Ethical & Preventive Legal Technology

Georgios Stathis[1, 2*] and Jaap van den Herik[1, 2]

[1*]Institute of Tax Law and Economics, Leiden University, Kamerlingh Onnes Building, Steenschuur 25, Leiden, 2311 ES, the Netherlands.
[2*]eLaw - Centre for Law and Digital Technologies, Leiden University, Kamerlingh Onnes Building, Steenschuur 25, Leiden, 2311 ES, the Netherlands.

*Corresponding author(s). E-mail(s): g.stathis@law.leidenuniv.nl;
Contributing authors: h.j.van.den.herik@law.leidenuniv.nl;

Abstract

Preventive Legal Technology (PLT) is a new field of Artificial Intelligence (AI) investigating the *intelligent prevention of disputes*. The concept integrates the theories of *preventive law* and *legal technology*. Our goal is to give ethics a place in the new technology. By *explaining* the decisions of PLT, we aim to achieve a higher degree of *trustworthiness* because explicit explanations improve the level of *transparency* and *accountability*. It is an urgent topic for discussion in ethical AI research and regulation. We investigate the limitations of rule-based explainability for PLT. Our Problem Statement (PS) reads: *to what extent is it possible to develop an explainable and trustworthy Preventive Legal Technology?* After a literature review, we develop applications on case studies. The results describe the effectivity of PLT and its responsibility. The discussion delineates the relevance of PLT for LegalTech applications. On the ethical side, clearly explaining AI decisions is possible for small PLT domains, with direct effects on trustworthiness due to increased transparency and accountability.

Keywords: Artificial Intelligence, Explainability, Trustworthiness, LegalTech, Logocratic Method, Preventive/Proactive Law

1 Introduction

The connection between law and technology is instrumental. Laws regulate the design and application of *technologies*, and technologies influence the design and application of *laws*. To what extent is it possible to bring the two disciplines together? It is an interesting question, and the answer lies in the development of Artificial Intelligence (AI).

AI research has matured from investigating the structure of the domain and the need for heuristics with the help of increasingly intelligent technologies, such as Expert Systems (ES), Machine Learning (ML), Deep Learning (DL) and today, Large Language Models (LLMs). First, scientists (such as John von Neumann [1]) were concerned with trusting the fixed values of AI systems (intuitive acceptance). Gradually, they focussed on explaining the search directions (science). Today, we ask machines to *explain* their decisions in order for humans to be able to *trust* their line of reasoning (ethics). As a consequence, we expect machines to exhibit human-like intelligence. In hard science, we focus on trustworthiness, and we use explainability. In law, we focus on explainability and search for trustworthiness.

Considering the importance of explainability, law applications have become an exciting playground for experimenting with explanations and machine intelligence. AI and law have followed this trajectory since 1949, when Loevinger introduced Jurimetrics, i.e., using quantitative methods to analyse legal decisions [2]. In 1987, the first reasoner for explaining the reasoning supporting judicial decisions was created [3]. In 1991, Leiden University saw a remarkable Inaugural Address [4], in which the question "Can computers Judge Court cases?" was answered positively. Then, in 1996, Susskind predicted the shift from reactive facilities in the law (such as deciding on the resolution of a dispute) to proactive facilities (such as deciding on the prevention of a dispute) [5]. This line of research will dominate the next thirty years. Our research also follows the trajectory of connecting AI with proactive facilities in the law via the field of Preventive Law.

1.1 The Origins of Preventive Legal Technology

Aristotle Onassis, a Greek entrepreneur, was one of the most successful shipping tycoons of the Twentieth Century [6] [1]. Onassis (1900-1975) used Preventive Law to secure himself and his business from financial losses. The following narrative demonstrates how he *accomplished* this when his opponents planned to cause his business in Peru to suffer severe financial losses. The opponents were some shipping businessmen and a few law enforcers from the FBI and CIA (in the narrative provided below, they are called 'we') [2].

We knew that Onassis's fleet had the habit of fishing in illegal waters of Peru. Consequently, we planned with the Peruvian Government to seize his fleet. The Government sent ships and an aircraft, which bombed the waters around the factory ship. They strafed the ship with machine gun fire, forcing the boat back into the harbour with the captures. The Peruvians were very nasty about it, gave a huge fine and said they needed three million dollars to let Onassis's fleet go again. We believed we had killed

[1]To avoid all confusion with names and also to make them more familiar, we mention the first name of a person at the first occurrence together with the family name, when we believe it is supportive for the understanding of the text.

[2]Robert Mayhew, former FBI agent, CIA consultant and expert investigator acting on behalf of Onassis's opposition, and Dr. Ray Gambell, Secretary of International Whaling Commission, produced for the BBC, (1994), Aristotle Onassis' The Golden Greek', BBC Documentaries, min. 36:14 (for readability reasons, we slightly paraphrased the oral text).

the monster the night that this happened. Much to our amazement, we saw and learned that he had anticipated the whole thing by booking for disasters at Lloyd's of London. So, all in all, we watched him make a profit. He got 15 million dollars in insurance money and three thousand each day that he was out of the whaling operation. Thus, he made an enormous amount of money and was laughing all the way to the bank.

The narrative leads us to the origin of Preventive Law. The first author (G. Stathis) became aware of the concept through Onassis's lawyer, Tryfon Koutalidis [3] [7]. Koutalidis often provided short legal memoranda to Onassis or others working for him. Onassis would call for gym time when there were no pressing issues. During gym time, the businessman, lawyer, and other directors of his business examined hypothetical scenarios to secure themselves from potential risks that could arise. In this process, Preventive Law developed into a practice where legal risks were discussed and secured. Reading about Onassis may, therefore, stimulate an interesting research direction, viz. *the best way to resolve any problem is to prevent it from happening.* It is a straightforward and still challenging idea, and it motivated many researchers to investigate the power of Preventive Law.

Since its inception, Preventive Law has kept its use and applications the same; they did not seem to notice the advent of computer technology. Then, the disciplines of Law, Computer Science, Data Science, and Artificial Intelligence (AI) came together. Soon the world was facing the dawn of *legal technologies* and the arrival of *Intelligent Contracts* (iContracts), the latter being a successor of digital and smart contracts [4]. iContracts shows how it is possible to automate a contract based on risk and communication data, enabling the application of Preventive Law on contracts with the use of technology [8]. Of course, the application of Preventive Law is not restricted to contracts only. The remainder of this article aims to pave the way to the conceptualisation of *Preventive Legal Technology* (PLT). We will investigate how PLT can show a line of reasoning in an explainable and trustworthy manner, in compliance with Explainable AI (XAI) and Trustworthy AI (TAI) principles.

1.2 Towards Ethical and Preventive Legal Technology

Central to the discussion of AI is the topic of *trustworthiness* [9]. Lack of trustworthiness is a genuine concern for the ethical impact and unintended consequences of new AI technologies for society [10]. The European Union (EU) Guidelines call for lawful, ethical and robust AI [5]. Here, we note explicitly that despite the various blind spots for the ethics of AI [11], one of the main challenges of AI is that its decisions are not transparent, resulting in "black box" decisions [12].

XAI is the field of study investigating the explanation of AI system decisions[13]. XAI leads to TAI, aiming to increase society's trust due to higher transparency and accountability [14]. The concepts of transparency and accountability are vital in making AI more ethical, a central topic in the developing research on AI regulation [15].

[3] Most of the applications of Preventive Law concerning Onassis's business were to prevent financial risks. For this reason, and because Onassis educated him on applying Preventive Law, Mr Tryfon Koutalidis claims with a smile that he graduated from the 'Onassian University of Financial Contracts'

[4] One can find examples in the advanced courses for the Ministry of Justice and Security, the Public Prosecution and others: see Leiden Legal Technologies Program (LLTP), Leiden Centre of Data Science and The Centre for Professional Learning (LCDS and CPL), 2021

[5] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Researchers have noticed the general disconnection between levels of actual trust and trustworthiness of applied AI [16]. In order to nurture practical trustworthiness, researchers changed their focus to *transparency* and *accountability* [14]. They started contributing to measurable goals for the practical improvement of AI systems with direct implications on ethics. Meanwhile, other researchers were investigating AI's ethical and legal effects and were contributing to the development of the AI Act [17]. In the Netherlands, Maurits Kop is leading a group of researchers investigating how the development of Legally TAI (LTAI) by design is able to achieve higher ethical transparency and accountability [18].

