Preventing disputes: preventive logic, law & technology

Stathis, G.

Chapter 5

Proactive Control Data

The Chapter addresses RQ4, which reads as follows:

> **RQ4:** *To what extent is it possible to generate quality Proactive Control Data to improve an Intelligent Contract?*

iContracts have many challenges, among which including the quality of data used. In our research we focus on generating and including quality Proactive Control Data (PCD) to improve iContracts. It is a novel research scope in literature. Currently, the legal system is more *reactive* than *proactive*, leading to high consequential legal costs. By shifting the focus to proactiveness, we discuss and improve upon available methodologies (the Bow-Tie Method and the Logocratic Method). Moreover, we examine PCD with the context of three technologies (Ontology Engineering, Software Engineering and LLMs) with the aim to demonstrate a higher degree of proactiveness in iContracts. Our research direction is threefold. First, we are able to generate PCD with the development of a prototype. Second, we show that impact of PCD on contract drafting is possible. Third, we show how the quality of PCD can be assessed and improved. The discussion highlights (1) the feasibility of the research with available technologies and (2) that its implementation depends on organisational considerations and resource allocation. From the results we may conclude that it is possible to implement our new ideas successfully.

The current chapter corresponds to the following publication:

Stathis, G., Biagioni, G., de Graaf, K. A., Trantas, A., and van den Herik, H. J. (2023a). The Value of Proactive Data for Intelligent Contracts. *World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), Intelligent Sustainable Systems, Springer Lecture Notes in Networks Systems (LNNS)*, 803:107–125

5.1 iContracts and Legal Prevention

iContracts are able to contribute to the reduction of contractual dispute resolution by helping to (1) minimise the likelihood of dispute resolution, and (2) reduce its complexity. The main idea is to help reduce operational expenses during the resolution of disputes [Stathis et al., 2023d]. Even though the literature clearly supports the benefits of iContracts for dispute resolution [McNamara and Sepasgozar, 2020], thus far no research has *measured* the extent to which iContracts contribute to that end. The reason is that researchers have not sufficiently focussed on *measuring* the proactive value of iContracts. To be precise, they have thus far not paid any attention to measuring the *explicit data* that contribute to the prevention of legal problems.

This Chapter aims to make the hidden data *transparent* and *explicit* by leveraging the EBTO, the primary ontology structure for managing contract risk in iContracts [Stathis et al., 2023b]. We start with an introduction of the basic concepts.

5.1.1 The Basic Concepts

In agreement with the EBTO terminology, we distinguish (1) *Proactive Controls* (hereinafter "Proactive Control Data" or "PCD", see Definition 5.1), which play an important role as soon as we have arrived at the identification of a (2) *Hazardous Event* and have produced an analysis of a (3) *Cause* [in the remainder of this article (1), (2) and (3) will be called "Proactive Data"].

*Definition 5.1 – **Proactive Data***

Proactive Data are a collection of data, which include Hazardous Event, Cause and Proactive Control data, that contribute towards the prevention of a Hazardous Event, within the context of the Bow-Tie Method.

Proactive Data *determine* the PCD necessary to prevent a contract from incurring legal risks. By measuring and leveraging PCD qualitatively and quantitatively, iContracts are able to maximise their value towards reducing the likelihood of dispute resolution, which is the main driver of consequential legal costs.

5.1.2 Towards Proactive iContracts

Our research scope follows the direction of the members of the school of PPL. They advocate for more *proactiveness* in contracting [Hietanen-Kunwald and Haapio, 2021].

In Chapter 2 we developed the Onassis Ontology, which provides deep insight into all data that can be generated with iContracts [Stathis et al., 2023d]. PCD forms part of such data. However, the *number* and *quality* of PCD that iContracts are able to generate for a more preventive automated contract is still unknown. More concrete insight into the quantity and quality of PCD that iContracts are able to generate is necessary. At that point, legal experts can be empowered with improved contract drafting. This may result in protecting contractors to a greater extent and consequently reducing reactivity in the field of contracting. As demonstrated in Chapter 3, the identification of PCD occurs during the risk analysis stage [Stathis et al., 2023b]. However, that research did not focus on any criteria to determine the *quality* of PCD. Hence, it is now necessary to investigate how to develop quality *assessment criteria* to measure PCD qualitatively. Following a qualitative analysis, the criteria can be measured quantitatively and will be able to impact contract drafting. A difficult point here is that in order to examine their impact on contract drafting, legal experts have divergent writing styles. Therefore, we are going to leverage LLMs [Brown et al., 2020] as a research methodology to reduce diversity in writing styles. LLMs present an opportunity to investigate the extent to which the Proactive Control-specific prompt engineering alters LLM's contract drafting for similar clauses and contexts in order to validate the impact of PCD.

5.1.3 Chapter Goals

All in all, our research goals are threefold. First, we explore whether the Onassis Ontology is able to generate PCD, by building a prototype validating (1) our ontology design, and (2) the generation of PCD. Second, we investigate the impact of the generated PCD on contract drafting via the use of LLMs to draft contracts, replacing the legal expert by a module with a higher degree of experimental accuracy. Third, we examine whether PCD can be qualitatively assessed so as to improve their quality.

The research in the Chapter aims at making progress by the following research activities:

1. introducing the value concept of Proactive Data for iContracts,
2. reporting on a prototype web application that uses the Onassis Ontology and the EBTO structure,
3. testing the use of LLMs as a methodology for reducing the variety of contract drafting styles,
4. measuring the PCD quantitatively in iContracts,

5. establishing qualitative assessment criteria for PCD,
6. supporting legal experts to improve the overall data management and its decision making during contract drafting within the context of iContracts for the purposes of reducing the likelihood of contractual disputes, and
7. investigating a direction for measuring the value of Proactive Data for iContracts.

5.1.4 Research Question 4

The foregoing discussion leads us to the following RQ4:

> **RQ4:** *To what extent is it possible to generate quality Proactive Control Data to improve an Intelligent Contract?*

5.1.5 Research Contribution

The Chapter shows how it is possible to promote the proactivity of iContracts by (a) identification, (b) quality evaluation and (c) automated analysis of PCD. The main prevention of data under the EBTO contributes towards a general prevention. The identification of PCD occurs with the use of argumentation. The quality of argumentation can be evaluated further with *argumentation theory*. With the advances of technology, such fast evaluation and a proper production of arguments, it has become possible via LLMs to produce front-ranked results. Our research shows the extent to which such automation is able to reduce the that an expert needs to produce quality PCD.

