.g., [20][24]) and transformer (e.g., [17][47]). For example, in [26], semi-synthetic norm pairs, consisting of neutral and contradictory cases, were converted into a matrix form in which the characters of one norm represent the rows and the characters of the other norm represent the columns. In order to emphasize similarities and differences between the norms, matching characters in the corresponding cells of the matrix are assigned a 1 and otherwise a 0. Using all annotated sentence pair matrices, a CNN was then trained to create a model for classifying norm pairs as conflicting and non-conflicting. Other approaches, such as [17] have trained a neural network, consisting of the decoder part of the Transformer architecture, with the SNLI dataset [52] to classify the inference between a premise and a hypothesis. They then used this trained neural network on a corpus containing a set of conflicting norms to see if they could be identified by the model.

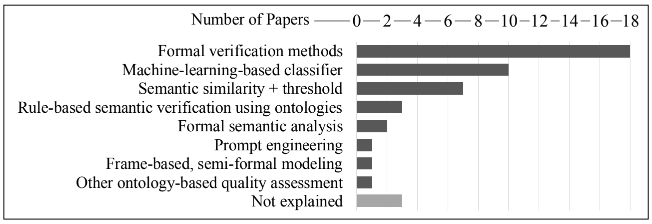

Fig. 12. Technical Approach for the Conflict / Inconsistency Detection

The third most common approach is based on a two-step procedure in which the similarity between the respective texts (or specific entities) is first measured and then used to detect conflicts / inconsistencies if it exceeds or falls below a previously defined threshold (e.g., [21][23][51]). Such semantic comparisons have been used in different ways, such as in [11], where a pair of constraints was classified as conflicting if the semantic similarity of the constraint related subjects is low, while the similarity of the tasks (actions) is high.

In addition to machine learning, semantic-based and formal verification approaches, rule-based semantic verification methods based on ontologies were also used, albeit much less frequently. For example, in [33], legal classes such as "obligation", "permission", "prohibition", their properties and the relationships between these classes are defined in an ontological model and the respective texts and their components are represented using an Resource Description Framework (RDF). A reasoning engine then uses both the RDF data and the ontological knowledge to analyze relationships between the rules and thereby identify overlaps and contradictions. Besides another ontology-based approach, in which inconsistencies are only one of 28 quality metrics [50], two formal semantic approaches using text analysis and information retrieval [42][43] as well as a semi-formal approach using a frame-based knowledge representation were also used to identify conflicts or inconsistencies [31]. In addition, a recent study [53] also investigated how prompt engineering can be used to detect and analyze changes in natural language regulation. Although this study does not focus on conflicts in particular, it enables human analysts to detect regulatory changes that can be understood as conflicts according to common interpretation approaches (see Section II.B).

IV. Conclusion

In this literature review, we have provided a structured overview of papers that aim to detect conflicts, contradictions, or inconsistencies within regulatory documents. As part of the literature analysis, we analyzed 46 papers in the 9 categories: "types and domains of regulatory documents", "form and language of the textual content", "number and origin of the contained inconsistencies / contradictions", "content selection", "sentence pairing", "scope of sentence comparison", "size of context window", "interpretation of inconsistencies / contradictions" and "technical approach for detection". For example, in the "types of regulatory documents" category, the majority of all studies focused on contracts and the norms they contain, followed by policies / guidelines, business rules and legal provisions. Standards, operating procedures and requirement specifications, on the other hand, were used less frequently. A large proportion of these documents covered either the medical domain or multiple legal contract areas, although two examples from car rental and an airline check-in process were also frequently represented. The underlying language of the documents was English, with the exception of two studies in which French and Dutch texts were used. The evaluation was typically carried out on a prepared data set that already contained only relevant sentences. However, if the upstream step of selecting the sentences to be analyzed from the raw texts was also covered, this was done either manually, with the help of a previously trained classifier, or using regular expressions. In contrast, the comparison process used to detect conflicts between the respective information depended heavily on whether the texts remained in their raw form or were transferred into a controlled form (or were initially available in a controlled form). If the latter was the case, typically no candidate pairs were formed for an information comparison, but all identified rule sets were transferred into formula or graph-based representations and then checked for fulfillment using formal verification methods. This approach was used most frequently, followed by rule-based candidate pair generation, where the semantic similarity between the texts must exceed a certain threshold or where certain terms must match in both texts. Besides the formal verification methods, most studies used machine learning approaches for the actual detection of conflicts, contradictions, or inconsistencies, in which annotated data was used to train a classifier. Other studies, however, also used rule-, semantic-, or ontology-based detection approaches. During this detection process, most approaches that were not based on controlled natural language compared one sentence with another sentence. In these cases, the definition of a conflict between two textual statements was mostly based on deontic logic and the mutual exclusivity of the principles involved (e.g. obligation vs. permission). In the case of the studies that focused on controlled natural language, however, specific types of logical conflicts were used, such as *redundancy, circularity,* or *inconsistency*. Regardless of the detection approach, the number of annotated conflicts was mostly less than 250 (only once 800-849 and once 6k-7k), with 87% of all cases being (semi-)synthetic conflicts.