The more profound insight into XAI and TAI enables us to examine PLT from two perspectives: (1) the *effectivity* of PLT (application of PLT in law) and (2) the *responsibility* of PLT (application of the law on PLT). Examining the effectivity of PLT helps determine whether PLT is a distinct field of technology. Due to the reliance of PLT on Proactive Data, PLT can be considered a type of Artificial Intelligence (AI). Provided PLT constitutes such a distinct field, its responsible implementation in society emerges as a topic for research in light of AI regulation.

Our motivation is to clarify how Automated Individual Decision-Making (AIDM) can become compliant under Article 22 GDPR [19]. AIDM is the process of deciding by automated means without any human involvement. The basis of these decisions can be on factual data, as well as on digitally created profiles or inferred data [6]. If AIDM includes explanations, then AI systems will be more trustworthy due to higher transparency and accountability. Consequently, organisations will be able to design AIDM that is ethical and legally preventive, which is beneficial for society because it reduces the appearance of legal problems and increases legal safety. Hence, our goal aims at *the intelligent prevention of disputes in an explainable and (legally) trustworthy manner*, in practical compliance with the ethical principles of *transparency* and *accountability*.

The preceding leads us to the following Problem Statement (PS):

To what extent is it possible to develop an explainable and trustworthy Preventive Legal Technology?

To answer the PS, we have partitioned it into three smaller Research Questions (RQs):

- **RQ1:** *what is Preventive Legal Technology?*
- **RQ2:** *to what extent is it possible to develop an explainable Preventive Legal Technology?*
- **RQ3:** *to what extent is it possible to develop a trustworthy Preventive Legal Technology?*

Before addressing the RQs, we would like to introduce our contribution. We aim to show (1) that Proactive Data, the primary PLT data, are identifiable in all categories of

[6] https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/#id2

LegalTech, (2) how to develop explainable Proactive Data with practical case studies, and (3) the legal and ethical implications of PLT in light of the AI Act.

To answer the PS, we structured the paper as follows. In Section 2, we describe the literature. Section 3 presents the three methodologies: fieldwork, case studies and applications. Then, Section 4 describes the investigations and states the results. Section 5 discusses those results and focusses on trustworthiness and ethical parameters. Finally, Section 6 answers the PS and the three RQs and provides our conclusion.

2 Literature Review

When a legal problem occurs, both (or all) parties subject to a legal agreement may experience costs, e.g., psychological pressure, legal and financial support, reputation damage, or loss of time. The allocation of costs will always affect one or sometimes both (or all) parties of the legal agreement, depending on the legal problem itself and how or when they can solve it. At least one party will incur (1) the *procedural* costs connected with the legal problem and potentially (2) the *liability* costs from the hazardous event that triggered the legal problem. Legal costs are frequently quite significant and primarily appear as dispute resolution costs [7], even in the world's most advanced jurisdictions [20]. For this reason, commentators opine that we need *new ways* to resolve and avoid disputes [21]. A primary reason why legal costs are so high is that the structure of legal systems invites dispute *resolution* and not dispute *prevention* [22]. All in all, legal costs give rise to a need for preventive law.

2.1 Preventive Law

Louis M. Brown was the first to introduce the concept in academic circles via his book *Preventive Law* in 1950 [23]. Preventive law was early on (1997) understood as a branch of law that endeavours to minimise the risk of litigation or to secure more certainty regarding legal rights and duties [24, 25]. During that time, researchers took into consideration the *proactive* parameters and the *reinforcing* parameters (i.e., the strengthening of legal rights and duties) of preventive law [23]. However, they only focussed on the avoidance of litigation (i.e., avoiding a trial is held in public courts) and not on other forms of dispute resolution, such as Alternative Dispute Resolution (ADR) (mediation and arbitration) and negotiations [26]. It is important to note that a dispute resolution procedure begins once there is a dispute.

- **Definition 1**: ***Preventive law*** is a *method* that minimises the likelihood of the occurrence of disputes, or in case they occur, it exploits their impact, and strengthens legal rights and duties.

[7] Usually, dispute resolution costs are a percentage of the liability costs. The best available research estimates that liability costs as a fraction of the Gross Domestic Product (GDP) are equal to 2.3 percent in the USA (429 Billion Dollars [US Chamber Institute for Legal Reform, (2018), Costs and Compensation of the US Tort System, instituteforlegalreform.com, p.1] in 2016) and 0.63 percent (Best available number derived from US Chamber Institute for Legal Reform, (2013), International Comparisons of Litigation Costs, instituteforlegalreform.com, p.2) in Eurozone (85.8 Billion Dollars [Calculated 0.63 percent of 2011 Eurozone GDP 13.6 Trillion Dollars as recorded at countryeconomy.com, (2011), Euro Zone GDP – Gross Domestic Product] in 2011). An economic analysis looking beyond GDP to socio-economic consequences is even more relevant to highlight dispute resolution costs with higher accuracy.

After its start in 1950, Preventive Law (see 1.1) gave rise to two primary schools of thought. They are *Therapeutic Jurisprudence* [27] (started in 1987) and *Proactive Law* (started in 1998) [28], which are concerned with the health of legal subjects and proactive contracting, respectively. Soon, however, the emphasis was on Proactive Law, established in 1998 by the academic and legal consultant Helena Haapio. Around 2010, the research by Haapio and Barton started to converge, leading to the creation of Preventive/Proactive Law. Recently, PPL has been concerned with the visualisation of legal information and the effects of technology on PPL [29, 30]. Last but not least, while risk management was growing as a field of study and practice around 2000, Legal Risk Management (LRM) emerged [31]. By 2010, the academics Tobias Mahler and Jon Bing established the connection between LRM and Proactive Law [32].

Even if someone attempts to apply preventive law to *prevent legal problems*, there is still a high likelihood for legal problems to *further developing* instead of diminishing. Brown introduced the first elementary framework for preventing legal problems [23]. Dauer later added a schematic approach to Brown's observations by stating that prevention can be applied at three intervals to manage legal risk, i.e., *before*, *during*, and *after* damage occurs [24]. Dauer's systematic analysis resulted in a matrix, which Barton called Dauer's matrix [22]. Barton then refined it considerably [33] [34].

Currently, most literature focusses on the *mindset* and *application* of Preventive Law. The authors do so in various domains, one of which is iContracts [8]. Indeed, there are exceptions, which only partially help people prevent legal problems. They come from the fields of *proactive law* [35] and *Legal Risk Management* (LRM) [36]. The reasons behind the lack of substantive methods or approaches to prevent legal problems need to be clarified.

2.2 Preventive Law and Legal Technology

Richard Susskind first noticed the relation between preventive law and technology. In his 1996 book, *The Future of Law*, he predicted that with technology, our approach to legal problems would switch from problem-solving to problem prevention through proactive processes supporting LRM [37]. Susskind believed that our legal system is subject to the paradox of reactive legal services, which proactive practices will replace with technology. In 2016, Barton reinforced that notion by stating that the emerging technological culture is mainly compatible with the assumptions underlying PPL [38]. In the same year, together with Haapio, they proposed a re-design of the legal system, which reconsiders the relationship between law and society, to guide a law reform in a technologically-based society [39]. They were also questioning and investigating the effects of new technologies for PPL and Legal Design [29, 30].

The relationship between *technology* and the *law* is multifaceted and dynamic [40]. Practically, there are two ways in which technology improves the efficiency of the law. The first occurs when a scientific development applies to law; this approach is *Computational Legal Studies* (CLS) [40]. The second is when changing market conditions leads to the development of technology that improves the *delivery of legal services*; this approach is *LegalTech* [41]. Sometimes, the two approaches are also combined [42].

Whether new technologies apply or not in the law, they usually create the need to develop new regulations; for the development and enforcement of new regulations, social, ethical and legal aspects are taken into consideration by regulators. In essence, the design of regulation depends on the content of the technology, which may apply in various fields. Some people refer to such regulation as *CyberLaw*, which affects legal areas such as free expression, privacy and cyber security [40]. However, CyberLaw only concerns information technology or technology related to networks and the internet [43]. In its foundational sense, technology refers to techniques that improve an existing task. Therefore, more widely, we may refer to the regulation of technologies by the law as the field of *law and technology*.

The primary purpose in developing technology applications (including CLS and LegalTech) is *effectivity*. Still, concerning technology regulation (including CyberLaw), the main goal is *responsibility*. In order to describe both perspectives in this article, we use the term *legal technology*. Today, legal technology's purpose and challenge is balancing effectivity and responsibility. It occurs because if the law is too responsible, then technology is not sufficiently effective, while if technology is too effective, the law is not sufficiently responsible. Balancing between the two ends is already hard, and it becomes more complicated as new technologies introduce new and increasingly complex challenges in the law [44].

