Ethical and preventive legal technology

Stathis, G.; Herik, H.J. van den

AI and Ethics (2025) 5:1069–1086
https://doi.org/10.1007/s43681-023-00413-2

ORIGINAL RESEARCH

Ethical and preventive legal technology

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

Received: 30 July 2023 / Accepted: 15 December 2023 / Published online: 18 March 2024

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 are expected to improve the level of *transparency* and *accountability*. Trustworthiness is an urgent topic in the discussion on doing AI research ethically and accounting for the regulations. For this purpose, we examine the limitations of rule-based explainability for PLT. Hence, our Problem Statement reads: *to what extent is it possible to develop an explainable and trustworthy Preventive Legal Technology?* After an insightful literature review, we focus on case studies with applications. The results describe (1) the effectivity of PLT and (2) its responsibility. The discussion is challenging and multivariate, investigating deeply the relevance of PLT for LegalTech applications in light of the development of the AI Act (currently still in its final phase of process) and the work of the High-Level Expert Group (HLEG) on AI. On the ethical side, explaining AI decisions for small PLT domains is clearly possible, 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 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

✉ Georgios Stathis
g.stathis@law.leidenuniv.nl

Jaap van den Herik
h.j.van.den.herik@law.leidenuniv.nl

[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

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, therefore, 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 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] [8]. 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 [7]. 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) and its applications. 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 so far are *not transparent*, resulting in "black box" decisions [12]. Below, we briefly introduce XAI and TAI. A literature review expands on the concepts in Sects. 2.4.1 and 2.4.2.

XAI is the field of study investigating the *explanation* of AI system decisions [13]. XAI is assumed to lead to TAI,

[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).

[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.

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, which is a central topic in the developing research on AI regulation according to the High-Level Expert Group on AI (HLEG) [15]. Researchers have noticed a 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 properly formulating 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 a higher ethical transparency and accountability [18].

The idea is that more profound insight into XAI and TAI will enable us to examine PLT from two different 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 to what extent PLT is a distinct field of technology. Due to the reliance of PLT on Proactive Data, PLT can be considered a special type of Artificial Intelligence (AI). It is a type of Predictive AI (probabilistic future event forecasting based on historical data) rather than Generative AI (new data creation in text or image).[6] 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 such decisions is on factual data, as well as on digital profiles or inferred data.[7] If AIDM includes explanations, then AI systems will be more trustworthy due to higher transparency and accountability. AIDM is particularly relevant for the EU's AI Act, which provides the right to consumers to launch complaints and receive meaningful explanations[8] (see also [20]). As a result, 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, we focus on *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 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 and Predictive AI.

To answer the PS, we structured the paper as follows. In Sect. 2, we describe the literature. Section 3 presents our three methodologies: fieldwork, case studies and applications. Then, Sect. 4 describes the investigations and states the results. Section 5 discusses those results and focusses on trustworthiness and ethical parameters. Finally, Sect. 6 answers the PS and the three RQs and provides our conclusion; then Sect. 7 provides further research suggestions.

2 Literature review

When a legal problem occurs, both (or all) parties subject to a legal agreement may experience costs, e.g., owing to 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 and how or when the parties can solve it. Let us start to remark that 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,[9] even in the world's most

[6] https://www.blueprism.com/resources/blog/generative-ai-vs-predictive-ai/.

[7] 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.

[8] https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

[9] 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% in the USA (429 Billion Dollars [US Chamber Institute for Legal Reform, (2018), Costs and Compensation of the US Tort System,

advanced jurisdictions [21]. For this reason, commentators opine that we need *new ways* to resolve and avoid disputes [22]. A primary reason why legal costs are so high is that the structure of legal systems invites dispute *resolution* and not dispute *prevention* [23]. All in all, legal costs give rise to a need for preventive law.

2.1 List of definitions

As assistance to the reader we provide a list of nine definitions, so that the content is clear and the precise meaning of the concepts is easy to consult. Definitions are in bold, concepts in italics and descriptions in normal letter type. For clarity in reading we will repeat the definitions when necesary.

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 (derived from [24]).

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 (derived from [5, 24].

Definition 3 *Legal Risk Management* is the management of risk (effect of uncertainty on objectives) related to legal, regulatory and contractual matters, and ranging from no rights to non-contractual rights to obligations [25].

Definition 4 *Artificial Intelligence* refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data, and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal [15].

Definition 5 An *Intelligent Contract* (or *iContract*) is a contract that is fully executable without human intervention [26, 36].

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 [28].[10]

Definition 8 *Interpretable Machine Learning* (IML) is a system that is possible to learn its working principles and outcomes in human-understandable language without affecting the validity of the system [29, 30].