In view of these analysis results, we identify the following five aspects as problematic and therefore relevant for future research:

- **Non-consideration of certain sentences.** A frequently used first step in conflict detection on regulatory documents is to identify those sentences that can be directly assigned to a permission, a prohibition, or an obligation due to the modal verbs they contain (e.g. *must*, *should*, *can*, *may*, etc.). However, there are also sentences that cannot be directly assigned to any of these three categories, but which nevertheless have regulatory relevance because they describe an important condition, an object state or a procedure. If these sentences are excluded from the analysis from the outset, certain contradictions or inconsistencies may never be detected.

- **Limited context window:** The fact that the majority of studies only compare individual sentences can lead to two disadvantages: Firstly, sentences that contain co-references to previous or subsequent sentences (e.g. if a sentence starts with *"This condition..."*) are typically not considered, even if they might have been relevant. Secondly, it is not possible to check whether a missing information responsible for an inconsistency or conflict is possibly hidden in one of the surrounding sentences. This would be the case, for example, if one of two compared sentences links an action to a specific main condition, but still mentions an exception condition (in the same sentence), while the other sentence only mentions the main condition (and the exception condition in a subsequent sentence). Similarly, one sentence might contain the prescribed procedure for a particular condition or situation in one sentence, while the other sentence divides it into two or three consecutive sentences. If such following information is not taken into account, this can lead to False Positives during the conflict detection.
- **Detection often rule-based or supervised:** Many of the detection approaches used in the studies are either semantic and rule-based or require training data for supervised machine learning. In both cases, the transferability to a broader set of heterogeneous regulatory documents is uncertain, as overfitting to the characteristic properties of the training data used in each case cannot be ruled out. In contrast, the commonly used controlled natural languages (CNLs) have the risk of information loss, as they attempt to reduce the complexity of the natural languages by introducing strict syntactic and semantic rules. However, this can mean that certain nuanced meanings or expressions present in the natural language can no longer be expressed in the controlled language.
- **Non-consideration of intermediate conflict stages:** In most studies, the definition of a conflict between two statements is based on deontic logic and the mutual exclusivity of the principles involved (e.g. obligation vs. permission). However, there are often very few, if any, examples provided that describe the different manifestations of these cases in more detail. In addition, there are also edge cases that do not represent a direct contradiction between two statements but should rather be regarded as a relevant discrepancy (e.g. if sentence A describes the prescribed steps of an action more precisely than sentence B, without contradicting it). These cases have hardly been considered in previous research, which makes it difficult to assess their detection probability in real-world scenarios.
- **Few conflict cases and often synthetic:** Only two studies had more than 250 cases of conflict in their data. In addition, most conflict cases (87%) were synthetically generated. This may affect the assessment of the detection quality, as it raises the question of whether the synthetically generated conflicts also reflect those that occur in practice. Furthermore, half of all studies demonstrated their approaches using only a few individual examples instead of evaluating them holistically using an annotated test data set.

Considering these trends and limitations can help to further define and develop future research on conflict detection in the regulatory domain. On the other hand, these findings may also be relevant for organizations and individuals working with regulatory documents in practice and considering the introduction of automated conflict detection, as they may need to be prepared for challenges that are barely addressed in current research. At the same time, the structured overview of existing approaches can enable them to select concrete methods for their own purposes, provided that these cover their needs and requirements.