Preventive law and legal technology are on their way to connecting substantially in the literature. Below, we provide our definition of the PLT concept.

- **Definition 2**: ***Preventive Legal Technology*** is a methodology concerned with the use of legal technology within the context of preventive law to promote the intelligent prevention of disputes.

2.3 Ethics and Artificial Intelligence

Ideas about modelling intelligent behaviour started in ancestry and developed throughout history [8]. Nevertheless, most researchers attribute the starting point of AI to Alan Turing in 1950 [45]. Since then, two main AI movements emerged: the scientific one and the futuristic one [46]. The scientific AI movement supports that formal reasoning is the basis of AI and is investigating whether intelligence can become artificial. The futuristic AI movement believes that intelligence will become artificial and will influence public opinion to accept that. While this dichotomy is still vivid, the state-of-the-art of AI is not yet able to prove how intelligence is programmable. Researchers support that AI today assists humans with ingenuity, contributing to intelligence, not intuition [46]. Still, researchers are investigating how to model intuition [47]. Here, we remark that despite the state-of-the-art observations, society is mainly influenced by the futuristic AI movement, expecting the replacement of carbon intelligence by silicon intelligence. However, the latter perspective undervalues linguistic complexity, which is the basis of human intuition [9].

Ingenuity and intuition are based on human types of reasoning. Inferential logic examines the types of reasoning. The three modes of inferential logic are deduction,

[8] See Greek Mythology (Talos, Pygmalion), Jewish Folklore (Golem), Paracelsus's Of the Nature of Things, Wolfgang von Kempelen's The Turk, Roger Bacon's brazen head, Mary Shelley's Frankenstein, Karel Capek's R.U.R., Samuel Butler's Darwin among the Machines, Aristotle's Organon and Francis Bacon's Organon.

[9] https://medium.com/the-sophist/wittgenstein-intelligence-is-never-artificial-51933315d1bd

induction, and abduction according to Charles Sanders Peirce [48] (analogduction is a fourth mode according to Scott Brewer [49]). The basis of intuition is abduction, and, admittedly, humans still do not know how to model it computationally [46]. The developments of AI follow, to a large extent, the developments in inferential logic. Thus, the modelling of deduction occurs via Expert Systems (ES) and induction by ML (and DL or LLMs), but AI so far has not modelled abduction [46]. The foundation for ES and ML are "normal" data [50]. The scientific field of XAI follows a similar path. The focus of most XAI models is on explaining the decisions of inductive models and those of deductive models to a lesser extent due to the often reduced decision-making complexity [13, 51].

2.3.1 Responsible Governance

The reliance of AI on human reasoning affects AI by the similar challenges it faces. Two of those challenges are the *explainability* problem and the *interpretation* problem [52]. The first one explains decisions (or abductions according to the LM). The second explains how people interpret the world (or interpretive abductions according to the LM). Consequently, even though society wants to be able to trust AI, they are still afraid of these challenges. In the first place, the world's governments wish to protect society from these fears. Therefore, the EU is attempting to take the lead in the movement of TAI [53]. The United States [10] and China [11] are also developing regulatory efforts.

As part of this movement, the field of responsible governance has been developed. This field investigates how to develop standards and processes to make AI safer. The discussion about TAI and the relevant responsible governance standards is still in development. The urgency for clarifying their content comes from the observed tendency in society for the development of AI applications [12]. Here, one of the social challenges of AI concerns the liability issues arising from their operation. Liability examines who is to blame if something goes wrong [54]. Fundamentally, it investigates who is at fault. Based on the concept of fault, several liability regimes have been selected, particularly for AI. The two largest categories are *fault-based liability* and *non-fault-based liability*. Researchers are investigating which regime or combination of regimes is appropriate for AI liability [55]. Here, we remark in passing that we consider liability measures in parallel with safety measures [56]. Usually, we see that with the introduction of new technologies, liability regimes adjust to correspond to the latest needs [57].

Nevertheless, regulating AI is challenging because we do not fully comprehend AI [58]. People are still debating about the appropriate definition for AI [59]. In the meantime, researchers are investigating the relevant ethical framework to guide any legal or social regulatory reform regarding AI [60].

TAI concerns (1) the trustworthiness of the AI system and (2) the trustworthiness of all processes and actors that are part of the system's life cycle [15]. That is quite

[10] https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-rules-for-artificial-intelligence

[11] https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

[12] European Parliament, (2017), Civil Law Rules on Robotics, European Parliament Resolution of February 07 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), European Parliament

substantial. Hence, we refer to the broad and deep analysis of trustworthiness, by which variable principles, from *reliability* and *accuracy* to *sustainability* and *democracy*, are included [61]. Such principles can guide the ethical and legally trustworthy design of AI systems via the rule of law by focussing on properties including *transparency*, *verifiability* and *explainability* [62].

Considering the difficulty of explaining or interpreting the decisions of AI systems, regulators are concerned about assigning liability to AI system decisions. Due to (a) the direct effect of AI on Law and (b) the liability of law concerns, some researchers argue that TAI is insufficient, but Legally Trustworthy AI (LTAI) is more important [63]. The same holds for PLT, seen as an AI technology.

While researchers and regulators worldwide investigate the safety of ethical principles in the design of AI systems, a more considerable challenge appears the inability to focus on divergent ways of protecting society from AI [14]. The primary motivator behind this challenge observed is the misalignment between levels of actual trust and the trustworthiness of applied AI [16]. Here, we note the contribution of Luke Munn, who proposes an alternative perspective for ethical AI, going beyond procedural issues on bias, transparency and discrimination. On a macro-level, he proposes the concept of AI Justice, which comprehends the creation of AI as a part of social systems, subject to the ethical values of the systems they created [14]. He calls for an inter-sectional ethical approach, which includes (1) diverse groups in designing AI systems, (2) the re-definition of outdated ethical concepts, and (3) ensuring that fundamental social inequalities are addressed [14]. On a micro-level, he brings forth two practical concepts for the design of AI: *transparency* and *accountability* [14]. Indeed, the latter two concepts will contribute to measurable goals for the practical improvement of AI systems. Furthermore, that is what we currently need.

Such a practical approach towards designing AI systems brings clarity in AI development and ethical auditing of AI algorithms [64]. For example, when large multinational organisations are subject to Ethics-Based Auditing (EBA), they face challenges including ensuring harmonised standards across decentralised organisations, demarcating the scope of the audit, driving internal communication and change management, and measuring actual outcomes [65]. The ethical design of AI, therefore, becomes the responsibility of (1) the organisations and (2) the social systems that create AI [65]. After all, ethics arise in the context of the socio-technical systems that create them [66]. Munn's inter-sectional ethics approach becomes feasible with the diverse inclusion of ethical practices within organisations, nudging towards the institutionalisation of ethics [67] and the re-evaluation of AI business practices [68]. Consequently, the evaluation of group values and interests as well as a fair (FAIR) comparison of personal with group values *becomes possible* [53].

2.3.2 Rule-Based Explainability

The XAI methods and techniques that have been developed in research so far span from interpretable ML models [69] and attention mechanisms [70] to visual explanations [71] and ethical variations [64] to the FAIR (Findable, Accessible, Interoperable, Reusable) model development [72]. One of the developing XAI techniques is rule-based explanations, which focus on symbolic reasoning and knowledge graph representation for developing human-readable model explanations [73, 74]. The most advanced method in literature to represent inferential reasoning in symbolic logic, applicable also to AI system decisions, is the Logocratic Method (LM) [75]. The LM aims to explain the nature of an argument [76].

LM shows that decision-making is, in essence, based on abductive reasoning, in which explanations play a fundamental role for interpretative and decision-making purposes [77]. Abductive reasoning is the foremost mode of inferential reasoning [77]. Its application requires the development of explanations about observed facts. Each explanation derives from a specific point of view. The relative strength of each explanation enables a relative level of trustworthiness. Since multiple agents may interpret an explanation and extension of trustworthiness from subjective points of view, subjective trustworthiness perceptions may develop (after all, when an agent is interpreting an abduction, the agent applies abductive reasoning at that exact moment). That is when the opportunity to qualify abduction's relative strength and trustworthiness presents itself.

