Reasoning on Privacy Policies

Yilian Huang[1,*], Cosimo Perini Brogi[1,*] and Rocco De Nicola[2,3]

[1]*IMT School for Advanced Studies Lucca, Italy*
[2]*Institute for Informatics and Telematics - CNR Pisa, Italy*
[3]*Cybersecurity National Laboratory, CINI - Roma, Italy*

Abstract

Privacy policies stated in natural language are often ambiguous and not amenable to direct machine processing. Controlled Natural Languages (CNLs) provide a simplified, structured way to express policy rules that is both easier for natural language processing (NLP) to parse and closer to formal representations used in automated reasoning (AR). By translating further CNL statements into logical systems, we can connect NLP with rigorous AR to verify privacy policies and related regulations. This paper proposes a three-layer methodology – based on NLP, CNL, and AR – for the formal verification of privacy policies, and describes our ongoing development of this approach.

Keywords

Privacy policies, Formal verification, Controlled Natural Languages, Automated reasoning, Natural Language Processing,

1. Introduction

Privacy policies are written statements that companies are required to provide to explain how users' personal information will be collected, used, stored, and shared by first-party and third-party organizations [1, 2]. These policies are commonly written in natural language. However, they are usually very long, filled with legal terminology, and challenging to understand without a legal background. Consequently, users tend to accept them uncritically without reading them. This reduces the companies' general commitment to transparency of privacy policies and increases privacy risks [3].

Still, privacy policies remain essential for protecting user rights and ensuring corporate accountability. In recent years, governments have increasingly emphasized personal data protection by implementing specific regulations, such as the EU GDPR (General Data Protection Regulation) and the Law Enforcement Directive (LED), aiming to provide consistent protection of personal data processed by law enforcement in EU institutions [4, 5].

Privacy policies must be precise and easy to understand because of their importance and relevance in our daily lives and societies. To achieve this, it seems relevant to adopt structured representations of policies using semi-formal languages – better if easy-to-read for humans – that can then be translated into formal languages – suitable for efficient automatic reasoning (AR).

Rigorous models for concepts such as obligations, permissions, prohibition and conditional allowance are also essential to really understand privacy policies, even though what is effectively permitted and/or required may be obscured – even to natural language processing (NLP) tools – by the legal jargon of policy rules. Systems of formal logic offer various types of mathematical languages capable of expressing those concepts, by means of, for instance, modal/deontic logics, which have clear syntax and semantics presentations in mathematical and computational terms [6, 7].

This paper presents a three-layer approach for translating natural language privacy policies into formal logic. We argue that a Controlled Natural Language (CNL) [8] is the ideal intermediate layer.

OVERLAY 2025: 7th International Workshop on Artificial Intelligence and fOrmal VERification, Logic, Automata, and sYnthesis, co-located with ECAI 2025, October 26, 2025, Bologna, Italy

[*] Corresponding authors.

yilian.huang@imtlucca.it (Y. Huang); cosimo.perinibrogi@imtlucca.it (C. Perini Brogi); rocco.denicola@iit.cnr.it (R. De Nicola)

0009-0004-6066-1289 (Y. Huang); 0000-0001-7883-5727 (C. Perini Brogi); 0000-0003-4691-7570 (R. De Nicola)

Indeed, a CNL bridges the gap between human language's ambiguity and formal logic's rigid syntax. This structured layer makes policies machine-readable for automated reasoning while remaining understandable to humans.

To bridge the gap between human readability and suitability for machine processing, we propose a three-step process:

1. Extraction: We apply current NLP tools and techniques to identify structural and linguistic features in privacy policies.

2. Disambiguation: Using this information, we translate policies into a CNL format with restricted grammar and vocabulary, thereby resolving ambiguities inherent in the original text.

3. Formalization: Finally, we map the CNL expressions – soundly and completely – into logical formulas of a suitable system (e.g., first-order, modal, or other logics suited for legal analysis [9]). This enables the use of optimized AR tools to verify the consistency and compliance of the legal agreement.

The remainder of this paper is structured as follows. First, we discuss current challenges in privacy policy analysis and emphasize the critical role of formal reasoning in addressing them. Next, we detail our mapping mechanism from natural language to (semi-)formal languages, examining two specific CNLs as candidates for the intermediate layer. Finally, we compare our NLP pipeline to a prior approach [10], highlighting our improvements and outlining future enhancements.