5.1.6 Research Structure

To answer RQ4, we structured the remainder of the Chapter as follows. In Section 5.2, the relevant literature is described. Section 5.3 presents the research methodology, which includes the use of a survey. Then, Section 5.4 describes our field work and formulates the results of our investigations. Thereafter, Section 5.5 discusses them. Finally, Section 5.6 provides an answer to RQ4 and yields our Chapter conclusion.

5.2 Relevant Literature

The section is structured into six parts. Subsection 5.2.1 introduces the literature on the topic of Preventive/Proactive Law. Then, Subsection 5.2.2 presents sources on the intersection of PPL and Legal Technology. Subsection 5.2.3 shows the relevant sources on Proactive Control Data. Subsection 5.2.4 discusses the qualitative assessment criteria options for PCD. Subsection 5.2.5 applies literature findings on LLMs as a contract drafting medium. Finally, Subsection 5.2.6 compares our finding with the literature on ontology engineering and linked data.

5.2.1 Preventive/Proactive Law

Chapter 2 already introduced to a sufficient degree the literature on PPL. Below we provide a summary to bring the text in line with the other five topics. The birth of PPL was in 1950, when Louis M. Brown introduced the concept of *preventive law* [Brown and Rubin, 1950]. Brown (1950) believed that preventive law concerns the cost difference between entering into and avoiding legal costs [Brown and Rubin, 1950]. He remarked that a complete avoidance of legal costs is not always possible; however, prevention is still an ever-present possibility. After preventive law took hold, it gave rise to two main schools of thought. One is *therapeutic jurisprudence* [Wexler, 2000], which is concerned with the health of legal subjects, and the other is *proactive law* [Haapio and Varjonen, 1998], which focusses on proactive contracting. During the past decade, the research of preventive law and proactive law started to converge, leading to the creation of the term Preventive/Proactive Law [Barton, 2016]. Recently, PPL is endowed with (1) the visualisation of legal information and (2) the effects of technology on PPL [Corrales et al., 2019a]. Last but not least, while risk management was growing as a field of study and practice around the year 2000, the field of Legal Risk Management (LRM) emerged [Iversen, 2004]. Gradually, the connection between LRM with PPL was established [Mahler and Bing, 2006].

5.2.2 Preventive/Proactive Law and Legal Technology

Susskind was the first to notice the relation between preventive law and technology [Susskind, 1996]. PPL researchers reinforced that notion by stating that the emerging technological culture is largely compatible with the assumptions underlying PPL [Barton, 2016]. Hence, they are proposing a re-design of the legal system, which reconsiders the relation between law and society, to guide a reform of law in a technologically-based society [Barton et al., 2016]. They are also questioning and investigating the effects of new technologies for PPL and legal design [Corrales et al., 2022]. Most of the research on PPL and legal technology focuses on legal design and smart contracts. In particular, the focus is on: (1) the fundamental consideration by Susskind [Susskind, 1996], namely that with legal technology, the law can become *more proactive*, and (2) the usual work by PPL researchers related to *legal design* and *smart contracts*. To date, no further progress has been made on integrating PPL with legal technology.

5.2.3 Proactive Control Data

The concept of PCD is based on the Bow-Tie Method terminology, which is introduced in Chapter 3. The Bow-Tie Method helps in performing visual legal risk analysis [Haapio and Siedel, 2013]. It is extensively used in enterprises, projects, and energy risk management. In such cases, the Bow-Tie Method is used for visualising risk in a holistic manner by taking into consideration *proactive* and *reactive* risk measures [1]. In this Subsection we only concentrate on the proactive risk measures, since they are directly relevant to the prevention of a hazardous event. Reactive measures play a mitigating role once a hazardous event has already occurred.

The Bow-Tie Method guides us through the *hazardous events* (henceforth sometimes '*event*'). For each *event* to occur there is at least one *cause*. Thus, each *hazardous* event has a *cause*. As such, knowledge on the *hazardous events* and *causes* is necessary before defining the *Proactive Controls*, of which the role is to minimise the likelihood of a *hazardous event* occurring.

PCDs help in identifying measures that reduce the likelihood for a hazardous event from occurring. The higher the quality of PCD, the less the likelihood that a hazardous event will take place. To identify the relevant PCD, it is necessary to characterise (1) a hazardous event, and (2) the sources that may lead to a hazardous event (which together amount to Proactive Data).

From a methodological perspective, in order to identify Proactive Data, the use of three methodologies is possible: (1) Scenario Planning [Amer et al., 2013], (2) Post-Mortem Analysis [Stålhane et al., 2003] and (3) Pre-Mortem Analysis [Klein, 2007, Adriaanse and van der Rest, 2020]. From a reasoning point of view, Proactive Data can be identified (whether one uses one of the three aforementioned methodologies or not) by applying investigations along the following line (a) deductive reasoning (also called *ingenuity*), (b) inductive reasoning (also called *ingenuity*), and (c) abductive reasoning (also called *intuition*) (together also known as *Inferential Reasoning*).

Successfully avoiding a hazardous event depends on the availability of prior Proactive Data [Niiniluoto, 1999], which at present is not systematically structured. Currently, deductive reasoning is modelled with expert systems, and inductive reasoning with machine learning, while abductive reasoning cannot be modelled yet [Larson, 2021]. Still, Van den Herik believes that intuition can be programmed [van den Herik, 2015, van den Herik, 2016]. However, as matters stand now, the programming of intuition is only in its experimental stage. A big step towards qualitatively equal results has been made by the develop-

[1] https://www.wolterskluwer.com/en/solutions/enablon/bowtie/expert-insights/barrier-based-risk-management-knowledge-base/the-bowtie-method

ment of Deep Learning (DL) [LeCun et al., 2015]. All in all, the last word has not been said. From 2018 to 2021 it was believed that DL would fully outclass other advanced techniques based on pattern recognition. Although, fully relying on (well chosen) pattern recognition was able to prove that DL process may overlook a winning possibility [2](which was a shock to public and researchers). Hence, DL is very good but not perfect.

For the moment we will accept that the automated identification of Proactive Data can be achieved with Expert Systems (ES), ML and DL. As an example of intuition programming, we point to the adjacent topic of *scenario planning*; within this area there exists an *intutitive* technology developed by Pandora Intelligence [de Kock, 2014]. Thus far, Pandora Intelligence relies on future scenario prediction based on historical and present data via technology that combines ES and ML [3].