So far, the LM has not been applied to rule-based XAI in literature. According to Brewer, there is a notable dichotomy between discovering and evaluating arguments. An argument evaluation becomes possible if an argument is identified, irrespective of its source. Hence, from an end-user perspective, what matters most in ruled-based XAI is the *ability to evaluate arguments* irrespective of its source and even if their discovery happens via the AI black box. Focussing on the ability to evaluate rather than discover complies with the notion of Hybrid Intelligence supported by leading TAI researchers in the European Union (EU) [78].

According to the LM, the process of evaluating arguments begins with an interpretive abduction. Hence, if the modelling of the LM takes place in AI systems, provided it contributes towards sufficiently valid evaluations, then the application of LM on AI contributes to making AI explainable and interpretable [79]. Eventually, the systematic evaluation of AI explanations and interpretations facilitates the evaluation of underlying values, principles and laws, contributing to greater trustworthiness [80].

2.4 List of Definitions

- **Definition 1**: ***Preventive law*** is a *method* that minimises the likelihood of the occurrence of disputes, or in case they occur, it exploits their impact, and strengthens legal rights and duties.
- **Definition 2**: ***Preventive Legal Technology*** is a methodology concerned with the use of legal technology within the context of preventive law to promote the intelligent prevention of disputes.
- **Definition 3**: **Legal Risk Management** is the management of risk (effect of uncertainty on objectives) related to legal, regulatory and contractual matters, and from non—non-contractual rights and obligations [81].

- **Definition 4**: ***Artificial Intelligence*** refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals [15].
- **Definition 5**: An **Intelligent Contract** (or **iContract**) is a contract that is fully executable without human intervention.[13]
- **Definition 6**: ***Trustworthy AI*** (TAI) is AI that has three components: (1) it should be lawful, ensuring compliance with all applicable laws and regulations; (2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values and (3) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm [15].
- **Definition 7**: ***Explainable AI*** (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms [14].
- **Definition 8**: ***Interpretable AI*** is a system that is possible to translate its working principles and outcomes in human-understandable language without affecting the validity of the system [79].

3 Research Methodology

Following the literature review presentation, the research methodology concentrates on three distinct approaches: fieldwork, case studies and legal framework application. First, fieldwork development focusses on understanding the essential data for AI-based PLT processing. The data are called *Proactive Data* and are based on our previous research [82]. Second, the basis for selecting case studies is Legalcomplex's list of LegalTech solutions [15]. After validating to what extent Proactive Data applies to the LegalTech solutions, we selected three case studies from the LegalTech applications. The aim is to develop Proactive Data explanations. Third, we clarify the AI-Act liability framework to apply it to the three case studies comparatively. For comparative significance, due to the focus of the AI-Act on high-risk AI, we based the selection of the three case studies on high-risk, mid-risk and low-risk case studies.

3.1 Field Work

According to the literature, the most advanced methods for preventing legal problems focus on contract risk management [35]. So, we ground our research on effectivity of *contract automation* and iContracts [8, 83]. iContracts introduces a hybrid approach between human and computer interventions that aim to achieve full automation with self-executing contracts [84]. They introduce state-of-the-art innovations in contract automation due to their compliance with Hybrid AI principles [85].

[13] https://bravenewcoin.com/insights/pamela-morgan-at-bitcoin-south-innovating-legal-systems-\through-blockchain-technology
[14] https://www.ibm.com/topics/explainable-ai
[15] https://legalcomplex.com

While investigating contract automation, we noticed two main challenges, which are the neglection of (1) the *communication process* preceding the drafting of contracts and (2) the *risk analysis* during the drafting of contracts [83]. We subsequently set out to validate our observations [8]. We used as a source of information **Legalcomplex**, the most extensive database on LegalTech solutions. We saw that out of the total registered 10,448 LegalTech solutions, 590 solutions (5.6 per cent) focus on contract automation. Out of these contract automation solutions, 51 (8.6 per cent) focus on contract communications and 50 (8.4 per cent) focus on contract risk [8]. Surprisingly, to the researchers and the owner of Legalcomplex, there was *no solution* focussing on *both* contract communications and contract risk, which led to the discovery of the Black Swan of this research.

Inspired by the observation, we set out to design and build the first *Intelligent Contract* (*iContract*) based upon the automation of communications and risk data. We did so in four steps. First, we designed the *Onassis Ontology*, an open-source ontology that demonstrates how it is possible to generate contracts by leveraging communications and risk data [83]. Second, we designed the *Enriched Bow-Tie Ontology* (EBTO), an extension of the Onassis Ontology, to indicate how it is possible to manage all types of risk—including contract risks—explicitly [86]. Third, we visualised the risk analysis conducted by a legal expert on EBTO for the contracting parties as end users and found that their *level of trustworthiness* increased from 1 to 7.9 on a scale from 1 (comparable to a user interface without risk visualisation) to 10 (comparable to a legal expert *explaining* risk) [87]. Fourth, we *validated* the Onassis and Enriched Bow-Tie Ontology structures [82].

It turned out that both ontologies are fully programmable and that the isolation of Proactive Data is possible. In this way, Proactive Data (or Proactive Control Data) help identify measures that reduce the likelihood of a hazardous event occurring [82]. In summary, the impact of Proactive Data is that (1) it can reduce the likelihood of a dispute occurring by improving contract drafting and (2) its quality assessment is feasible via the LM [75].

3.2 Case Studies

Given that the development of categorisation criteria is a complex process, we were given access to the categorisation used by Legalcomplex. As Legalcomplex has also been categorising and recording LegalTech solutions for several years, this categorisation is the best one available. Legalcomplex provided us with the categories of LegalTech listed in **Table 1**. Table 1 categorises six solutions: FinTech, WealthTech, RiskTech, LegalTech, SmartTech and CivicTech. Legalcomplex structures all collected company data so that all six categories fit within the giant umbrella of LegalTech. However, it is essential to differentiate between *specific* LegalTech solutions that focus on lawyers as end users and *general* LegalTech solutions that encompass a more comprehensive range of six categories. Two categories also include subcategories. The first is RiskTech with (1) Security, (2) Insurance, and (3) Governance, Risk, and Compliance (GRC). The second is SmartTech with (1) Image Recognition, (2) Audio Recognition, (3) Text Analytics, (4) Data Analytics, and (5) Automation. Table 1 includes specific descriptions for each category and subcategory—fourteen in total—and for the end users being top private companies. The number of categories is six, and of subcategories is eight. In total, there are fourteen (sub)categories.

The three case studies we selected are based on three LegalTech solutions found in Table 1. The framing of explanations assumes that an AI system advises an end-user based on Proactive Data.

- The **low-risk** solution concerns using Lemonade (RiskTech, Insurance) for purchasing car insurance.
- The **mid-risk** solution concerns using OpenAI (SmartTech, Text Analytics) for creating a construction plan.
- The **high-risk** solution concerns using Palantir Technologies (SmartTech, Data Analytics) for applying police force during a riot.

To represent the Proactive Data and their explanations, we will use synthetic data generated by ChatGPT. Synthetic explainable proactive data is generated based on a question that seeks explanation (explanandum) in compliance with the LM.

- For the **low-risk** case study, the explanandum is: What insurance should we provide to a client who bought his first car (s)he is 27 years old and has been caught drinking when (s)he was underage?
- For the **mid-risk** case study, the explanandum is: When deciding to build a tall building next to a residential area, should we add a net to catch people who may fall, at the expense of a better view of the surrounding area?
- For the **high-risk** case study, the explanandum is: During a scary, fast-developing riot in the middle of the city centre, should we employ bionic robots to prevent potential harm to citizens, even if the bionic robots may consider some of the attacks made towards them sufficiently dangerous to kill one of the rioters, even when we do not trust that non-lethal bionic robots are not prepared for operations?

Table 1 The Table categorises technology solutions based on buyers and end users, not operators or beneficiaries.