2. The Need for Formal Versions of Privacy Policies

Privacy policies stated in natural language pose challenges that can be categorized into four core dimensions: *readability*; *comprehensibility*; *ambiguity*; and *accessibility* [11, 12].

According to [13], lengthy privacy policies, full of jargon and advanced legal vocabulary, would cost people eight to twelve minutes to finish *reading*. In that study, researchers also estimated, relying on experimental data, that if all American internet users were to read every privacy policy word-for-word, the whole country would cumulatively spend approximately 54 billion hours doing so. In addition, *comprehending* these policies requires at least 13.6 years of education on average.

Regarding *ambiguity*, natural languages inherently allow for words or phrases to have multiple meanings and interpretations, which may be exploited in potentially misleading ways. For example, malicious companies may intentionally utilize ambiguous phrases to mislead users and hide the potential risk of their policies [14].

Accessibility also remains a challenge. Although most websites and organizations directly provide privacy policies, users often struggle to find relevant information. Even when policies are accessible, information vital to users' rights — such as data storage or sharing — may be buried within and scattered across complex and lengthy texts [15, 16, 17].

2.1. Compliance Checking and Policies Verification

We propose automatically transforming privacy policies into formal languages to reduce the ambiguity inherent in natural language statements. Formal languages provide strict and precise symbols and rules, allowing natural language policies to be broken down into clear logical components. By expressing policies via a rigorous syntax, ambiguity can be significantly reduced, if not eliminated. Crucially, a mathematical representation also enables rigorous implementation of three key tasks: (a) *compliance verification*; (b) *comparison*; (c) *change detection.*

a. Verification of Compliance with Legal Regulations. Legal regulations like the General Data Protection Regulation (GDPR) [18] and the California Consumer Privacy Act (CCPA) [19] require companies and organizations to provide clear, transparent, understandable, and accessible privacy policies for users. Their policies must include information such as the type of personal information they collect and how long they will keep it. Verifying whether a company's privacy policy complies with these regulations is difficult when written in natural language for the abovementioned reasons: length, technical and legal terminology, accessibility limitations.

By translating privacy policies into formal languages, such as those of logical systems, or semi-formal, such as CNLs, comparing policy statements with regulatory requirements becomes a feasible automated task. Formal languages with (more or less strict) constrained structure enables the computerized detection of missing information and ambiguous, contradictory, or non-compliant rules.

As expressive CNLs, close to human language, may retain many natural-language features, they can still contain ambiguous or hard-to-parse expressions. Resolving those ambiguities often requires sacrificing some of the original wording or nuance, and it typically forces manual compliance checks and verification. That manual validation is time-consuming, error-prone, and impractical to scale. Fully formalizing policies as logical formulas of a chosen formal language avoids these issues by producing unambiguous, machine-checkable specifications that support automated verification and scalable enforcement.

b. Comparison of Cross-jurisdictional and Cross-company policies. With increasing globalization, firms routinely operate across multiple jurisdictions and provide online services to users worldwide [20]. Consequently, it is essential to evaluate whether a company's privacy policy complies with each jurisdiction's legal requirements. Since data-protection regimes differ in scope, definitions, and obligations, organizations expanding internationally must determine the applicable foreign rules and adapt their policy statements accordingly [21, 18].

Manual cross-jurisdictional comparison of privacy policies is laborious, error-prone, and scales poorly as the number of target markets increases. Representing privacy policies in a shared, formally structured language with broad international acceptance would allow such comparisons to be framed as syntactic and deductive problems, amenable to automated consistency checks, formal verification, and more efficient compliance workflows.

c. Detection of Privacy Policy Changes. Another problem is that companies have the right to update their privacy polices regularly [22]. Although they notify users of such changes, there is a widely observed tendency among users to ignore these changes. Thereby, users also neglect the potential influence these changes may have on processing and safeguarding their personal information.

When policies are expressed formally, each statement becomes more precise and well-defined. This makes it possible to compare versions and to detect changes reliably – for example, determining whether a user's rights have been strengthened or weakened can be reduced to a pattern-matching task (e.g. regular expressions) or addressed through deductive verification. This kind of automated detection – potentially supported by tailored human-machine interaction interfaces – can help both users and regulators to monitor how companies manage personal information and plan to use it in the future, which vastly increases the transparency of privacy policies and provides a precise record how a company's policy evolves over time.