5.2.4 Quality Assessment of Proactive Data

One way to assess the quality of Proactive Data is to examine their identification process, which may be viewed as argumentation schemes open to interpretation within the context of forming contracts. In the 1990s, Pierre Schlag examined the interpretative nature of constitutions, and developed a theory which may be useful in this direction, as long as we examine the nature of contracting within the sphere of constitutional legal theory. Schlag believes that interpretation is key to recognizing the ontological emptiness of constitutions (meaning conceptual vagueness) [Schlag, 1996], which also holds for contracts. Hence, inquiring about the nature of the ontology of a contract can end in a perpetual process. An interesting explanation is given by H.L.A. Hart, who argued already over sixty years ago that the meaning of law is generally *clear*, *certain*, and *stable* at its core, but less so at its penumbra [Hart, 1958]. This leads to *vagueness* and *open-texture*, meaning the abstract meaning of legal terminology, which is up to today an unresolved issue in legal literature [Soames, 2012, Escher, 2021, Culver, 2004].

As a consequence, legal experts have learned to deal with the law in non-ontological manners. Those manners may include technical, normative, or epistemic approaches [Schlag, 1996]. As a result, legal experts have learned to deal with the law from the perspective of *legal pragmatism* [Schlag, 1996]. Legal pragmatism means that in order to solve legal problems, a legal analyst should use everyday tools that come to hand such as precedent, tradition, legal text, and social policy [Farber, 1988]. As a result, the law leads away (1) from the ontological

[2] https://mashable.com/article/google-s-alphago-wins-final-go-game-against-lee-sedol

[3] https://www.pandoraintelligence.com

to the epistemic, then (2) from the epistemic to the normative, and nowadays (3) from the normative to the technical [Schlag, 1996]. Taking into consideration the context of ontological emptiness of contracts, legal arguments can hardly be absolute as the closer they reach the ontological nature of the contract, the larger the role of emptiness. Currently, however, as contracts have become more concrete—*though not necessarily more true*—the closer they have come to the technical aspects of the contract, the better we can assess them. Hence, the closer the identification of Proactive Data comes to technical contractual parameters (rather than ontological), the greater the degree of certainty about their quality. Below we discuss the content identification of Proactive Data, i.e., how it becomes possible (A) as well as we comment on the *defeasible nature* of argumentation (B).

A. Contextual Identification of Proactive Data

To identify Proactive Data, as mentioned above, it is imperative to apply inferential reasoning. Inferential reasoning is normally applied within the context of implicit argumentation. The *Logocratic Method* developed by Scott Brewer [Brewer, 2011] aims to explain the nature and two main uses of arguments. (1) One can evaluate an argument to determine the degree to which premises provide evidential support, by examining how much support do they provide for inferring that the the conclusion is true. (2) One can also evaluate an argument to determine the degree to which its premises provide agonal support for its conclusions (that is, whether the argument is strong in some one more more fora of arguments contests, such as in litigation; by definition, this is an assessment of the argument's agonal virtue.). As Brewer contends (drawing upon John Dewey), the LM is a system of analysis where *"it is a whole whose wholeness is particularly tied to the interrelations between its parts; it has elements that have some independent existence; those elements have formally specifiable relations and the relations form a structure"* [Brewer, 2022].

B. Defeasible Arguments

The highest possible degree of the evidential virtue of an argument (see 5.2.4(A) above) is validity [Walton, 1996]. Validity is the property of an argument such that, whenever all the premises of the argument are true, the conclusion must also be true. Not all arguments have this degree of evidential virtue, that is, not all arguments are valid [Walton et al., 2008]. One may identify a genus of arguments that do not have the property of validity. One species in this genus is arguments that, although they an be fairly interpreted as deductive arguments, nevertheless are invalid (one example is arguments that commit the fallacy of "denying the antecedent," reasoning from 'If P then Q' and 'not-P' to 'not-Q') [Brewer, 2018]. The other species in this genus are defeasible arguments [Brewer, 2018].

As explained in Logocratic terms: *"A defeasible argument from premises 1–n to conclusion h is one in which it is possible that the addition of some premise(s), n+1, to 1–n, can undermine the degree of evidential warrant that premises 1–n provide for h. As this definition indicates, the only kind of argument that is indefeasible is a valid deductive argument."* [Brewer, 2018]

Due to the ontological emptiness of contracts, there will always be premises which can only be examined probabilistically. Bart Verheij demonstrated how the representation of defeasible arguments within the context of the law is possible and how the benefits of this practice are helpful for improved argumentation [Verheij, 2003]. Larry Simon contends that *"as we confront the multiple language-meanings permitted by many of the open-textured provisions of the Constitution, the only apparent standard we can bring to bear in evaluating competing arguments for one or another interpretative methodology (...) is the extent to which they promote a good and just society"* [Simon, 1985].

Our conclusion is that when following the argumentative analysis of Proactive Data, their value ultimately depends largely on the extent to which they contribute to the prevention of a hazardous event. Although, we need to be aware that in making assessments of risk in preventive abduction, one relies on explanations (abductions) that in turn rely on inductive generalizations and specifications, and all types of induction are defeasible.

5.2.5 Large Language Models and Contract Drafting

Recently, LLMs emerged in the field of Natural Language Processing (NLP), enabling transformative advances in diverse applications such as machine translation, sentiment analysis, and text summarisation [Brown et al., 2020]. By leveraging vast amounts of training data and employing advanced neural architectures, such as the transformer [Vaswani et al., 2017], LLMs have demonstrated remarkable proficiency in both analytical and generative tasks. The three main types of LLMs based on transformers are the following [Cai et al., 2022]. First, Encoder-Only, such as BERT [4], that utilises only a transformer encoder to generate contextual relations. Second, Encoder-Decoder, such as BART[5], which introduces a combination in using an encoder to process input text and an decoder to process ouptut text. Third, Decoder-Only, such as GPT [6], which uses only a decoder to produce contextually relevant outputs based on given prompt.

[4] `https://github.com/google-research/bert`
[5] `https://huggingface.co/docs/transformers/model_doc/bart`
[6] `https://openai.com/chatgpt/`

The two general uses of LLMs are analysis and generation. The main challenge with analysis is classification, while the main issue with generation is language. Our research is mostly concerned with classification or the conversion of text into knowledge structures, such as ontologies. For our purpose, encoder-only models are very relevant due to the high effectivity in domain-specific knowledge representation and the translation of abstract legal language into knowledge structures [Limsopatham, 2021]

A notable difference between encoder-only models and models that involve decoding, is that the encoder models do not suffer from hallucinations. In fact, it is possible to reduce hallucinations during decoding by focus on encoding. To further improve accuracy, using fine-tuned decoder models is also possible. To that end, three methods that can be used to reduce inaccuracies and improve factuality, are Retrieval-Augmented Generation (RAG), TruthfulQA and TruLens. RAG combines retrieval models with generative models to enhance accuracy and relevance of generated text by grounding it on knowledge structures [Wu et al., 2024, Es et al., 2023]. TruthfulQA is a benchmark to evaluate the accuracy and truthfulness of LLM output [Lin et al., 2021], and TruLens is a framework designed to track and explain ML decisions via interpretability and transparency [Datta et al., 2022].