Definition 9 *Proactive Data* are (1) *Proactive Controls*, which play an important role as soon as we have arrived at the identification of (2) a *Hazardous Event* and have produced an analysis of (3) a *Cause* within the context of the terminology of the Bow-Tie Method and aim to prevent a contract from incurring legal risks [31].

2.2 Preventive law

Louis M. Brown was the first to mention the concept in academic circles via his book *Preventive Law* in 1950 [24]. Up to 1997, Preventive Law (see Definition 1) was understood as a branch of law that endeavours to minimise the risk of litigation or to secure more certainty regarding legal rights and duties [32, 33]. During that period, researchers took into consideration the *proactive* parameters and the *reinforcing* parameters (i.e., the strengthening of legal rights and duties) of preventive law [24]. 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 [27]. It is important to note that a dispute resolution procedure begins once there is a dispute (see Definition 1).

After its start in 1950, Preventive Law (see Sect.1.1) gave rise to two primary schools of thought. They are *Therapeutic Jurisprudence* [34] (started in 1987) and *Proactive Law* (started in 1998) [35, 36], 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

Footnote 9 (continued)

instituteforlegalreform.com, p. 1] in 2016) and 0.63% (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% 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.

[10] https://www.ibm.com/topics/explainable-ai.

Barton started to converge, leading to the creation of Preventive/Proactive Law (PPL). Recently, PPL has been concerned with the visualisation of legal information and the effects of technology on PPL [37, 38]. Meanwhile, we saw around 2000 that risk management was growing as a field of study and practice; consequently, Legal Risk Management (LRM) emerged [39] (see Definition 3). By 2010, the academics Tobias Mahler and Jon Bing established the connection between LRM and Proactive Law [40].

Even if a good, willing person attempts to apply preventive law to *prevent legal problems*, there is always a high likelihood for legal problems to *further developing* instead of diminishing. Hence, Brown introduced the first elementary framework for preventing legal problems [24]. 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 [32]. Dauer's systematic analysis resulted in a matrix, which Barton called Dauer's matrix [23]. Barton then refined it considerably [41, 42].

Currently, most literature focusses on the *mindset* and *application* of Preventive Law. The authors do so in various domains, one of which is iContracts [7] (see Definition 5). Indeed, there are exceptions that only partially help people prevent legal problems. Some of them come from the fields of *Proactive Law* [43] and *Legal Risk Management* [44] (see Definition 3). Obviously, there is a need to clarify the reasons behind the lack of substantive methods or approaches to prevent legal problems.

2.3 Preventive law and legal technology

Richard Susskind was the first to emphasise 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 [5]. He believed that our legal system is subject to the paradox of reactive legal services, where technology will support proactive practices (see Definition 9). In 2016, Barton reinforced that notion by stating that the emerging technological culture is mainly compatible with the assumptions underlying PPL [45]. In the same year, Barton and Haapio proposed a re-design of the legal system, which reconsiders the relationship between law and society, in order to guide a law reform in a technology-based society [46]. They were also questioning and investigating the effects of new technologies for PPL and Legal Design [37, 38].

The relationship between *technology* and the *law* is multifaceted and dynamic [47]. 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 called *Computational Legal Studies* (CLS) [47]. The second is when the changing of market conditions leads to the further development of technology that improves the *delivery of legal services*; this approach is called *LegalTech* [48]. Sometimes, the two approaches are also combined [49].

Whether new technologies do or do not apply 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 considered 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 may affect various legal areas, e.g., freedom of expression, privacy, and cyber security [47]. However, in itself, CyberLaw solely concerns information technology or technology related to networks and the internet [50]. In its foundational sense, technology refers to techniques that improve the execution of 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*. Nevertheless, concerning technology regulation (including CyberLaw), we may note that a higher-ranked purpose is *responsibility*. In order to describe the perspective of both purposes in this article, we use the term *legal technology*. Today, legal technology's purpose (and challenge) is to balance effectivity and responsibility. We see that if the law is too responsible, technology is not sufficiently effective, while if technology is too compelling, the law is not sufficiently responsible. Balancing between the two ends is already hard, and it becomes even more complicated as new technologies introduce new and increasingly complex challenges in the law [51]. The literature shows that preventive law and legal technology are on their way from being connected to substantially entangled (see Definition 2).

2.4 Ethics and artificial intelligence

Many ideas about modelling intelligent behaviour started in ancestry and were further developed throughout history.[11] Nevertheless, most researchers attribute the starting point of AI (see Definition 4) to Alan Turing in 1950 [52]. Since then, two main AI movements emerged: the scientific one and the futuristic one [53]. The scientific AI movement supports the idea that formal reasoning is the basis of AI and

[11] 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.