ACKNOWLEDGMENT

This work was funded and supported by the Institute for AI Safety and Security at the German Aerospace Center (DLR – Deutsches Zentrum für Luft- und Raumfahrt, Institut für KI-Sicherheit).

REFERENCES

[1] V. H. Peemöller, and J. Kregel, Grundlagen der Internen Revision: Standards, Aufbau & Führung. Germany, Erich Schmidt Verlag, 2014.

[2] J. Berwanger, and S. Kullmann, Interne Revision – Funktion, Rechtsgrundlagen und Compliance (2nd Edition). Wiesbaden, Springer Gabler, 2012.

[3] E. Knapp, Interne Revision und Corporate Governance. Aufgaben und Entwicklungen für die Überwachung (2nd Edition). Germany, Erich Schmidt Verlag, 2009.

[4] M. Hashmi, "A methodology for extracting legal norms from regulatory documents", In: 2015 IEEE 19th International Enterprise Distributed Object Computing Workshop, pp. 41–50, IEEE, 2015.

[5] P. K. Chittimalli, and K. Anand, "Domain-independent method of detecting inconsistencies in sbvr-based business rules. In Proceedings of the International Workshop on Formal Methods for Analy-sis of Business Systems, pp. 9–16, 2016.

[6] S. Mitra, and P. K. Chittimalli, "A Systematic Review of Methods for Consistency Checking in SBVR-based Business Rules", In: DIAS/ EDUDM@ISEC, 2017.

[7] G. Morley, "Regulatory Writing", Medical Writing, 23, pp. 58–59, 2014.

[8] C. Sai, K. Winter, E. Fernanda, and S. Rinderle-Ma, "Detecting deviations between external and internal regulatory requirements for improved process compliance assessment", In: International Conference on Advanced Information Systems Engineering, pp. 401–416, Cham: Springer Nature Switzerland, 2023, June.

[9] N. Madaan, G. Singh, S. Bedathur, and A. Kumar, "A Change Tracking Framework for Financial Documents", In: 2018 International Conference on Services Computing (SCC), pp. 81–88. IEEE, 2018.

[10] D. Collarana, T. Heuss, J. Lehmann, I. Lytra, G. Maheshwari, R. Nedelchev, T. Schmidt, and P. Trivedi, "A question answering system on regulatory documents", In: Legal Knowledge and Information Systems, pp. 41–50, IOS Press, 2018.

[11] K. Winter, and S. Rinderle-Ma, "Detecting constraints and their relations from regulatory documents using nlp techniques", In: OTM Confederated International Conferences "On the Move to Meaningful Internet Systems", pp. 261–278, Springer, Cham, 2018.

[12] Y. Koreeda, and C. D. Manning, "ContractNLI: A dataset for document-level natural language inference for contracts", 2021, arXiv preprint arXiv:2110.01799.

[13] S. Fenech, G. J. Pace, and G. Schneider, "Clan: A tool for contract analysis and conflict discovery", In: International Symposium on Automated Technology for Verification and Analysis, pp. 90–96, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, October.

[14] S. Fenech, G. J. Pace, and G. Schneider, "Automatic conflict detection on contracts". In: International colloquium on theoretical aspects of computing, pp. 200–214, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, August.

[15] S. M. Montazeri, N. K. Roy, and G. Schneider, "From contracts in structured english to CL specifications", 2011, arXiv preprint arXiv: 1109.2657.

[16] P. Rosso, S. Correa, and D. Buscaldi, "Passage retrieval in legal texts", The Journal of Logic and Algebraic Programming, 80(3–5), 139–153, 2011.

[17] A. Y. Ichida, and F. Meneguzzi, “Detecting Logical Relation in Contract Clauses”, 2021, arXiv preprint arXiv:2111.01856.

[18] S. Azzopardi, G. J. Pace, F. Schapachnik, and G. Schneider, “Contract automata: an operational view of contracts between interactive parties”, Artificial Intelligence and Law, 24, 203–243, 2016.

[19] G. Schneider, “Specification and Verification of Normative Documents”, In: Formal Methods for Software Engineering: Languages, Methods, Application Domains, pp. 307–343, Cham: Springer International Publishing, 2022.