2.2. Controlled Natural Languages: A Viable Middle Ground

Although formal languages can precisely represent the semantics of privacy policy and support automated reasoning and compliance verification, their complex and technical syntax makes them difficult for everyday users to read and understand. To address the challenge of explaining the outcomes of formal reasoning and policy analysis to general users, it is helpful to integrate the proposed mathematical approach with an intermediate layer to guarantee both accuracy and readability of privacy statements. Controlled natural languages (CNLs) – that is, restricted and simplified subsets of natural languages with clearly defined grammar and limited lexicon [23] – are well-suited for this purpose.

Two widely used controlled natural languages are Attempto Controlled English (ACE) and Processable English (PENG). Both support automated translation into first-order logic via Discourse Representation

Structures (DRSs) [24, 25].

To achieve our goal, it is essential to have rigorous counterparts of concepts such as obligations, prohibitions, and conditional permissions, as reflected in several privacy statements, including the following example from [26][1]:

> You *must* provide your mobile phone number and basic information *to create* a WhatsApp account.

First-order logic (FOL) is inadequate for correctly formalizing this privacy statement because its expressive power cannot capture the statement's core semantic components: obligation and purpose. The word 'must' introduces a deontic modality – a notion of duty or necessity – that falls outside the purely descriptive, truth-functional scope of FOL, which concerns what is true, not what *ought* to be true. A dedicated deontic logic would be required to handle this. Furthermore, the phrase 'to create' implies a purposeful action that results in a state change, a dynamic process that FOL's static nature is ill-equipped to model. Capturing the relationship between an action and its outcome necessitates a more expressive framework, such as temporal or dynamic logics [27, 28]. More generally, since privacy statements articulate rules governing behaviour rather than a simple assertion of fact, their accurate formalization demands a logic capable of reasoning about both modality and dynamics.

ACE provides constructs with modal auxiliaries for expressing obligation, permission, and prohibition in the construction rules;[2] PENG provides temporal ordering expressions[3]. Nonetheless, these constructs are not directly mapped in the FOL/DRSs translation mechanism provided by the currently available tools for those CNLs. Furthermore, legal documents are known to contain qualitatively different concepts – such as defaults and exceptions that can be effectively captured by systems of non-monotonic logic [9]. Recent ACE releases demonstrate a variant support expressive enough for simple non-monotonic normative reasoning [30, 31]. Therefore, one of our future goals is to identify a "CNL environment" (or suite) that can fully express the meaning of policy rules while ensuring a sound and complete translation of the identified language into formal logics.

CNLs could be used to design a Question-Answering (QA) chatbot as a practical byproduct of such an intermediate layer. Users could ask questions about policies, and the chatbot will quickly search for the corresponding policy rule[4], and then answer the original policy rule in a clear and concise CNL format, validated by its translation into logical statements in the background. Such a tool would inform users about privacy policies more effectively than a quick and distracted glance at them before accepting the legal agreements they propose.

3. First Steps in Translating Privacy Policies

To translate privacy policies into CNLs or formal languages, it is crucial to capture the core semantic meaning of sentences, beginning with basic Subject–Verb–Object (SVO) triplets [32]. This structure identifies the agent performing the action (the subject), the recipient of the action (the object) and the semantic relation between them (the verb). Therefore, an automated tool capable of accurately extracting this structure from a sentence is an essential and basic component of the entire formal verification process. The effectiveness and completeness of CNLs' translation into logical formulas depend heavily on the accuracy of this step. With this in mind, we will discuss the methodologies we are considering to achieve this goal.

The work in [10] serves as our baseline. It presents a system for translating social network policies into CNLs using an ontology-based approach that systematically processes the original text to produce a machine-readable format: the controlled natural language CNL4DSA [33]. The process commences with an automated NLP phase, where the policy is parsed to analyze its syntactic structure, tokenized

[1] The emphasis from the extract is ours.

[2] We refer the reader to the website of ACE construction rules: ACE 6.7 Construction Rules.

[3] The temporal order principle implemented in PENG is discussed in [29, §4].