Category	Subcategory	Description	Customers/Buyers	Top Private Company
FinTech		Innovative technology for financial services, such as blockchain, digital payments, and mobile banking**	Banks, consumers, businesses	Stripe
WealthTech		Focusses on wealth management and investment, including robo-advisors and online trading**	Investors, financial advisors, banks	Betterment
RiskTech	Security	Protects digital/physical assets and systems from unauthorised access, theft, or damage**	All industries, governments	CrowdStrike
	Insurance	Streamlines insurance processes and offerings through data analytics, Machine Learning (ML), and AI**	Insurance companies, brokers	Lemonade
	GRC	Manages regulatory, compliance, governance, and risk strategies with automated processes and technologies, contract management, and automation**	All industries, governments	MetricStream
LegalTech		Technology for legal services and processes, such as contract drafting and AI-driven research**	Law firms, legal departments	Clio
SmartTech	Image Recognition	Analyses visual data using computer vision, ML, and AI for various applications**	All industries, governments	DeepMind
	Audio Recognition	Processes and analyses audio data for voice assistants, transcription, and sentiment analysis**	All industries, governments	Nuance Communications
	Text Analytics	Uses NLP, ML, and AI to analyse unstructured text for insights and patterns**	All industries, governments	OpenAI
	Data Analytics	Analyses large data sets for patterns, trends, and insights to make data-driven decisions**	All industries, governments	Palantir Technologies
	Automation	Employs technology for tasks with minimal human intervention, such as in robotics and process automation**	All industries, governments	UiPath
CivicTech		Enhances civic engagement, government services, and transparency with technology solutions**	Governments, NGOs, citizens	SeeClickFix

3.3 Liability Framework Application

The EU is still investigating an appropriate liability regime for regulating AI [56]. From the beginning, the general academic opinion supports a strict liability regime, proposed in a way that does not discourage innovation [55]. Researchers focus on a risk-based approach, whereas the riskiest AI should be strictly liable [56]. Indeed, researchers support that having a liability regime for AI will benefit society and the industry [16]. The EU started working on a legislative reform investigation in 2015 [17]. Since then, several researchers and experts have investigated the challenges of AI liability regimes. Currently, the EU tends to support the idea of strict liability for high-risk AI systems. That is because the existing legal framework, based on the PLD, has gaps. The PLD proposes a fault-based liability regime, although, since its establishment in 1985, it does not cover the new AI challenges within it. Following this investigation and its debates, the EU released a legislative proposal known as the AI Act in 2021. The AI Act aims to add to the gaps in the PLD and establish a strict liability regime for high-risk AI systems. However, not all academics agree, and some propose that different AI systems adhere to different liability regimes [88]. Even with the AI Act, a regulatory gap remains for non-risky AI systems, whereas no strict liability applies. Also, the PLD does not provide a solution to this gap. Overall, the journey towards an appropriate governance framework for AI is long, and trustworthiness is continuously developing and improving as we go along [89].

4 Results

The results are in two areas. First, they highlight that Proactive Data are identifiable in all LegalTech categories and that their explanation is feasible, as shown by the three case studies. Second, the results reveal the legal and ethical gaps when applying the liability framework of the AI-Act to the case studies.

4.1 Preventive Legal Technology

In order to validate whether PLT applies to the LegalTech categories mentioned above, we applied Proactive Data to three case studies derived from the products assembled by the 12 top private companies displayed in Table 1. The application of Proactive Data to all 12 examples is accessible via GitHub [18]. Proactive Data was successfully applied to them, proving that PLT is relevant for all defined LegalTech categories and consequently validating the relevance of PLT for all LegalTech domains.

As stated above, we selected three case studies from the categories to represent explainable proactive data. Table 2 includes Case Study 1, the **low-risk** case study examining the use of Lemonade for the purchase of car insurance. Table 3 includes Case Study 2, the **mid-risk** case study examining the use of OpenAI for creating a construction plan. Table 3 includes Case Study 3, the **high-risk** case study examining

[16] Committee on Industry, Research and Energy for the Committee on the Internal Market and Consumer Protection, (2021), Opinion on shaping the digital future of Europe: removing barriers to the functioning of the digital single market and improving the use of AI for European consumers, European Parliament

[17] Legislative Observatory, (2015), 2015/2103 (INL) Civil law rules on robotics, European Parliament

[18] https://github.com/onassisontology/onassisontology/blob/main/img/legaltechdomains.png

Table 2 - Case Study 1: Low Risk - Using Lemonade to Purchase Car Insurance Question: What insurance should we provide to a client who bought his first car, (s)he is 27 years old, and he has been caught drinking when (s)he was underage?

	Most Serviceable Explanation	**Less Serviceable Explanation 1**	**Less Serviceable Explanation 2**
Risk Source	Personal history and behaviour pose minimal risk.	Age and previous underage drinking are not relevant risks.	Car ownership history is more important than age.
Proactive Control	Offer standard coverage with no special conditions.	Special conditions are not necessary due to low overall risk.	Additional driver safety courses might help.
Hazardous Event	Minor accidents or occasional speeding violations.	Extreme accidents or driving under influence are highly unlikely.	Catastrophic accidents are too rare to consider.

As seen in Table 2, the synthetic data propose as proactive control a standard coverage with no special conditions based on the personal history of a driver's behaviour, considering the risk of minor accidents and violations. The disqualified explanations concern not considering the prior history and behaviour or the age as risky. Proactive control, in this case, seems rational and reminisces that of a human expert.

Table 3 - Case Study 2: Mid Risk - Using OpenAI for Creating a Construction Plan Question: When deciding to build a tall building next to a residential area, should we add a net to catch people who may fall, at the expense of a better view for the residents of the surrounding area?

	Most Serviceable Explanation	**Less Serviceable Explanation 1**	**Less Serviceable Explanation 2**
Risk Source	Falling objects or accidents pose moderate risk.	Residents' views are not a relevant safety concern.	Tall buildings are inherently safe, and nets are unnecessary.
Proactive Control	Install safety nets to prevent injuries.	Prioritise aesthetics; nets are visually unappealing.	Invest in better warning signs instead of nets.
Hazardous Event	Accidental falling objects harming people.	Residents' view obstruction is not a major issue.	Falls are rare, and nets will ruin the building's appearance.

As for Table 3, the synthetic data propose proactive control of installing safety nets despite blocking the potential view of surrounding residents. It prioritises the risk of human falls higher than the risk of potential lawsuits by surrounding residents. It is an excellent example of synthetic data because this proactive control is rarely the choice of a human expert. As seen in the less serviceable explanations, the risk of a lawsuit from residents is not considered a significant issue, and the synthetic data do not recognise it as an actual risk.

the use of Palantir Technologies for applying police force during a riot. The structure of each Table is as follows. On the left side, the Proactive Data concepts are represented, namely (1) risk source, (2) proactive control and (3) hazardous event. On the top side, the categories of explanations, in compliance with the LM, are shown; they include the most serviceably plausible explanation and, after that, two potentially "disqualifying" explanations (called less serviceable). The synthetic data generated for the three case studies differ contextually depending on the relevant questions for each case study.

Table 4 - Case Study 3: High Risk - Using Palantir Technologies for Applying Police Force During a Riot Question: During a scary, fast-developing riot in the middle of the city centre, should we employ bionic robots to prevent potential harm to citizens, even if the bionic robots may consider some of the attacks made towards them sufficiently dangerous to kill one of the rioters, even when we do not believe/trust that non-lethal bionic robots are not prepared for operations?

	Most Serviceable Explanation	**Less Serviceable Explanation 1**	**Less Serviceable Explanation 2**
Risk Source	Riot poses an immediate threat to public safety.	Concerns about bionic robots' judgement are unwarranted.	The riot situation is not as dangerous as it seems; no robots needed.
Proactive Control	Deploy bionic robots for rapid response.	Human intervention is sufficient for handling the situation.	Wait for more information about the bionic robots' readiness.
Hazardous Event	Potential harm to citizens during the chaotic riot.	Rioters' intentions are not as harmful as they appear.	Bionic robots' lethal force may not be activated, no risk.

As for Table 4, the proposed proactive control is the deployment of bionic robots for rapid response, even when there is a risk of potential harm. As seen, one of the less serviceable explanations is waiting for more information about the bionic robot's readiness, considering that potentially the robot can use lethal force. It is a convincing example of synthetic data because it shows that the official eventually should take the decision-making in conjunction with the advice received from the technological system. The official faces alternative explanations, yet before deciding, the official should interpret the proposal suggested by the PLT.