In relation to contract drafting, employing fine-tuning techniques would allow a decoder to generate contextually relevant and legally accurate language, for enabling the automation of contract drafting and reducing the time and cost associated with manual contract creation [Chalkidis et al., 2020]. Following this line of action, incorporating established legal principles via encoding can support the decoding in generating text that is comprehensible for specialised human objectives [Nay, 2023]. Essentially, the higher degree of classification by means of encoding, especially for legal tasks which are vastly complicated, can help improve the outcomes of a decoder. To that extent, the relevance of ontology engineering is high and can be leveraged so that a decoder algorithm attains a higher degree of classificatory accuracy and improved encoding performance vi ontology-based encoding.

Beyond algorithmic considerations, deployment parameters should also be taken into consideration. Ensuring the security and privacy of sensitive legal data during the training process is important (such privacy concern mostly cloud-based systems and not necessarily on-premise systems). This necessitates the implementation of secure and privacy-preserving machine learning techniques [Abadi et al., 2016]. Additionally, the explainability and interpretability of LLM-generated content are critical concerns, as legal professionals must be able to comprehend and justify the rationale behind the generated text [Arrieta et al., 2020].

5.2.6 Ontology and Linked Data

An *ontology* refers to a formal domain model in which concepts and relationships between concepts are described [López et al., 2012]. The classes and relationships in an ontology can be used for organising contract, risk, and proactive control data in a contract definition. Each distinct ontology class and relationship has properties and descriptions that explicitly define their meaning (i.e., *semantics*), allowing different possible contract users (legal experts, contractors, laymen, automated software systems and databases) to interpret them consistently and unambiguously. Relationships in an ontology allow its users to see how contract details (e.g., scope, contractors, questions, signature), risk, and proactive control instances in the text of a contract are interrelated; for example, "*Contract X has risks Y and proactive control Z*", and thereby the relationships will improve traceability between contract data. The instantiations of an ontology, the actual contract text, contractors, scope, signature, risk, and proactive controls, can be stored as triples (subject, predicate, object: *"contract1 has_id 1"*, *contract1 rdf:type Contract*, *contract1 hasRisk risk1*, etc.). As a result, they can be generated, processed, and accessed systematically and assist encoding models with improving analysis results via improved classification.

Ontologies and linked data may positively influence contract automation by offering structured representation of contractual concepts that may facilitate contractual interpretation and connection with multiple legal sources, demonstrating benefits for either automated compliance or reasoning [Palmirani et al., 2018, Athan et al., 2013]. Regarding risk management, (1) ontologies contribute to standardisation and (2) linked data enable integration with information across data-bases [El-Ghalayini, 2017, García and Gil, 2020]. The closest work to proactive controls, in accordance with the bow-tie method, lies with ontologies aiming to formalise the bow-tie terminology and linked data concerned with identifying hazards across multiple industries [Bloem and Reniers, 2019, Koren et al., 2021].

5.3 Research Methodology

This Section presents the methodology of our research. The methodology concerns four main topics: (5.3.1) using a case study for iContracts Proactive Data, (5.3.2) the development of an iContracts prototype , (5.3.3) the development of an LLM experiment, (5.3.4) the application of the LM when designing data for Proactive Data and (5.3.5 the development of a quantitative experiment.

5.3.1 Case Study

Our case study will focus on an agreement between a freelancer and a client (contractors). It is motivated by our previous case studies as seen in Chapters 3 and 4. The legal expert will (a1) define the scope of the agreement, (a2) conduct the risk analysis, (a3) define the legal questions for the two contractors, and (a4) visualise the risk analysis next to the questions. Then, the contractors will (b1) answer the legal questions and (b2) the legal expert may process potential modal information before the contract is generated, and (b3) will send the result to the contractors. The case study will concentrate on a specific risk, i.e., the payment risk. An example of applying the EBTO to a payment risk case study has already been presented in Chapter 4 [Stathis et al., 2023c]. Its visualisation is accessible via GitHub [7].

5.3.2 Prototype Development

We have built a prototype web application that uses the Onassis Ontology and EBTO structure to guide users in identifying legal risks and proactive controls during the negotiation and generation of a contract [8]. For a visual overview of the user interface and the user interaction we refer to [9]. The prototype contains several web pages with input forms and interaction elements to interactively draft a freelancer contract, based on a text template and contract-specific questions. After the freelancer and the client have answered the contract-specific questions, a legal expert uses those answers to fill in the text in a template contract. The expert may add legal risks and possible proactive controls next to a visible predefined set of specified risks. Additional questions can be interactively asked to both the client and the freelancer. These questions can be about the (new or predefined) risks, proactive controls, or about initial questions on the contract to which client and freelancer provided conflicting answers, e.g., about the milestones or the payment, or about the need for negotiation/mediation. Finally, the questions, answers, risks, controls, and legal text are stored as data consisting of semantic subject+predicate+object triples (in Turtle *.ttl format) specified according to the Onassis Ontology and EBTO (or 'model'). We explain the stored data, which help in isolating PCD, in more detail with examples of data actually generated by the prototype in the Results Section in 5.4 [10].

[7] https://github.com/onassisontology/onassisontology/b\lob/main/img/Visualisation.png

[8] https://github.com/onassisontology?tab=repositories

[9] https://github.com/onassisontology/icontracts-front-end/blob/main/README.md

[10] See Appendix and Github: https://github.com/onassisontology/icontracts-back-end

5.3.3 Large Language Models Experiment

According to Chapter 2, text generation can be applied to minimise the involvement of the legal expert during the contract drafting process [Stathis et al., 2023d]. Instead of using legal experts as an experimental subject, who have *inconsistent* contract drafting styles, we are going to leverage ChatGPT, which has a more consistent contract drafting style and which is more measurable for research purposes.

ChatGPT is an AI chatbot developed by OpenAI and launched in November 2022 [11]. It is built on top of OpenAI's GPT Plus family of LLMs and has been fine-tuned using both supervised and reinforcement learning techniques [12]. ChatGPT can be used to generate text data, which includes drafting contracts. Two limitations decoder transformers are facing relevant to our study are as follows. (1) The training data is vast and perhaps over-exhaustive, which means potentially PCD are already fed in the algorithm (however, we are not in a position to know that). (2) Decoder models lack knowledge-driven intuition which would otherwise be leveraged by a legal expert to draft a contract with greater safety based on PCD.