[20] J. P. Aires, R. Granada, J. Monteiro, R. C. Barros, and F. Meneguzzi, “Classifying norm conflicts using learned semantic representations”, 2019, arXiv preprint arXiv:1906.02121.

[21] J. P. Aires, D. Pinheiro, V. S. D. Lima, and F. Meneguzzi, “Norm conflict identification in contracts”, Artificial Intelligence and Law, 25(4), 397–428, 2017.

[22] J. P. Aires, and F. R. Meneguzzi, “A deep learning approach for norm conflict identification”, In: Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2017, Brasil, 2017.

[23] J. P. Aires, J. Monteiro, R. Granada, and F. Meneguzzi, “Norm conflict identification using vector space offsets”, In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, IEEE, 2018.

[24] J. P. Aires, R. L. Granada, J. Monteiro, R. C. Barros, and F. R. Meneguzzi, “Classification of contractual conflicts via learning of semantic representations”, In: Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2019, Canadá, 2019.

[25] J. P. Aires, R. L. Granada, and F. R. Meneguzzi, “ConCon: a contract conflict identifier”, In: Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems, 2019, Canadá, 2019.

[26] J. P. Aires, and F. Meneguzzi, “Norm conflict identification using a convolutional neural network”, In: Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XIII: International Workshops COIN 2017 and COINE 2020, Sao Paulo, Brazil, May 8–9, 2017 and Virtual Event, May 9, 2020, Revised Selected Papers, pp. 3–19, Springer International Publishing, 2021.

[27] S. Huang, J. Sun, and R. Li, “NeuralConflict: Using neural networks to identify norm conflicts in normative documents”, Expert Systems, 41(6), e13035, 2024.

[28] S. Azzopardi, A. Gatt, and G. J. Pace, “Integrating natural language and formal analysis for legal documents”, 2016.

[29] A. Khoja, M. Kölbl, S. Leue, and R. Wilhelmi, “Automated Consistency Analysis for Legal Contracts”, In: International Symposium on Model Checking Software, pp. 1–23, Cham: Springer International Publishing, 2022, May.

[30] J. Gordon, “A General Theory of Contract Conflicts with Environmental Constraints”, JURI SAYS, 83, 2020.

[31] M. Araszkiewicz, and T. Zurek, “Identification of Legislative Errors Through Knowledge Representation and Interpretive Argumentation”, In: International Workshop on AI Approaches to the Complexity of Legal Systems, pp. 15–30, Cham: Springer International Publishing, 2018, December.

[32] M. Araszkiewicz, E. Francesconi, and T. Zurek, “Identification of Legislative Errors”, In: Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law, pp. 2–11, 2023, June.

[33] M. Araszkiewicz, E. Francesconi, and T. Żurek, “Identification of contradictions in regulation”, Legal Knowledge and Information Systems, pp. 151–160, IOS Press, 2021.

[34] A. R. Alshamsan, and S. A. Chaudhry, “Detecting Privacy Policies Violations Using Natural Language Inference (NLI)”, In: 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1–6, IEEE, 2022, December.

[35] D. B. Raykar, L. T. JayPrakash, and K. V. Dinesha, “An Iterative and Incremental Approach to Address Regulatory Compliance Concerns in Requirements Engineering”, In: Advanced Computing: 10th International Conference, IACC 2020, Panaji, Goa, India, December 5–6, 2020, Revised Selected Papers, Part II 10, pp. 323–335, Springer Singapore, 2021.

[36] D. Solomakhin, E. Franconi, and A. Mosca, “Logic-based reasoning support for SBVR”, Fundamenta Informaticae, 124(4), 543–560, 2013.

[37] P. K. Chittimalli, K. Anand, S. Pradhan, S. Mitra, C. Prakash, R. Shere, and R. Naik, “BuRRiTo: a framework to extract, specify, verify and analyze business rules”, In: 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 1190–1193, IEEE, 2019, November.

[38] K. Anand, P. K. Chittimalli, and R. Naik, “An automated detection of inconsistencies in sbvr-based business rules using many-sorted logic”, In: Practical Aspects of Declarative Languages: 20th Interna-tional Symposium, PADL 2018, Los Angeles, CA, January 8–9, 2018, Proceedings 20, pp. 80–96, Springer International Publishing, 2018.