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 [53]. However, many researchers are investigating how to model intuition [54]. 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. Indeed, at this moment, the latter perspective may undervalue linguistic complexity, which is the basis of human intuition[12] [55, 56]. In logic, ingenuity is modelled by deduction or induction; and intuition via abduction [57, 58]. Admittedly, humans still do not know how to model abduction computationally [53]. The developments of AI follow, to a large extent, the developments in logic with modelling intelligence.

2.4.1 Explainable artificial intelligence

The modelling of *deduction* occurs via Expert Systems (ES) and *induction* via ML (and DL or LLMs), but AI so far has not modelled *abduction* [53]. The fundamental elements for ES and ML are "normal" data [59]. With this knowledge, we are ready for the next step: Expainable AI (XAI) (see Definition 7). XAI follows a similar path as modelling deduction and induction. 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, 60].

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 [61]. The first one explains decisions. The second explains how people interpret the world.

The XAI methods and techniques that have been developed in research so far span from rule-based explanations and attention mechanisms [62] to visual explanations [63], Interpretable ML (IML) models [64] (see Definition 8), and ethical variations [65] to the FAIR (Findable, Accessible, Interoperable, Reusable) model development [66]. 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 (see Sect. 4) [67, 68]. The most advanced method in literature to represent explanations, applicable also to AI system decisions, is the Logocratic Method (LM) [69], which explains the nature of arguments [70] (to be discussed in Sect. 7).

2.4.2 Trustworthy artificial intelligence

Addressing the seven challenges (precisely defined by the HLEG [15]) is a priority for Trustworthy Artificial Intelligence (TAI) (see Definition 6). Here, as part of the TAI movement, the field of responsible governance, with an eye on transparency and accountability, has been developed. The TAI field investigates how to develop standards and processes to make AI safe (or at least safer). The discussion about TAI and the relevant responsible governance standards is currently in development. The urgency for clarifying their content comes from the observed tendency in society for the development of AI applications.[13] 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 [71]. 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 [72]. Here, we remark in passing that we consider liability measures in parallel with safety measures [73]. Usually, we see that with the introduction of new technologies, liability regimes adjust to correspond to the latest needs [74].

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 substantial. For interested readers, we refer to the broad and deep analysis of trustworthiness, by which variable principles, from *reliability* and *accuracy* to *sustainability* and *democracy*, are included [75]. Such principles may guide the ethical and legally trustworthy design of AI systems via the rule of law by focussing on properties including *transparency*, *verifiability* and *explainability* [76].

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 [77]. The same holds for PLT, which is seen as an AI technology.

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

[13] 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.

2.4.3 Regulating artificial intelligence

Even though society wants to be able to trust AI, they are still afraid of the positive answers to the challenges posed by XAI and TAI. In the first place, all governments in the world wish to protect their society from suffering fears. Therefore, the EU is attempting to take the lead in the movement of TAI [78] via the research of the HLEG.[14] In passing, we note that the United States[15] and China[16] are also developing regulatory efforts. The European Commission (EC) has spearheaded research on this topic with a White Paper by identifying potential risks of AI affecting society from fundamental rights and privacy to industrial safety and legal liability [79]. Due to the significant focus on the Ethics of AI, an increasing amount of research in guidelines has been discussed in academic and political circles, leading to what some call the "AI ethics boom" [80]. The results are so far accessible in the Product Liability Directive (PLD) and the Artificial Intelligence Liability Directive (AILD).[17]

Regulating AI is challenging because we do not fully comprehend AI [81]. People are still debating about the appropriate definition for AI [82].[18] In the meantime, researchers are investigating the relevant ethical framework to guide any legal or social regulatory reform regarding AI [83]. We observe the three primary regulatory efforts in China, the US and Europe.

China has opted for an incremental regulatory approach following the developments of AI. The three core regulatory initiatives from the People's Republic of China are PRC Regulation I, regulating recommendation algorithms; PRC Regulation II, regulating synthetically created content; and PRC Draft Regulation III, recommending regulation on Generative AI [84]. Despite an observed difference in the motivations supporting the regulatory initiative from the Chinese Government, specific regulatory parameters are also observed in the Western (US and EU) efforts, paving the way for some consensus in international AI regulation [84].

The US is in the process of developing regulation for AI.[19] Following the Executive Order of President Biden, taking into consideration the opinions of leading executives from AI institutions in the US,[20] the direction of the US about regulating AI is becoming clearer.[21] The regulatory direction aims at strengthening AI governance, advancing responsible AI innovation, and managing risks from the use of AI, without adopting a risk-based approach as the EU proposes. The consensus of the democratic party, supported by Majority Leader Schumer is inclined to support a "SAFE Innovation Framework for AI", where AI is seen as central driver for US economic growth and protecting American values.[22] The US seems to head towards the direction where business organisations have a significant degree of freedom to develop AI, for so long as they comply with fundamental safety principles.