[4] NLP tools for semantic embedding and semantic searching might be used, but we will discuss them in future work.

into sentences and words, and refined by removing actions, predicates, and standard stopwords. The system then automatically tags the remaining words and extracts instances of subjects, objects, actions, and predicates. These extracted instances are subsequently used to populate a pre-defined ontology, creating a formal, structured representation of the policy's rules. The final stage involves translating this ontological representation into CNL4DSA[5]. Their framework is *semi*-automatic and relies on crucial human intervention at several key stages: a human operator must provide a pre-defined list for word replacement, supply any additional custom stopwords for removal, supply a dictionary to train the Unigram Tagger, and initially define the core ontology, including its classes and properties, before the automated processing can commence.

By leveraging several NLP tools for syntactic *and* semantic analysis, we argue that many of the steps previously requiring manual definition can now be automated. These NLP tools provide more accurate dependency parsers capable of identifying SVO triplets even in complex sentence structures. Building on the general methodology of [10], we propose a lightweight NLP pipeline that enhances their workflow by automatically extracting the triplets. Figure 1 visually compares the two workflows.

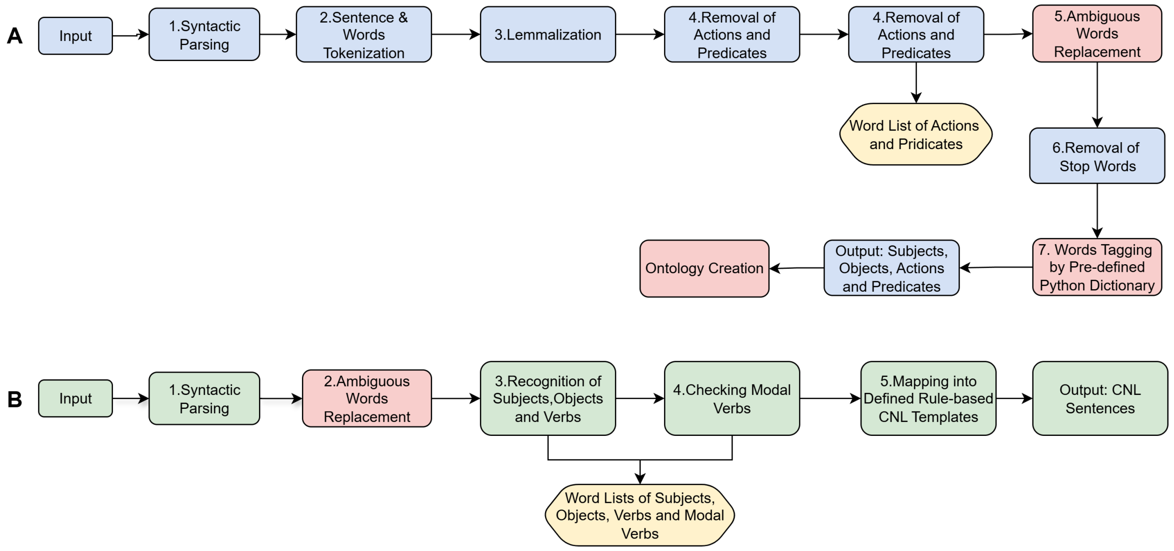
A
Input
1.Syntactic Parsing
2.Sentence & Words Tokenization
3.Lemmalization
4.Removal of Actions and Predicates
4.Removal of Actions and Predicates
5.Ambiguous Words Replacement
Word List of Actions and Pridicates
6.Removal of Stop Words
7. Words Tagging by Pre-defined Python Dictionary
Output: Subjects, Objects, Actions and Predicates
Ontology Creation
B
Input
1.Syntactic Parsing
2.Ambiguous Words Replacement
3.Recognition of Subjects,Objects and Verbs
4.Checking Modal Verbs
5.Mapping into Defined Rule-based CNL Templates
Output: CNL Sentences
Word Lists of Subjects, Objects, Verbs and Modal Verbs

Figure 1: Flowchart A, sourced from [10], illustrates the pipeline we are currently improving; Flowchart B presents our proposed enhancement. Red boxes indicate that this step requires manual intervention; the remaining colored boxes represent automated steps.