[39] K. Anand, S. Mitra, and P. K. Chittimalli, “Semantic Search and Query Over SBVR-based Business Rules using SMT based Approach and Information Retrieval Method”, In: ENASE, pp. 47–58, 2019.

[40] S. Mitra, K. Anand, and P. K. Chittimalli, “Identifying Anomalies in SBVR-based Business Rules using Directed Graphs and SMT-LIBv2”, In: ICEIS (2), pp. 215–222, 2018.

[41] C. Kacfah Emani, “Automatic detection and semantic formalisation of business rules”, In: The Semantic Web: Trends and Challenges: 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25–29, 2014. Proceedings 11, pp. 834–844, Springer International Publishing, 2014.

[42] H. Hematialam, “Knowledge Extraction and Analysis of Medical Text with Particular Emphasis on Medical Guidelines”, (Doctoral dissertation), The University of North Carolina at Charlotte, 2021.

[43] W. Zadrozny, H. Hematialam, and L. Garbayo, “Towards semantic modeling of contradictions and disagreements: A case study of medical guidelines”, 2017, arXiv preprint arXiv:1708.00850.

[44] W. Zadrozny, “Detecting and representing contradictions and disagreements in medical guidelines”, In: MedRACER 2018 and WOMoCoE 2018, pp 3–5, 2018.

[45] R. Tsopra, J. B. Lamy, and K. Sedki, “Using preference learning for detecting inconsistencies in clinical practice guidelines: methods and application to antibiotherapy”, Artificial intelligence in medicine, 89, pp. 24–33, 2018.

[46] A. Galopin, J. Bouaud, S. Pereira, and B. Séroussi, “Using an ontological modeling to evaluate the consistency of clinical practice guidelines: application to the comparison of three guidelines on the management of adult hypertension”, In: e-Health–For Continuity of Care, pp. 38–42, IOS Press, 2014.

[47] L. van Leijenhorst, A. P. de Vries, T. Habben Jansen, and H. Wertheim, “SOPalign: A Tool for Automatic Estimation of Compliance with Medical Guidelines”, In: European Conference on Information Retrieval, pp. 307–312, Cham: Springer Nature Switzerland, 2023.

[48] K. Angelov, J. J. Camilleri, and G. Schneider, “A framework for conflict analysis of normative texts written in controlled natural language”, The Journal of Logic and Algebraic Programming, 82(5–7), 216–240, 2013.

[49] É. Cota, L. Ribeiro, J. S. Bezerra, A. Costa, R. E. da Silva, and G. Cota, “Using formal methods for content validation of medical procedure documents”, International journal of medical informatics, 104, 10–25, 2017.

[50] J. L. de la Vara, H. Bahamonde, and C. Ayora, “Assessment of the quality of the text of safety standards with industrial semantic technologies”, Computer Standards & Interfaces, 88, 103803, 2024.

[51] H. Ren, Y. Cai, M. Zhang, W. Hao, and X. Wu, “Standard-Oriented Standard Knowledge Graph Construction and Applications System”, In: Web and Big Data: 5th International Joint Conference, APWeb-WAIM 2021, Guangzhou, China, August 23–25, 2021, Proceedings, Part II 5, pp. 452–457, Springer International Publishing, 2021.

[52] S. R. Bowman, G. Angeli, C. Potts and C. D. Manning, “A large annotated corpus for learning natural language inference”, 2015, arXiv preprint arXiv:1508.05326.

[53] S. Abualhaija, M. Ceci, N. Sannier, D. Bianculli, L. Briand, D. A. Zetzsche, and M. Bodellini, “AI-enabled Regulatory Change Analysis of Legal Requirements”, In: 32nd IEEE International Require-ments Engineering Conference. IEEE, 2024.

[54] A. L. Bonifácio, and W. A. Della Mura, “Automatically running experiments on checking multi-party contracts”, Artificial Intelligence and Law, 29, 287–310, 2021.

[55] J. Wei, and K. Zou, “EDA: Easy data augmentation techniques for boosting performance on text classification tasks”, In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.

[56] S.-A. Sadat-Akhavi, Methods of Resolving Conflicts Between Treaties, volume 3 of Graduate Institute of International and Development Studies. Martinus Nijhoff Publishers, 2003.