The command we provided to ChatGPT *without* explicitly mentioning PCD reads as follows: "Write a payment clause for a freelancer contract." The command we provided to ChatGPT that *explicitly* mentions PCD reads: "Write a payment clause for a freelancer contract that includes PCD1, PCD2, PCD3, etc." In order to measure the content differences we used open-source text comparison technology [13].

5.3.4 Logocratic Method

The application of the Logocratic Method on the qualitative evaluation of PCD is possible for so long as we see an identified PCD as an argument. According to the LM, there are four modes of inference: (1) deduction, (2) induction, (3) abduction and (4) analog-duction. Of these four, abduction is the "first among equals" [Brewer, 2022] in that abduction plays a role *within* virtuous inductions and *within* virtuous analog-ductions and there are several types of abduction that operate within legal reasoning, including legal abduction, interpretive abduction and rule abduction. Abduction also plays a vital role in risk analysis. In this thesis, I identify a new type of explanatory argument which I will refer to as *preventive abduction*.

[11] https://en.wikipedia.org/wiki/ChatGPT
[12] https://en.wikipedia.org/wiki/ChatGPT
[13] https://www.diffchecker.com/text-compare/

According to the logocratic explanation of abduction, its "meta-abduction", an abduction has two main components: (1) a four step pattern of inference and (2) the concept of a *point of view* (=meaning the set of judgments, methods and axiological aims within which every abduction takes place), which is *explanatory* in nature [Brewer, 2020, Brewer, 2022]. Below we attempt to represent only the essential elements of Brewer's theory of abduction for our case study [Brewer, 2022].

Premise ε_1	Θ [some phenomenon to be explained, the *explanandum*]
Premise(s) $\varepsilon_{2n\text{-}m}$	For each candidate Φ_i, '$\Phi_i \surd \rightarrow \Theta$' is true. ['$\Phi_i \surd \rightarrow \Theta$' is the *plausibly serviceable explanation conditional*, read as 'if explanans Φ_i were true or otherwise warranted, it would provide a *plausibly serviceable explanation* of Θ.']
Premises ε_3 and ε_4	For candidate Φ_n, '$\Phi_n \surd\surd \rightarrow \Theta$' is true. ['$\Phi_n \surd\surd \rightarrow \Theta$' is the *most serviceable explanation conditional*, that is, the one member of the set of proposed explanations that, in the abductive reasoner's judgement, is the *most serviceable among the set of plausibly serviceable explanations*. This step is constituted by the disconfirmation of all of those plausibly serviceable explanations identified in the articulation of Premise(s) $\varepsilon_{2n\text{-}m}$ until one, Φ_n is 'left standing' to be endorsed as *the most* serviceable explanation. The Logocratic explanation of abduction, like the accounts that regard abduction as inference to a single best explanation among those that are plausible, regards all abductions as instances of what some philosophers refer to as *contrastive inferences*.[67]]
Conclusion h	Φ_n [Φ_n is the explanation identified in step ε_3 that is settled on as *the* explanation, the *explanans* of the explanandum.]

Figure 5.1: Abstract Structure of Abduction

On a practical level, to apply the LLM we start by identifying arguments (Enthymeme) and then by fairly formally representing them; whereas we judge as matter of interpretation, whether the representation is based on as deduction, induction, abduction or analog-duction [Brewer, 2022]. According to LM's distinctive conception of the concept of the enthymeme, an enthymeme is a proposition whose logical form is not explicit in its original mode of presentation [14]. Consider for example what we may call the Socrates text: *Socrates is a man, so he will die.*

The first interpretative decision to be made is whether there is an argument at all, even though its logical form is not explicit (that is, it is an argument enthymeme) [Brewer, 2022]. Suppose we do believe that there is a set of premises to support a set of conclusions. If that is our interpretative judgment we might represent the argument as follows, using E(n) to label a premise and H(n) to represent the conclusion [Brewer, 2022].

Example for deduction

1. E1 = All men are mortal
2. E2 = Socrates is a man
3. H = Socrates is mortal

Example for induction

1. E1 = X1 is a man and X1 is mortal
2. E2 = X2 is a man and X2 is mortal
3. ...
4. E1000 = X1000 is a man and X1000 is mortal
5. H1/E1001 = All men are mortal
6. E1002 = Socrates is a man
7. H2 = Socrates is mortal

[14] Brewer discusses the distinct conception of the enthymeme and compares it to other conceptions in [Brewer, 2022]. The two most important differences between the LM conception and other conceptions is that on the LM conception both rules and arguments can be enthymematic, whereas under one version of the classical conception only arguments can be enthymematic. The other important difference is that on the classical conception of the enthymeme only syllogistic arguments are understood to be enthymematic, while on the LM conception any argument in any of the four modes of inference can be enthymematic.

The precise point of view we take when interpreting or explaining a PCD as argument is the point of view of prevention, driven by the unique axiology of prevention which requires a PCD to prevent a hazardous event from occurring. Thus we may in fact label our argument evaluation of PCD, or any other argument evaluation aimed at prevention, as *preventive abduction*. Central to preventive abduction is the role of deduction and induction, since, according to Brewer, any type of abduction is based either on deduction or induction. The evaluation of a deductive preventive abduction is based on the indefeasible evidential strength of the premises offered for the conclusion, while the evaluation of inductive preventive abduction is based on the defeasible evidential strength of the premises offered for the conclusion, which may be examined according to probabilistic criteria, where probabilities sum to less than 1. By evidence we may refer to any type of propositional supporting data for the formation of a PCD, relative to propositional articulation of specific hazardous events and risk causes. Moreover, provided that more domains than just the legal domain study proactive data, analog-duction may be used as well for the evaluation of a preventive abduction. An analog-duction may examine similar or dissimilar elements between at least two proactive controls in parallel domains, and derive insights to evaluate the relative strength of a preventive abduction, or even improve it.

5.3.5 Quantitative Experiment

Our quantitative experiment investigates the reliability of the Onassis Ontology and the EBTO via the application of the Cohen's Kappa (CK) coefficient [Warrens, 2014]. CK coefficient helps us measure inter-rater reliability on the basis of quantitative data [Vieira et al., 2010]. CK is one of multiple quantitative methods used in examining usability design [van Kuijk and Staats, 2012, van Kuijk et al., 2019]. It shows the agreement percentage between two independent raters concerning a body of data [Scholtes, 2024]. Cohen's Kappa is similar to correlation coefficients, it can range from 0 to plus 1, where 0 represents the amount of agreement that can be expected from random chance, and 1 represents perfect agreement between the raters [McHugh, 2012]. Essentially, CK uses statistics to measure the degree of agreement between two independent raters who are tasked with classifying data. To calculate the CK coefficient, the rater's data are collected on a contingency table, and then the following formula is used as shown in Figure 5.1.