Our pipeline begins by parsing the syntactic dependency structure of sentences. Based on the dependency relations, it identifies SVO triplets and separately stores them in three lists. This automated SVO extraction directly replaces the manual dictionary construction and tagging steps required in the baseline approach. After extracting the triplets, our pipeline searches for nearby modifiers, such as adjectives and adverbs, to provide a full noun phrase rather than single words. This step is another improvement and ensures that modifiers are included in each component of the SVO triplet to preserve the original semantic meaning of the sentence. Table 1 showcases some examples of SVO-outputs from our automated prototype.

We also aim to fully automate the pronoun replacement step of [10] by using coreference resolution tools to address pronominal ambiguities [36]. At the current stage of the project development, our attempt has faced some challenges that we are still addressing and will elaborate on in Section 4 below.

Despite this temporary limitation, the current version of our NLP pipeline performs well on corpora of limited complexity, including sentences with embedded structures (i.e., sentences nested within other sentences) and non-trivial modifiers of SVO-items, showing promising potential for advancing our research project.

[5] We note in passing that a variation of that CNL, named CL4SPL, is already implemented in the theorem prover Maude, as used for the formal verification of product line engineering [34]. An automated formalization in CNL4DSA of railway requirements is discussed in [35].

Original Sentences	Extracted SVO Structures	Remarks
We are committed to protecting your privacy and ensuring a secure and enjoyable experience on our platform.	• (We, are, —) • (We, committed, —) • (We, protecting, your privacy) • (We, ensuring, a secure experience on our platform)	Correctly handles linking verb "are", extracts infinitive and gerund structures after "committed to" and conjunctions.
This Privacy Policy outlines how we collect, use, disclose, and safeguard your personal information when you use our online streaming service, including our website, applications, and any related services.	• (Privacy Policy, outlines, how we collect/use/disclose/safeguard...) • (we, collect, —) • (we, use, —) • (we, disclose, —) • (we, safeguard, your personal information) • (you, use, our online streaming service)	Handles embedded noun clauses as objects; identifies internal multiple SVOs within them.
We collect various types of information to provide and improve our Service, personalize your experience, and communicate with you.	• (We, collect, various types of information) • (We, provide, —) • (We, improve, our Service) • (We, personalize, your experience) • (We, communicate with, you)	Handles coordinated actions and prepositional objects (e.g., "communicate with you") accurately.

Table 1
Examples of Subject-Verb-Object (SVO) Extraction and Minimal Structural Analysis

4. Conclusions

To enhance verification and compliance checks of privacy policies, we advocate for the systematic adoption of mathematical logics – be they classical or non-classical – for formal reasoning on those legal agreements. A combination of various systems of formal logic provides a solid foundation for policy verification, and their implementations can be leveraged to achieve high confidence in privacy policy compliance and formal reasoning. To address common challenges to rigorous analysis of those policies, such as ambiguous expressions and opaque structure formulation, we recognize that a direct translation of policies into logic formulas is not feasible. Moreover, the formal syntax of logic languages in some cases might be too technical and complex for everyday users of online services to read and understand.

We thus propose the adoption of Controlled Natural Languages (CNLs) as a structured intermediary between the original policy text and their purely formal representation. Specifically, we advocate using Natural Language Processing (NLP) techniques to automatically extract relevant linguistic structures from policy statements and translate them into clear, simplified CNL expressions. These expressions preserve the semantic meaning of the original rules and can be further transformed into formal logics formulae for automated reasoning (AR) and compliance verification. In most cases, CNLs effectively balance linguistic precision and human readability.

Future work. The workflow and NLP pipeline proposed here aims to automate steps from [10] that previously required manual intervention, while preserving the full semantic meaning of sentences when converting into a CNL. Although our initial results are promising, applying basic coreference resolution tools has revealed limitations in the prototype pipeline. A key dilemma arises depending on when the resolution is applied. If performed *before* subject-verb-object (SVO) triplet extraction, pronouns are sometimes replaced with incorrect referent, adversely affectings the text's syntactic structure and compromisings downstream analytical tasks. However, if performed *after* SVO extraction, the process lacks the necessary context to link anaphora to their antecedents, making substitution impossible. We therefore acknowlege that further work is required to overcome these issues, particularly through evaluation on larger and more diverse corpora of privacy policies.