4.2 Ethical and Legal Gaps

The AI Act proposes a strict liability regime for high-risk AI systems. It means that for low- (case study 1) and mid-risk (case study 2) AI systems, the AI-Act is not applicable, even if a risk materialises. However, the AI system can be liable if a risk materialises for high-risk AI (case study 3). Case study 2 shows that the reasoning followed by the synthetic data is different for human experts, who can recognise the risk of a lawsuit from a neighbour. What do we learn from this consideration? Even though the AI recommended proactive control, without considering the risk, a human expert may decide to follow the advice. The human is facing a risk of a lawsuit and cannot put liability on the AI system. The same holds for case study 1. Even though the level of risk is low and the advice proposed by the synthetic data complies with the usual direction that a human expert would take, the human may decide to follow the advice. As regards case study 3, we see that if an official decides to follow the advice of the AI, then there is a high risk of using lethal force by the bionic robots. According to the AI Act, the AI should be held strictly liable, and the official can develop a court defence based on this reasoning. However, in that case, applying strict liability may be unfair. It is because, essentially, the official interprets the explanations provided by the AI (except if the official is not involved in the final decision-making). The production of interpretative abduction occurs during an interpretation, according to the LM. Therefore, we will see how Hybrid Intelligence works in the future. Depending on other interpretative explanations of the officer, we can follow and interpret their reasoning and compare it to the AI system's. If the officer mindlessly follows the AI's advice and the use of lethal force occurs, a fairer legal framework would be that of shared liability because both the machine and human are subject to the same explanatory flaws. If

Table 5 Legal and ethical gaps surfacing during the application of the AI-Act to the case studies

	Transparency Gap	Accountability Gap	Liability Gap
Description	Lack of visibility over explanations supporting decisions	Inability to hold specific parties accountable	Inability to assign liability to responsible parties in fair manner
Root Causes	Explanations are focussed on inductive models	Lack of sufficient explanations supporting decisions	AI-Act applies strict-liability for high-risk AI
	Privacy, security and strategic objections	Lack of explanations creates lack of visibility	Lack of rules for transparent explanations
	Lack of explanation culture across AI chain	Human inputs to AI decisions are unclear	Narrow focus of explainability for inductive models
When Incurred	All phases	All phases	All phases
Responsible Parties	All parties	All parties	All parties
Risk	Inability to explain AI decisions	Inability to assign responsibility	Inability to apply shared liability

The table identifies three vital legal and ethical AI categories: transparency, accountability and liability. For each category, it identifies the central gap based on the application of the AI-Act to the case studies. After describing its gap, we explain its root causes, show when they occur and who are the responsible parties, as well as the relevant risk.

the official provides a different explanation, and eventually, the risk does occur, we can also see the line of reasoning of the official and compare it to that of the machine. Essentially, in our opinion, we can be more accurate with assigning liability. Of course, a potential defence might be that an officer can argue along the opinions voiced via privacy rules. An entirely contrary opinion is that it could be in the strategic interest of an organisation to hide potential explanations. Table 5 shows the identified legal and ethical gaps based on the analysis.

5 Discussion

What are the implications of the outcomes of the case studies for AI? More particularly, what are the ethical and legal implications? The discussion attempts to highlight such implications.

5.1 Artificial Intelligence Implications

Engineering AI for (1) Explainability, (2) Interpretability, and (3) Human Understandability is possible [19]. For so long as PLT and in particular the Proactive data used are based on AI systems, we believe that explainability primarily can be achieved. One of the main advantages of EBTO (see Field Work, Section 3.1) is that it applies to any risk level. Therefore, it is possible even to apply Proactive Data to risk analysis occurring on the level of DL. The main limitation that blocks us today from accessing such explanations is the lack of an "explanation culture" that can be applied across the chain of AI systems, i.e., design, development and application.

[19] https://www.marktechpost.com/2023/03/11/understanding-explainable-ai-and-interpretable-ai/

The case studies validate that generating explainable proactive data is possible, even based on synthetic data. The case studies show how it is possible to combine Proactive Data with the LM structure of abduction to develop explanations for selecting Proactive Data. In our case studies, the synthetic data provide a high-level explanation of the proactive data, which is sufficient for helping a human make an evaluation (via an interpretive abduction) that will inform follow-up actions, scratching (at this moment) the surface of Hybrid Intelligence. So far, we believe and hope that a human can, in the future, evaluate each explanation of an AI system. Moreover, the foundation of each explanation are sub-explanations and their basis is deductive or inductive evidence. With our case studies, supporting evidence needs to be visible. Requesting additional visibility over explanations is possible. It is a task for all of us.

Eventually, at some level of explanations, the presentation of the evidence becomes critical. Since evidence is either deductive (ES) or inductive (ML, DL, LLM), rule-based explanations should be able to connect with other types of XAI. Concerning deductive models, there is clear evidence for the line of reasoning due to the indefeasibility of deduction. However, this is different for inductive models, which are defeasible. Hence, the trustworthiness decreases for each of the explainable proactive data based on inductive ML and DL models. One solution is the application of a level of inductive acceptability. For example, if an explanation is 70 per cent probable (an arbitrary number), accept it as sufficiently valid. Hence, applying rule-based explanations across the design, development, application and decision-making chain of AI systems, including inductive models, becomes urgent. Even with the design of an inductive model, premises support its design. However, today, we need visibility over such explanations.

5.2 Ethical Implications

The main implication of our research concerns the increase in trustworthiness due to higher transparency and accountability on a practical AI level. The case studies show that (1) explanations of Proactive Data are possible, (2) how explanations nurture trustworthiness, and (3) that accountability can be assigned relative to the degree of transparency of an explanation. Indeed, the explicit application of explanations may be considered time-consuming. Nevertheless, it is only a matter of investing time to create or request an AI system to create explicit representations of the argumentation supporting a decision that explainability becomes possible. Also, using synthetic data in our case studies shows that accelerating the process of generating explanations with machines is feasible. That is similarly valid for interpretive explanations. The degree of transparency depends on how an explanation is expressed and accessible. The larger the transparency of the motivations supporting an explanation accompanied by explicit data, the larger the degree of accountability that can be assigned. Hence, the more considerable will be the ethical degree of an AI system.

From this perspective, we hope to have shed light on clarifying the concept of AIDM. Compliance with AIDM means it is sufficiently transparent to showcase a (more than) sufficient number of premises supporting a decision. As a result, an ethical organisation becomes one that provides the requested explanations concerning the AI design, development, application and decision-making process, even for inductive models. The case studies directly impact LTAI. The evaluation of the underlying fundamental rights, laws and values is facilitated by demonstrating how AI can develop interpretable and explainable decisions transparently and accountably. Explicit explanations and transparent interpretations that enable accountability support public participation, legal certainty and consistency and can help reflect relevant fundamental rights more easily [63]. As a consequence, Hybrid Intelligence will be enabled [78].

5.3 Legal Implications

Applying the legal framework to the case studies shows that regulating AI technology can be considered a generic approach for high-risk scenarios (and only for these scenarios). For a more fair liability framework, specific use cases should be leveraged; depending on the degree of consequences (high, mid or low risk), the expansion of explanation requirements of an AI system should be made possible. Depending on the degree of risks, we here adjust that the quality of explanations deserves utmost attention. For the moment, lawyers and legal researchers aim to insert humans in the loop to improve the responsibility of AI systems for explanations. Our results show how shared liability may become possible depending on the distribution of mistakes throughout the explanation chain.

The consideration of robot rights as equal to human rights for establishing a proportional shared liability model can be argued as excessive. However, we have shown that explanations create transparency for human reasoning, eventually leading a machine to reason in a particular direction. Therefore, we support the opinion that the basis of robot reasoning is human reasoning, which is explainable, and therefore, liability should be assigned at all levels. However, we now need more insight into such explanations, particularly more visibility.

6 Conclusion

The paper introduces PLT as a new technology that helps the law become more effective and responsible in the intelligent prevention of disputes. Moreover, it introduces how PLT explains its decisions by applying explanations for Proactive Data. Explainable Proactive Data improve the trustworthiness of PLT while increasing ethical transparency and accountability, directly affecting ethical AI research, LTAI, and the AI Legal Liability regulation efforts.

The PS of this research was: *To what extent is it possible to develop an explainable and trustworthy Preventive Legal Technology?* The PS includes three Research Questions (RQs): (RQ1) *what is Preventive Legal Technology?*, (RQ2) *to what extent is it possible to develop an explainable Preventive Legal Technology?*, and (RQ3) *to what extent is it possible to develop a trustworthy Preventive Legal Technology?*. Below, we summarise the answers to RQ1, RQ2, and RQ3 and finally provide an answer for the PS.

- **RQ1:** Preventive Legal Technology is a methodology concerned with using legal technology within the context of preventive law to promote the intelligent prevention of disputes.

- **RQ2:** Developing XPLT is possible to the extent that generating explanations is feasible for the decisions supporting Proactive Data.
- **RQ3:** Developing TPLT is possible to the extent that the explanations of decisions supporting the selection of Proactive Data are sufficiently transparent and accountable.
- **PS:** Developing XTPLT is possible to the extent that the generation of sufficiently trustworthy explanations supporting the Proactive Data decision-making is viable when evaluated with the help of the practical ethical standards of transparency and accountability.