- **P***o* refers to the observed agreement (how many times raters agree), and
- **P***o* refers to the expected agreement by chance, calculated on the basis that each rater will assign each data point randomly.

$$\kappa = \frac{p_0 - p_e}{1 - p_e},$$

Figure 5.2: Cohen's Kappa Coefficient Formula

Our quantitative experiment is designed with data from the Contract Understanding Atticus Dataset (CUAD) [15]. After downloading the CUAD dataset, we randomly selected relevant clauses as data. Such clauses act as paragraphs, within sections within a contract in accordance with the Onassis Ontology. The key question was: *what are the variables that update the paragraphs based on a contractor conversation?* The extraction of such variables, according to the Onassis Ontology, happens after a contractor answers a question. However, not all variables derive from straight-forward question-answering. In more complex cases negotiations take place.

For this quantitative analysis we generated conversation data via ChatGPT based on a clause in order to mirror a real-life negotiation. The ChatGPT prompt was as follows: "create a fictional negotiation between two contracting parties regarding the following clause in 100 words (please do not provide an amended clause - I only need the negotiation) and then separately clarify the main risk of the clause in 25 words: [insert CUAD Clause]"

Thereafter, we asked the raters five questions:

1. Do you identify variables in the conversation to update the clause? (Yes / No)
2. Are the variables equal to or more than three? (Yes / No)
3. Is the likelihood of the risk happening high? (Yes / No)
4. Is the impact of the risk happening high? (Yes / No)
5. Are the identified variables sufficient proactive controls to prevent the risk from occurring? (Yes / No)

The particular structuring of the questions was subject to 7 rounds of trial and error provided that initial questions were not showing a high agreement (in fact, initially 3 to 4 rounds showed 0.0 agreement). Eventually, as you will see in the results (Section 5.4.3) we only arrived to a sufficient agreement (above 70

[15] https://www.kaggle.com/datasets/konradb/atticus-open-contract-dataset-aok-beta/data

percent) only after reducing the options of answers (either yes or no), reducing the words in the negotiation and reducing the complexity of the questions.

In total, there was 10 clauses with each one relating to a risk, section and a conversation, presented to the raters as an excel file (they are all accessible in Appendix 4A). We had two raters named Rater 1 and Rater 2. Each rater provided a reply in a separate excel sheet. The final answers of the raters are provided in Table 5.1 (Rater 1) and Table 5.2 (Rater 2) below and are also accessible in the Appendices.

5.4 Research Results

The section presents our research and then the four results. First, the generation and isolation of PCD via the experiment is presented (5.4.1). Second, the impact of PCD on contract drafting based on the LLM experiment is shown (5.4.2). Third, the application of the LM on an example of Proactive Data is presented (5.4.3). Fourth, we calculate the CK coefficient (5.4.4).

5.4.1 Proactive Control Data Generation and Isolation

We designed and built a prototype web application that uses the Onassis Ontology and EBTO structure to guide users in identifying legal risks and proactive controls during the negotiation and drafting stages of a contract. The source code for the prototype web application is accessible via Github or the Appendices (see Appendix 4A [16]) [17] [18]. It contains the Docker specification for installing, running, and hosting the website, including a docker-compose script that can be used to start the front-end, back-end, and underlying database in a single command. Several screenshots stored in the readme.md of the front-end repository on Github and in the Appendices (see Appendix 4A [19]) [20] show the user interface and exemplify the user interaction of the web application.

The prototype validates two main points. First, (1a) the development of iContracts based on the Onassis Ontology and EBTO structure is possible, (1b) the integration of APIs is possible. Second, that the extraction of isolated PCD is possible via the integration of the EBTO in the Onassis Ontology structure.

[16] https://github.com/onassisontology/onassisontology/blob/main/Appendices_PhD.pdf

[17] https://github.com/onassisontology/icontracts-back-end

[18] https://github.com/onassisontology/icontracts-front-end/blob/main/README.md

[19] https://github.com/onassisontology/onassisontology/blob/main/Appendices_PhD.pdf

[20] https://github.com/onassisontology/icontracts-front-end/blob/main/README.md

No.	Identify Variables	Variables ≥ 3	Risk Likelihood	Risk Impact	Sufficient Controls
1	Yes	Yes	Yes	Yes	No
2	Yes	Yes	Yes	Yes	No
3	Yes	No	Yes	Yes	Yes
4	Yes	Yes	Yes	Yes	No
5	Yes	Yes	Yes	Yes	No
6	Yes	Yes	Yes	Yes	Yes
7	Yes	Yes	Yes	Yes	Yes
8	Yes	No	Yes	Yes	No
9	Yes	Yes	Yes	Yes	No
10	Yes	Yes	Yes	Yes	No

Table 5.1: Rater 1

No.	Identify Variables	Variables ≥ 3	Risk Likelihood	Risk Impact	Sufficient Controls
1	Yes	Yes	Yes	Yes	No
2	Yes	No	Yes	Yes	Yes
3	Yes	No	Yes	Yes	Yes
4	Yes	No	Yes	Yes	No
5	Yes	Yes	Yes	Yes	No
6	Yes	Yes	Yes	Yes	Yes
7	Yes	Yes	Yes	Yes	Yes
8	Yes	Yes	Yes	Yes	No
9	Yes	Yes	Yes	Yes	No
10	Yes	Yes	Yes	Yes	No

Table 5.2: Rater 2

5.4.2 Impact on Contract Drafting

The experiment we conducted with ChatGPT demonstrated that the PCD-specific prompt engineering influences the generation of text by altering its contents by more than ninety (90) percent. The alterations in the text included ten (10) content removals and fourteen (14) content additions. It shows that the content (including wording and grammar) of the contract have been significantly altered, with its new version including more PCD-specific semantic information expressed in syntactically adequate sentences. The results can be accessed via [21].

The experiment validates the impact PCD has on contract drafting. The PCD-specific prompt engineering provides additional protection via the explicit inclusion of PCD-based clauses. Having validated the impact of PCD on contract drafting, the ChatGPT API (Application Programming Interface) can be integrated with the prototype web application for a higher level of automation.

Our experiment can be further explored within the context of recent advances with RAGAs and prompt engineering to improve the contextual quality of produced output. This occurs because as research advances, the contextual relations of textual ouput relative to specific prompting within specific knowledge based also increases as a result.