Our immediate practical goal is to evaluate the prototype pipeline described here using a real-world corpus of privacy policies. This corpus will allow us to assess its overall stability and accuracy. A key part of this work involves exploring more sophisticated approaches to coreference resolution, aiming to reduce ambiguity and enhance contextual consistency during the preliminary NLP stages.

A second crucial step will be to examine how effectively a given CNL can capture the subtleties of rules found in real-world policies. We therefore plan to identify CNLs that are not only highly expressive and readable but can also be automatically mapped to the syntax of a logic system suitable for automated deductive reasoning.

Consequently, a comprehensive formal analysis of legal texts like privacy agreements requires an integrated system capable of consistently capturing various modes of reasoning, including modal, constructive, non-monotone, and defeasible logic. At present, no such integrated system exists [37][6]. A core aspect of our research involves identifying or developing a formal logic system asuitable for automation. The goal is to enable automatic verification of privacy statements that encompasses complex concepts such as exceptions, temporal relations, and the deontic aspects of these agreements.

Acknowledgments

This work was partially supported by the project SERICS – Security and Rights in CyberSpace PE0000014, financed within PNRR, M4C2 I.1.3, funded by the European Union - NextGenerationEU (MUR Code: 2022CY2J5S, CUP: D67G22000340001).

Declaration on Generative AI

While preparing this work, the authors used Generative AI tools to check grammar and spelling. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] D. J. Solove, P. Schwartz, Information privacy law, Aspen Publishers, Inc., 2011.

[2] S. Zimmeck, The information privacy law of web applications and cloud computing, Santa Clara Computer & High Tech. LJ 29 (2012) 451.

[6] General-purpose interactive theorem provers may provide a handy tool for an implementation baseline of different systems for non-classical reasoning and interactions, as witnessed by e.g. [38, 39, 40, 41].

[3] A. Acquisti, R. Gross, Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook, in: International workshop on privacy enhancing technologies, Springer, 2006, pp. 36–58.

[4] C. J. Hoofnagle, B. Van Der Sloot, F. Z. Borgesius, The European Union General Data Protection Regulation: What it is and what it means, Information & Communications Technology Law 28 (2019) 65–98.

[5] E. Kosta, F. Boehm, The EU Law Enforcement Directive (LED): A Commentary, Oxford University Press, 2024.

[6] J. Garson, Modal Logic, in: E. N. Zalta, U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy, Spring 2024 ed., Metaphysics Research Lab, Stanford University, 2024.

[7] P. McNamara, F. Van De Putte, Deontic Logic, in: E. N. Zalta, U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy, Fall 2022 ed., Metaphysics Research Lab, Stanford University, 2022.

[8] T. Kuhn, A survey and classification of controlled natural languages, Computational linguistics 40 (2014) 121–170.

[9] L. Lawniczak, C. Benzmüller, Logical Modalities within the European AI Act: An Analysis, CoRR abs/2501.19112 (2025). URL: https://doi.org/10.48550/arXiv.2501.19112. doi:10.48550/ARXIV.2501.19112. arXiv:2501.19112.

[10] I. K. Tanoli, M. Petrocchi, R. De Nicola, Towards Automatic Translation of Social Network Policies into Controlled Natural Language, in: 2018 12th International Conference on Research Challenges in Information Science (RCIS), IEEE, 2018, pp. 1–12.

[11] A. Adhikari, S. Das, R. Dewri, Natural language processing of privacy policies: A survey, arXiv preprint arXiv:2501.10319 (2025).

[12] A. Adhikari, S. Das, R. Dewri, Evolution of composition, readability, and structure of privacy policies over two decades, Proc. Priv. Enhancing Technol. 2023 (2023) 138–153. URL: https://doi.org/10.56553/popets-2023-0074. doi:10.56553/POPETS-2023-0074.

[13] A. M. McDonald, L. F. Cranor, The Cost of Reading Privacy Policies, Isjlp 4 (2008) 543.

[14] J. R. Reidenberg, J. Bhatia, T. D. Breaux, T. B. Norton, Ambiguity in Privacy Policies and the Impact of Regulation, The Journal of Legal Studies 45 (2016) S163–S190.

[15] G. R. Milne, M. J. Culnan, H. Greene, A Longitudinal Assessment of Online Privacy Notice Readability, Journal of Public Policy & Marketing 25 (2006) 238–249. URL: https://doi.org/10.1509/jppm.25.2.238. doi:10.1509/jppm.25.2.238. arXiv:https://doi.org/10.1509/jppm.25.2.238.