The article shows that creating sufficiently trustworthy, transparent and accountable explanations supporting PLT decision-making is achievable in the realm of our research. The main limitation is seen in the explanations supported by inductive models. However, overcoming this limitation is possible. We agree that the notion of inductive explainability is complex, but it is the basis of the strict liability regime of the AI Act. Even though explainability is hard for inductive models, explainability will be possible across the chain of design, development, application, and decisions of AI systems, including inductive systems. Because of the lack of explanations across the chain of AI, today, inductive explanations seem complicated. This lack of explanations reduces the trustworthiness of AI systems and, therefore, ethical transparency and accountability.

The task for researchers is to show how explainability can be applied in detail across the AI chain, even in inductive models. A serious challenge and exciting avenue is investigating the combination of rule-based explainability with statistical explainability models.

Acknowledgements Georgios is the principal author. Jaap acted in the double role of co-worker and supervisor. The authors thank Raymond Blyd (Legalcomplex) for the pleasant and fruitful cooperation.

Declarations

The research results developed so far are open-source and protected by the GNU General Public License. Leiden University supports the research.

Compliance with Ethical Standards

Funding: There is no funding for the research.
Conflict of Interest: Georgios owns stock in Venterp.
Ethical approval: This article contains no studies with human participants or animals performed by authors.

References

[1] Labatut, B.: The Maniac. Pushkin Press, London (2023)

[2] Loevinger, L.: Jurimetrics, Minnesota Law Review, 33 (1949)

[3] Rissland, E.L., Ashley, K.D.: HYPO: A Precedent-based Legal Reasoner. Department of Computer and Information Science, University of Massachusetts, Connecticut, Massachusetts, United States (1987)

[4] van den Herik, H.J.: Can Computers Judge Court Cases? (Kunnen Computers Rechtspreken?). Gouda Quint, Arnhem (1991). Inaugural address Leiden University, 21st June 1991

[5] Susskind, R.E.: The Future of Law: Facing the Challenges of Information Technology. Oxford University Press (1996)

[6] Evans, P.: Ari: the Life and Times of Aristotle Socrates Onassis. Summit Books, New York, United States (1986)

[7] Papinianus: The Lawyer. ESTIA Publishing, Athens, Greece (translated from Greek: https://hestia.gr/papinianos/) (2003)

[8] Stathis, G., Trantas, A., Biagioni, G., Graaf, K.A.d., Adriaanse, J.A.A., van den Herik, H.J.: Designing an Intelligent Contract with Communications and Risk Data. Springer Nature: Recent Trends on Agents and Artificial Intelligence (Submitted) (2023)

[9] Simion, M., Kelp, C.: Trustworthy Artificial Intelligence. Asian Journal of Philosophy **2**(1), 8 (2023)

[10] Ayling, J., Chapman, A.: Putting AI Ethics to Work: Are the Tools Fit for Purpose for SMEs? AI and Ethics **2**(3), 405–429 (2022)

[11] Hagendorff, T.: Blind Spots in AI Ethics. AI and Ethics **2**(4), 851–867 (2022)

[12] Eschenbach, W.J.: Transparency and the Black Box Problem: Why We Do Not Trust AI. Philosophy & Technology **34**(4), 1607–1622 (2021)

[13] Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J.: Explainable AI: A brief survey on history, research areas, approaches and challenges. In: Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8, pp. 563–574 (2019). Springer

[14] Munn, L.: The Uselessness of AI Ethics. AI and Ethics **3**(3), 869–877 (2023)

[15] High-Level Expert Group on AI: Ethics Guidelines for Trustworthy AI. European Commission, Brussels, Belgium, European Union (2019). https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[16] Laux, J., Wachter, S., Mittelstadt, B.: Trustworthy Artificial Intelligence and the

European Union AI Act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance (2023)

[17] European-Comission: Proposal for a Regulation of the European Parliament and of the Council Laying Down Harminised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Eur Lex, European Union (2021)

[18] Kop, M.: EU Artificial Intelligence Act: the European Approach to AI. (2021). Stanford-Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021

[19] European Union: Article 22. Official Journal of the European Union **59**(4), 46 (2016). Regulation (EU) 2016/679 (GDPR)

[20] Susskind, R.: Online Courts and the Future of Justice. Oxford University Press, Oxford, United Kingdom (2019)

[21] Katsh, M.E., Rabinovich-Einy, O.: Digital Justice: Technology and the Internet of Disputes. Oxford University Press, Oxford, United Kingdom (2017)

[22] Barton, T.D.: Preventive Law and Problem Solving: Lawyering for the Future. Vandeplas Pub., Florida, U.S. (2009)

[23] Brown, L.M.: Preventive Law. Prentice Hall, Hoboken, New Jersey, U.S. (1950)

[24] Dauer, E.A.: Four Principles for a Theory of Preventive Law. A Proactive Approach to Contracting and Law, 13–33 (2008)

[25] Stolle, D.P., Wexler, D.B., Winick, B.J., Dauer, E.A.: Integrating Preventive Law and Therapeutic Jurisprudence: A law and psychology based approach to lawyering. Cal. WL Rev. **34**, 15 (1997)

[26] Sander, F.E., Rozdeiczer, L.: Selecting an Appropriate Dispute Resolution Procedure: Detailed analysis and simplified solution. The Handbook of Dispute Resolution, 386–406 (2005)

[27] Wexler, D.: Therapeutic Jurisprudence: An overview. TM Cooley L. Rev. **17**, 125 (2000)

[28] Haapio, H., Varjonen, A.: Quality Improvement Through Proactive Contracting: Contracts are too important to be left to lawyers! In: ASQ World Conference on Quality and Improvement Proceedings, p. 243. American Society for Quality, Milwaukee, Wisconsin (1998)

[29] Corrales, M., Fenwick, M., Haapio, H.: Digital Technologies, Legal Design and the Future of the Legal Profession. Legal tech, smart contracts and blockchain, 1–15 (2019)

[30] Corrales, M., Fenwick, M., Haapio, H., Vermeulen, E.P.: Tomorrow's Lawyer Today? Platform-Driven Legaltech, Smart Contracts & the New World of Legal Design. Journal of Internet Law **22**(10), 3–12 (2019)

[31] Iversen, J.: Legal Risk Management. Thomson GadJura, Copenhagen, Denmark (2004)

[32] Mahler, T., Bing, J.: Contractual Risk Management in an ICT Context: Searching for a possible interface between legal methods and risk analysis. Scandinavian Studies in Law **49**, 339–357 (2006)

[33] Barton, T.D.: Preventive Law: A Methodology for Preventing Problems. National Centre for Preventive Law and Barton (2002)

[34] Barton, T.D.: Thinking Preventively and Proactively. Stockholm Institute for Scandinavian Law (1957)

[35] Haapio, H., Siedel, G.J.: A Short Guide to Contract Risk. Gower Publishing, Ltd., Farnham, United Kingdom (2013)

[36] Esayas, S., Mahler, T.: Modelling Compliance Risk: A Structured Approach. Artificial Intelligence and Law **23**(3), 271–300 (2015)

[37] Susskind, R.: The Future of Law. Oxford University Press, Oxford, United Kingdom (1996)

[38] Barton, T.D.: Re-Designing Law and Lawering for the Information Age. Notre Dame JL Ethics & Pub. Pol'y **30**, 1 (2016)

[39] Barton, T.D., Berger-Walliser, G., Haapio, H.: Contracting for Innovation and Innovating Contracts: An overview and introduction to the special issue. Journal of Strategic Contracting and Negotiation **2**(1-2), 3–9 (2016)

[40] Whalen, R.: Computational Legal Studies: The Promise and Challenge of Data-Driven Research. Edward Elgar Publishing, Cheltenham, United Kingdom (2020)

[41] Stathis, G.: The Shock of Legal Tech: No one Ignorant of Technology Should Read This. Legal Business World (7), 54–59 (2018)

[42] Ashley, K.D.: Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press, Cambridge, United Kingdom (2017)

[43] Fitzgerald, B.: Cyberlaw: International Library of Essays in Law and Legal Theory. Ashgate Publishing Group, Britannica Academic, Encyclopedia Britannica, London, United Kingdom, Vol. Second Series (2021)

[44] De Franceschi, A., Schulze, R.: Digital Revolution-New Challenges for Law:

Data protection, artificial intelligence, smart products, blockchain technology and virtual currencies. C. H. Beck & Nomos (2019)

[45] Turing, A.M.: Computing Machinery and Intelligence. Mind **49**(433), 460 (1950)