5.4.3 Quality Assessment

To apply the LM to the Proactive Data of the visualisation of payment risk we used a specific example of a proactive control from the case study, with the purpose of demonstrating how it can be applied to more use cases. Essentially, interventions mimics a Retrieval-Augmented Generation Algorithm (RAGA) which combine retrieval models with generation models for more accurate and relevant text.

Our proactive control example is the *timeline*, which prevents payment risk by helping parties agree about a schedule for the (expectation of) delivery of milestones. The application of the LM to the timeline example can be accessed via [22].

This example demonstrates that the quality assessment of a proactive control is possible via all three modes of inference, including deduction, induction, and abduction. Moreover, it makes clear that both the Onassis Ontology and EBTO structure as well as the LM are based on First Order Logic (FOL), therefore the application of the LM can be engineered towards an ontology with the

[21] `https://github.com/onassisontology/onassisontology/blob/main/LLMexperiment`

[22] `https://github.com/onassisontology/onassisontology/blob/main/logocraticmethod`

purpose of achieving a higher level of automation. Due to the high reliance of the LM on the data required to validate the enthymemes, such automation might be preferred assuming the availability of data at a larger scale.

5.4.4 Cohen's Kappa Coefficient

The first step to calculating the CK coefficient is adding the ratings into a contingency table, which you can see in Table 5.4.

	Yes	**No**
Yes	39	2
No	2	7

Table 5.3: Contingency Table

After applying the CK formula on the contingency table we arrive to a **0.73** CK coefficient. A CK value of 0.73 indicates high degree of agreement between the raters. A rough guideline to interpret the CK coefficient values is as follows:

- Less than 0: Poor agreement
- 0.00 - 0.20: Slight agreement
- 0.21 - 0.40: Fair agreement
- 0.41 - 0.60: Moderate agreement
- 0.61 - 0.80: Substantial agreement
- 0.81 - 1.00: Almost perfect agreement

A CK of 0.73 falls under the "substantial agreement" range, suggesting there is sufficiently strong agreement between the raters. It is in fact a relatively good result showing there is consistency in the raters judgment.

5.5 Discussion

The discussion concentrates on (5.5.1) prototype feasibility, (5.5.2) the impact of Proactive Control Data on contract drafting, (5.5.3) Proactive Data, (5.5.4) the value of Proactive Data and (5.5.5) the inter-rater reliability on the basis of the CK coefficient experiment results.

5.5.1 Prototype Feasibility

The prototype shows that it is feasible to generate PCD in a linked open data format. It also shows that it is possible to link such data with LLM APIs, such as ChatGPT's API. Hence, technologically speaking, our research is practically feasible. One main obstacle is the lack of available data and friendly end-user interfaces that will support legal experts with the application of the EBTO and the LM. As a direct consequence, it becomes imperative to examine the development of such innovation more deeply in commercial settings, next to furthering the scientific development of the theory in academia.

The application of the LM is also technologically possible in particular for inferencing purposes on the ontology. Its implementation can be automatically executed via ontology reasoners. The LM will then use the inferencing system (as employed in FOL or any other form of deductive logic [Deontic, Modal, Propositional, Predicate]). Ontology Web Language - Description Logic (OWL DL) (i.e., the semantics used to build the Onassis Ontology and the EBTO) follows the description logic which is a branch of FOL. The inferencing system used in the LM mirrors the one followed by OWL, and there are also reasoners (e.g., PELLET, HERMIT) implemented that perform the exact same inferencing displayed by the LM [Singh and Karwayun, 2010]. The added value of the LM is the contribution it provides to quality assessment, beyond the assessment of consistency and contradictions that automated reasoners are able to achieve today. This is an innovative way to carry out an ontological quality assessment.

5.5.2 Impact of Proactive Control Data on Contract Drafting

During the application of PCD on prompt engineering it became evident that the generated text will change significantly. A higher level of detail, focused on the explicit PCD requested, is produced by the LLM. However, the quality of the drafted contract may not be ensured. Hence, a review by a legal expert is necessary. The examination of the quality of contract drafting remains a difficult task, even after the significant improvement of an automatically generated contract. To assess the quality of the drafted contract, it is also necessary to implement methods for assessing the quality of rules for generating the contract.

An example of such a method that can be applied in this case—and which also follows OWL DL—is the Calculemus-Flint Method which is being developed at TNO, the Netherlands Institute for Applied Scientific Research. The Calculemus-Flint Method makes explicit rule-based interpretations via an action-based interpretation instead of using a deontic-based interpretative framework. As a result, rules become explicit, explainable, and understandable from an action-oriented perspective (Actor A does X) instead of the traditional ruleoriented perspective (Agents Y should behave Y). The software, ontology, and documentation relating to the Calculemus-FLINT Method is accessible at the following GitLab repository[23].

5.5.3 Proactive Data Validity

The application of the EBTO to a case study has successfully generated Proactive Data. However, the quality of the generated Proactive Data is uncertain. The LM can help investigate the quality of the Proactive Data based on the examination of proactive data as argumentation schemes. Here, we admit that the application of the EBTO is a time-consuming process and the application of the LM further increases the time investment of a legal expert. This is the reason why (1) the implemented reasoners can be leveraged for assessing consistency and contradictions and (2) the quality assessment can be implemented only if necessary.

Moreover, the application of the LM does not necessarily guarantee that Proactive Data will be of high quality, provided that even the LM has limitations. All arguments are based on premises that can hardly be expressed in their totality provided they are subject to an infinite number of further underlying premises [Havenel, 2008]. Therefore, as the inferential structure deepens, complexity also increases and makes the task of representing reality in absolute terms eventually impossible.

Automating the quality assessment of Proactive Data is both relevant and possible with available technologies, assuming the availability of data. Hence, it becomes eminent to apply this innovation in practice and to perform further experiments with the tools at hand.

5.5.4 Value of Proactive Data

To measure the value of proactive data we can either approach it qualitatively or quantitatively. From a qualitative point of view, we see that Proactive Data is significantly influential in helping reduce the risk of disputes and minimising the subsequent legal costs. The goal of this Subsection is to demonstrate how its generation is possible, its impact, and how it can be further assessed and improved qualitatively.

Moreover, it would also be useful to examine their value from a *quantitative* perspective. That is possible by the numerical analysis of relatively good quality Proactive Data. Indeed, any quantitative measurement is quite limited due to a lack of appropriate data. It is therefore difficult to make an estimation that comes close to truth. We can only hope to arrive at a reasonable estimation

[23] `https://gitlab.com/normativesystems`

with follow-up experimental research by other researchers. Therefore, the analysis we will make for now is restricted to the *payment risk* case study and is as follows.