[16] B. Krumay, J. Klar, Readability of privacy policies, in: A. Singhal, J. Vaidya (Eds.), Data and Applications Security and Privacy XXXIV - 34th Annual IFIP WG 11.3 Conference, DBSec 2020, Regensburg, Germany, June 25-26, 2020, Proceedings, volume 12122 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 388–399. URL: https://doi.org/10.1007/978-3-030-49669-2_22. doi:10.1007/978-3-030-49669-2_22.

[17] H. Ding, S. Zhang, L. Zhou, P. Yang, Readability analysis of privacy policies for large-scale websites: A perspective from deep learning and linguistics, in: IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles, SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta 2022, Haikou, China, December 15-18, 2022, IEEE, 2022, pp. 1748–1753. URL: https://doi.org/10.1109/SmartWorld-UIC-ATC-ScalCom-DigitalTwin-PriComp-Metaverse56740.2022.00249. doi:10.1109/SMARTWORLD-UIC-ATC-SCALCOM-DIGITALTWIN-PRICOMP-METAVERSE56740.2022.00249.

[18] P. Voigt, A. Von dem Bussche, The EU General Data Protection Regulation (GDPR), A practical guide, 1st ed., Cham: Springer International Publishing 10 (2017) 10–5555.

[19] S. L. Pardau, The California Consumer Privacy Act: Towards a European-style privacy regime in the United States, J. Tech. L. & Pol'y 23 (2018) 68.

[20] J. Manyika, S. Lund, W. DC, J. Bughin, Digital globalization: The new era of Global Flows (2016).

[21] R. Creemers, G. Webster, Translation: Personal Information Protection Law of the People's Republic of China — Effective Nov. 1, 2021, 2021. Last revised 2021-09-07.

[22] P. M. Schwartz, D. Solove, Notice & choice, in: The Second NPLAN/BMSG Meeting on Digital

Media and Marketing to Children, volume 7, 2009, pp. 1–6.

[23] J. Bhatia, T. D. Breaux, Towards an Information Type Lexicon for Privacy Policies, in: 2015 IEEE eighth international workshop on requirements engineering and law (RELAW), IEEE, 2015, pp. 19–24.

[24] S. Guy, R. Schwitter, The PENG ASP system: architecture, language and authoring tool, Lang. Resour. Evaluation 51 (2017) 67–92. URL: https://doi.org/10.1007/s10579-016-9338-7. doi:10.1007/S10579-016-9338-7.

[25] N. E. Fuchs, First-Order Reasoning for Attempto Controlled English, in: M. Rosner, N. E. Fuchs (Eds.), Controlled Natural Language - Second International Workshop, CNL 2010, Marettimo Island, Italy, September 13-15, 2010. Revised Papers, volume 7175 of *Lecture Notes in Computer Science*, Springer, 2010, pp. 73–94. URL: https://doi.org/10.1007/978-3-642-31175-8_5. doi:10.1007/978-3-642-31175-8_5.

[26] WhatsApp, Privacy policy, https://www.whatsapp.com/legal/privacy-policy?lang=en, 2021. Effective: Janunary 4, 2021.

[27] V. Goranko, A. Rumberg, Temporal Logic, in: E. N. Zalta, U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy, Summer 2025 ed., Metaphysics Research Lab, Stanford University, 2025.

[28] N. Troquard, P. Balbiani, Propositional Dynamic Logic, in: E. N. Zalta, U. Nodelman (Eds.), The Stanford Encyclopedia of Philosophy, Fall 2023 ed., Metaphysics Research Lab, Stanford University, 2023.

[29] R. Schwitter, English as a Formal Specification Language, in: Proceedings. 13th International Workshop on Database and Expert Systems Applications, IEEE, 2002, pp. 228–232.

[30] N. E. Fuchs, Reasoning in Attempto Controlled English: Non-monotonicity, in: B. Davis, G. J. Pace, A. Z. Wyner (Eds.), Controlled Natural Language - 5th International Workshop, CNL 2016, Aberdeen, UK, July 25-27, 2016, Proceedings, volume 9767 of *Lecture Notes in Computer Science*, Springer, 2016, pp. 13–24. URL: https://doi.org/10.1007/978-3-319-41498-0_2. doi:10.1007/978-3-319-41498-0_2.