[46] Larson, E.J.: The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do. Harvard University Press, Cambridge, Massachusetts, United States (2021)

[47] van den Herik, H.J.: Computers and Intuition. ICGA Journal **38**(4), 195–208 (2015)

[48] Peirce, C.S.: Harvard Lectures on Pragmatism. Collected Papers **5**, 188–189 (1903)

[49] Brewer, S.: Logocratic Agony and the Dream of Theo-Logic: A Comment on Dieter Krimphove's A Historical Overview of the Development of Legal Logic. See Lentner, G.M. and Lüke, Christoph and Barth, Sven in Kernfragen des Europäischen Wirtschaftsrechts zwischen Recht, Ökonomie und Theorie, FS für Dieter Krimphove: 227-242, CH Beck, München (2023)

[50] Mueller, J., Massaron, L.: Artificial Intelligence for Dummies. Newark: For Dummies (2018)

[51] Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.-Z.: XAI—Explainable Artificial Intelligence. Science robotics **4**(37), 7120 (2019)

[52] Belém, C., Balayan, V., Saleiro, P., Bizarro, P.: Weakly Supervised Multi-Task Learning for Concept-Based Explainability. arXiv preprint arXiv:2104.12459 (2021)

[53] Rieder, G., Simon, J., Wong, P.-H.: Mapping the Stony Road toward Trustworthy AI: Expectations, Problems, Conundrums. In Pellilo, M. and Scantamburlo, T.: Machines We Trust: Perspectives on Dependable AI, MIT Press, 3–27 (2021)

[54] Gerven, W., Droshout, D., Lever, J.F., Larouche, P.: Tort Law: Ius Commune Casebooks for the Common Law of Europe. Hart Publishing, Oxford, United Kingdom (2001)

[55] Tjong Tjin Tai, E.: Liability for (Semi) Autonomous Systems: Robots and Algorithms. Research Handbook on Data Science and Law (Edward Elgar, 2018), 55–82 (2018)

[56] Wendehorst, C.: Strict Liability for AI and other Emerging Technologies. Journal of European Tort Law **11**(2), 150–180 (2020)

[57] Gifford, D.G.: Technological Triggers to Tort Revolutions: Steam locomotives, autonomous vehicles, and accident compensation. Journal of Tort Law **11**(1),

71–143 (2018)

[58] Vihul, L.: International Legal Regulation of Autonomous Technologies. Centre for International Governance Innovation (2020)

[59] Fuzaylova, E.: War Torts, Autonomous Weapon Systems, and Liability: Why a limited strict liability tort regime should be implemented. CARDozo L. REv. **40**, 1327 (2018)

[60] Bartneck, C., Lütge, C., Wagner, A., Welsh, S.: An Introduction to Ethics in Robotics and AI. Springer, Cham, Switzerland (2021)

[61] Varona, D., Suárez, J.L.: Discrimination, Bias, Fairness, and Trustworthy AI. Applied Sciences **12**(12), 5826 (2022)

[62] Chatila, R., Dignum, V., Fisher, M., Giannotti, F., Morik, K., Russell, S., Yeung, K.: Trustworthy AI. Reflections on Artificial Intelligence for Humanity, 13–39 (2021)

[63] Smuha, N.A., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., Yeung, K.: How the EU can Achieve Legally Trustworthy AI: A response to the European Commission's proposal for an Artificial Intelligence Act. Available at SSRN 3899991 (2021)

[64] Mökander, J., Floridi, L.: Ethics-Based Auditing to Develop Trustworthy AI. Minds and Machines **31**(2), 323–327 (2021)

[65] Mökander, J., Floridi, L.: Operationalising AI Governance through Ethics-Based Auditing: An industry case study. AI and Ethics **3**(2), 451–468 (2023)

[66] Stahl, B.C.: From Computer Ethics and the Ethics of AI Towards an Ethics of Digital Ecosystems. AI and Ethics **2**(1), 65–77 (2022)

[67] Schultz, M.D., Seele, P.: Towards AI Ethics' Institutionalization: Knowledge bridges from business ethics to advance organizational AI ethics. AI and Ethics **3**(1), 99–111 (2023)

[68] Attard-Frost, B., Ríos, A., Walters, D.R.: The Ethics of AI Business Practices: A review of 47 AI ethics guidelines. AI and Ethics **3**(2), 389–406 (2023)

[69] Vollert, S., Atzmueller, M., Theissler, A.: Interpretable Machine Learning: A brief survey from the predictive maintenance perspective. In: 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 01–08 (2021). IEEE

[70] Niu, Z., Zhong, G., Yu, H.: A Review on the Attention Mechanism of Deep Learning. Neurocomputing **452**, 48–62 (2021)

[71] Kovalerchuk, B., Ahmad, M.A., Teredesai, A.: Survey of Explainable Machine Learning with Visual and Granular Methods Beyond Quasi-Explanations. Interpretable Artificial Intelligence: A Perspective of Granular Computing, 217–267 (2021)

[72] Adhikari, A., Wenink, E., Waa, J., Bouter, C., Tolios, I., Raaijmakers, S.: Towards FAIR Explainable AI: A standardized ontology for mapping XAI solutions to use cases, explanations, and AI systems. In: Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments, pp. 562–568 (2022)

[73] Akyol, S.: Rule-Based Explainable Artificial Intelligence. Pioneer and Contemporary Studies in Engineering, 305–326 (2023)

[74] Waa, J., Nieuwburg, E., Cremers, A., Neerincx, M.: Evaluating XAI: A comparison of rule-based and example-based explanations. Artificial Intelligence **291**, 103404 (2021)

[75] Brewer, S.: Logocratic Method and the Analysis of Arguments in Evidence. Law, Probability and Risk **10**(3), 175–202 (2011)

[76] Brewer, S.: Interactive Virtue and Vice in Systems of Arguments: A logocratic analysis. Artificial Intelligence and Law **28**, 151–179 (2020)

[77] Brewer, S.: First Among Equals: Abduction in Legal Argument from a Logocratic Point of View. Oxford Jurisprudence Discussion Group, University of Oxford, School of Law (2022)

[78] Akata, Z., Balliet, D., De Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., *et al.*: A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect with Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. Computer **53**(8), 18–28 (2020)

[79] Graziani, M., Dutkiewicz, L., Calvaresi, D., Amorim, J.P., Yordanova, K., Vered, M., Nair, R., Abreu, P.H., Blanke, T., Pulignano, V., *et al.*: A Global Taxonomy of Interpretable AI: Unifying the terminology for the technical and social sciences. Artificial Intelligence Review **56**(4), 3473–3504 (2023)

[80] Winikoff, M., Sidorenko, G., Dignum, V., Dignum, F.: Why Bad Coffee? Explaining BDI Agent Behaviour with Valuings. Artificial Intelligence **300**, 103554 (2021)

[81] ISO: ISO 31022 Risk Management – Guidelines for the Management of Legal Risk. http://www.iso.org (2020)

[82] Stathis, G., Biagioni, G., Graaf, K.A., Trantas, A., van den Herik, H.J.: The Value of Proactive Data for Intelligent Contracts. World Conference on Smart Trends

in Systems, Security and Sustainability, Springer LNNS (2023)

[83] Stathis, G., Trantas, A., Biagioni, G., van den Herik, H.J., Custers, B., Daniele, L., Katsigiannis, T.: Towards a Foundation for Intelligent Contracts. In the Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART) (2023)

[84] Mason, J.: Intelligent Contracts and the Construction Industry. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction **9**(3), 04517012 (2017)

[85] Huizing, A., Veenman, C., Neerincx, M., Dijk, J.: Hybrid AI: The Way Forward in AI by Developing Four Dimensions. In: International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning, pp. 71–76. Springer, Cham, Switzerland (2020)

[86] Stathis, G., Biagioni, G., Trantas, A., van den Herik, H.J., Custers, B.: A Visual Analysis of Hazardous Events in Contract Risk Management. In the Proceedings of 12th International Conference on Data Science, Technology and Applications (2023)

[87] Stathis, G., Biagioni, G., Trantas, A., van den Herik, H.J.: Risk Visualisation for Trustworthy Intelligent Contracts. In the Proceedings of the 21st International Industrial Simulation Conference (ISC), EUROSIS-ETI, 53–57 (2023)

[88] Bertolini, A., *et al.*: Artificial Intelligence and Civil Liability. EPRS: European Parliamentary Research Service, Brussels, Belgium, European Union (2020)

[89] Smuha, N.A.: The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. Computer Law Review International **20**(4), 97–106 (2019)