According to the freelancers union, the payment risk for a freelancer and a client today is seventy-one (71) percent [24]. Hence, according to the EBTO case study on payment risk, the likelihood for payment risk according to available data (see our quantitative experiment) seems to be 0.7[25]. The current question is the transformed to: how is this percentage affected once a freelancer agreement includes the identified PCD? To arrive at an acceptable estimation we would need to measure the extent to which each PCD reduces the likelihood for the hazardous event of non-payment to occur. Then we know for which test we require data. Hence, it becomes evident that the application of our results in practice is relevant and even more promising for further experimentation. By measuring the value of Proactive Data quantitatively, it becomes possible to conduct an economic analysis of (1) how an investment in Proactive Data reduces subsequent legal costs, and (2) how a reallocation of investments targeted towards dispute prevention can save millions of people from unnecessary legal costs. All in all, the more available data we can access, the better the nature of our quantitative analysis will be.

5.5.5 Inter-Rater Reliability

The CK coefficient of 0.73 is a relatively good result. When looking deeper into the questions and their context we note that despite a high level of agreement, the level is only relative to those specific questions and their abstraction level. For as long as we attempted to ask more complicated questions the level of agreement observed was lower. In fact, we found that the higher level of complexity of the questions, the lower was the level of the agreement. Moreover, the same applies to the data provided to the raters. Initially, we provided longer negotiations that lead to higher degree of rating discrepancies due to larger body of data, thus higher likelihood of difference in the answers of the raters. The more we arrived to reduced level of data complexity and question complexity, the easier it was to achieve more agreement. This only shows how much more work we need to put into experimenting with inter-rater agreement, by classifying further the data sets as well as the questions to as specific as possible categories for a particular domain in examination. The higher the specificity in data and questions, the higher the level of inter-rater agreement.

[24] https://blog.freelancersunion.org/2016/03/28/add-your-ious-worlds-longest-invoice/

[25] https://github.com/onassisontology/onassisontology/b\lob/main/img/Visualisation.png

5.6 Chapter Conclusion

This section presents: (5.6.1) the answer to the RQ, (5.6.2) further research suggestions, and (5.6.3) the progress of the research.

5.6.1 Answer to RQ4

The RQ4 is:

> **RQ4:** *To what extent is it possible to generate quality Proactive Control Data to improve an Intelligent Contract?*

Regarding RQ4 we provide the following answer. Proactive Data is *valuable* in shifting iContracts towards minimising the likelihood of hazardous events, including that of a dispute that leads to consequential legal costs. The extent to which Proactive Data impacts an Intelligent Contract depends on (1) their quantitative identification and (2) qualitative assessment, and (3) the use of relevant technologies that integrate the risk assessment (based on communication) data when a contract is generated.

Our contribution shows that the plain generation of PCD is possible with the available technologies. Moreover, it shows that PCD can be *quantitatively generated* with the application of the EBTO and can be *qualitatively assessed* with the LM to a certain extent. Attempting to achieve a higher degree of quality with the further application of the LM, runs the risk that the efficiency will be reduced. Even though the application of the LM does not guarantee "absolute truth", its application is highly valuable and preferable—or as the well-known statistician George Box stated: "all models are wrong, but some are useful" [Box, 2013]. Available technologies are already sufficient in implementing the findings of our research. Hence, the handling of Proactive Data in iContracts depends (a) on the specific application preferences of an organisation, (b) their resource allocation, and (c) issues related to technological innovation. Therefore, the answer to the RQ4 is that "(1) the generation of PCD is possible, (2) their impact on contract drafting is significant, and (3) the generation of quality PCD is sufficiently possible within organisational conditions".

5.6.2 Further Research

In relation to further research, three key research areas appear to be relevant. The first is the conditional abductive reasoning automation of Proactive Data on iContracts via the use of the LM within the context of LLM technology. Within this research scope it is possible to examine with a higher degree of certainty whether intuition is indeed implementable in the LM and to what extent (having in mind Van den Herik's research statement on the possibility of programming intuition [van den Herik, 2016]). The second is the quality assessment of LLM-generated contract text via the integration of the Calculemus-Flint Method with the Onassis Ontology. The third is the conducting of experiments to generate data to measure the quantitative value of Proactive Data. Since the quality assessment is higher in order of priority for improving iContracts, our follow-up research will focus on the first of the identified research areas.

In relation to the technological pathway forward, decoder only LLM models (such as BERT) are successfully extracting ontologies and other Knowledge-Based Structures (KBS) from (legal) textual documents. State-of-the-art research indicates that such KBS can be used successfully to keep the drafted (legal) text and conversations more factual (e.g. reduce hallucinations significantly) by either generating background and system prompts from these KBS or to convert components of the KBS or document collections (in combination with search) into vector representations that are merged into the conversational vectors (Retrieval Augmented Generation). Applying such experiments to derive parts of the Onassis ontology from legal documents or to use the Onassis Ontology to control the drafting or conversations from hallucinating, will help us add value to the validity and impact of our research.

5.6.3 Research Progress

The progress of this research is that we have indicated how it is possible to practically generate Proactive Data quantitatively as well as examine them qualitatively. Moreover, the research is a step forwards because it develops one of the first practical prototypes of iContracts that shows the generation of Proactive Data in linked open data format is possible. Additionally, the LM, which is still a developing method in literature, is applied to a proactive case study rather than its traditional application on litigation arguments, which follows the reactive nature of legal systems. The research also shows how it is possible to measure the value of Proactive Data for iContracts, which can help in scaling up iContract technology innovation in commercial settings, depending on architectural choices for improving the ratio of quality and efficiency. Moreover, the research is encouraging because it combines multiple FOL methodologies and technologies, and it shows that the automation of the aforementioned results is possible. Finally, our research introduces the value of Proactive Data and proposes a direction for measuring their value both technologically and economically.

CRediT Author Statement

Below I would like to give credit to all persons involved.

Stathis, G., Biagioni, G., de Graaf, K. A., Trantas, A., and van den Herik, H. J. (2023a). The Value of Proactive Data for Intelligent Contracts. *World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), Intelligent Sustainable Systems, Springer Lecture Notes in Networks Systems (LNNS)*, 803:107–125

Stathis, G.: Conceptualization, Methodology, Writing - Original Draft, Investigation, Visualization, Validation, Project Administration, Data Curation, Funding Acquisition, Writing - Review & Editing; **Biagioni, G.**: Methodology, Validation, Investigation, Writing - Review & Editing; **de Graaf, K.A.**: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Original Draft; **Trantas, A.**: Writing - Review & Editing; **van den Herik, H.J.**: Writing - Review & Editing, Supervision.