[31] G. Garzo, A. Palumbo, Human-in-the-Loop: Legal Knowledge Formalization in Attempto Controlled English, in: 13th International Symposium on Digital Forensics and Security, ISDFS 2025, Boston, MA, USA, April 24-25, 2025, IEEE, 2025, pp. 1–6. URL: https://doi.org/10.1109/ISDFS65363.2025.11011971. doi:10.1109/ISDFS65363.2025.11011971.

[32] A. Papaluca, D. Krefl, S. M. Rodriguez, A. Lensky, H. Suominen, Zero-and Few-Shots Knowledge Graph Triplet Extraction with Large Language Models, arXiv preprint arXiv:2312.01954 (2023).

[33] I. Matteucci, M. Petrocchi, M. L. Sbodio, CNL4DSA: a controlled natural language for data sharing agreements, in: S. Y. Shin, S. Ossowski, M. Schumacher, M. J. Palakal, C. Hung (Eds.), Proceedings of the 2010 ACM Symposium on Applied Computing (SAC), Sierre, Switzerland, March 22-26, 2010, ACM, 2010, pp. 616–620. URL: https://doi.org/10.1145/1774088.1774218. doi:10.1145/1774088.1774218.

[34] S. Gnesi, M. Petrocchi, Towards an executable algebra for product lines, in: Proceedings of the 16th International Software Product Line Conference - Volume 2, SPLC ’12, Association for Computing Machinery, New York, NY, USA, 2012, p. 66–73. URL: https://doi.org/10.1145/2364412.2364424. doi:10.1145/2364412.2364424.

[35] M. H. ter Beek, A. Fantechi, S. Gnesi, G. Lenzini, M. Petrocchi, Can AI help with the formalization of railway cybersecurity requirements?, in: T. Margaria, B. Steffen (Eds.), Leveraging Applications of Formal Methods, Verification and Validation. REoCAS Colloquium in Honor of Rocco De Nicola - 12th International Symposium, ISoLA 2024, Crete, Greece, October 27-31, 2024, Proceedings, Part I, volume 15219 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 186–203. URL: https://doi.org/10.1007/978-3-031-73709-1_12. doi:10.1007/978-3-031-73709-1_12.

[36] R. Sukthanker, S. Poria, E. Cambria, R. Thirunavukarasu, Anaphora and Coreference Resolution: A Review, Information Fusion 59 (2020) 139–162.

[37] A. Steen, C. Benzmüller, What are non-classical logics and why do we need them? An extended interview with Dov Gabbay and Leon Van Der Torre, KI-Künstliche Intelligenz 38 (2024) 17–23.

[38] A. Bilotta, M. Maggesi, C. Perini Brogi, L. Quartini, Growing HOLMS, a HOL Light Library for Modal Systems, volume 3904, 2025, p. 41 – 48. URL: https://ceur-ws.org/Vol-3904/paper5.pdf.

[39] C. Benzmüller, Faithful Logic Embeddings in HOL—Deep and Shallow (Isabelle/HOL dataset) (2025).

[40] C. Perini Brogi, M. Maggesi, Analysing collective adaptive systems by proving theorems, in: T. Margaria, B. Steffen (Eds.), Leveraging Applications of Formal Methods, Verification and Validation. REoCAS Colloquium in Honor of Rocco De Nicola - 12th International Symposium, ISoLA 2024, Crete, Greece, October 27-31, 2024, Proceedings, Part I, volume 15219 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 223–237. URL: https://doi.org/10.1007/978-3-031-73709-1_14. doi:10.1007/978-3-031-73709-1_14.

[41] M. Maggesi, C. Perini Brogi, Rigorous analysis of idealised pathfinding ants in higher-order logic, in: T. Margaria, B. Steffen (Eds.), Leveraging Applications of Formal Methods, Verification and Validation. Rigorous Engineering of Collective Adaptive Systems - 12th International Symposium, ISoLA 2024, Crete, Greece, October 27-31, 2024, Proceedings, Part II, volume 15220 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 297–315. URL: https://doi.org/10.1007/978-3-031-75107-3_18. doi:10.1007/978-3-031-75107-3_18.