In the EU, the EU AI Act has taken multiple forms over several iterative attempts to clarify how to regulate AI - with the latest text adopted in 2023 [85], which is further amended (in December 2023) following the most recent negotiations.[23] The current new version amends the original proposal of the European Committee and seems to be a serious candidate for the AI Act. Our expectation now is that the AI Act is (almost) in its proposal stage after the December 2023 discussions, which are validated by the latest press release.[24] A final acceptance is further supported by a preparatory voting in a committee meeting in which it was decided to bring the then current version of the proposal to the first official meeting of the General EU Committee meeting where the final voting on acceptance may take place. If it still unexpectedly does not happen, then postponements should take place owing to the elections of the EU. Concentrating on the contents of the AI Act, we here remark that central to the EU is the topic of *safety*, which is visible by the reliance of the AI Act on progressing the *product safety regulation*. The EU regulatory proposal distinguished from its start among unacceptable risk (total ban), high risk (higher degree of regulation), and limited risk AI systems (voluntary transparency standards), and foundation models (registration in EU database) for Generative AI models (copyright disclosures, prohibition of illegal content) [20, 85]. Article 4a (1) is relevant to rule-based explainability,

[14] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

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

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

[17] https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence.

[18] https://www.euractiv.com/section/artificial-intelligence/news/oecd-updates-definition-of-artificial-intelligence-to-inform-eus-ai-act/.

[19] https://www.whitehouse.gov/omb/briefing-room/2023/11/01/omb-releases-implementation-guidance-following-president-bidens-executive-order-on-artificial-intelligence/.

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

[21] https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

[22] https://www.democrats.senate.gov/news/press-releases/majority-leader-schumer-delivers-remarks-to-launch-safe-innovation-framework-for-artificial-intelligence-at-csis.

[23] https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

[24] https://www.consilium.europa.eu/nl/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/.

which requires developers and AI users to use their best efforts for each principle of transparency as laid out in the regulation [85]. According to the press release, consumers have a right to request for meaningful explanations. Such a right further solidifies the justification for our research on rule-based explainability.[25] Moreover, the press release provides updates on general purpose AI systems and clarifies further high-risk and prohibited AI systems.[26] The EU AI Act follows a more robust risk management approach than prior research efforts from the EU bodies due to supporting research by the EC's White Paper [79] and the HLEG.[27]

When combining the Chinese, American and European approaches, we may find similarities and differences. Research shows that, in general, all regulatory approaches agree on fundamental risks and requirements [86]. The risks are black-box models, privacy violations, bias, and discrimination; the requirements are algorithmic transparency, human understandable explanations, privacy-preserving algorithms, data cooperatives, and algorithmic fairness [86]. However, in conclusion, the three regulations differ on the specific regulatory approach (e.g., risk-based vs non-risk-based) and the nature of compliance standards.

2.4.4 Transparency and accountability

While researchers and regulators worldwide investigate the safety of ethical principles in the design of AI systems, a straightforward and concrete challenge appears because of our inability to focus on one (or a few) of the divergent ways of protecting society from AI [14]. The primary motivator behind this challenge is the misalignment between (1) levels of actual trust and (2) the trustworthiness of applied AI [16]. As a direct follow-up, we mention the contribution by 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 advocates 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 - if possible to be realised - will bring clarity in AI development and ethical auditing of AI algorithms [65]. For example, when large multinational organisations are subject to Ethics-Based Auditing (EBA), they will 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 [87]. The ethical design of AI will then be the guideline to (1) the organisations and (2) the social systems that create AI [87]. All in all, ethics will then arise in the context of the socio-technical systems that create them [88]. Munn's inter-sectional ethics approach becomes feasible with the diverse inclusion of ethical practices within organisations, nudging towards the institutionalisation of ethics [89] and the re-evaluation of AI business practices [90]. Consequently, evaluating group values and interests *becomes possible* as well as a fair comparison of the personal values with group values [78]. In practice, despite the desire from developers and designers of AI systems to adopt more practically ethical approaches, a gap is observed with systematic practices that can direct their operations despite multiple attempts to make sense of the regulatory requirements [91, 92].

It is due to the challenge of translating ethical concepts into practical solutions for AI development and the implementation of the results that the field of AI Ethics-By-Design has emerged [93, 94]. The field addresses vital ethical concerns in the ethical development of AI, such as: *how can and should we develop ethically-aware AI agents whose behaviour is adaptable to socio-ethical contexts* [95]? To nurture such development, the experts involved in AI development should find consensus in the principal values to guide the design of AI systems [96, 97]. Designing such ethically aware AI agents will impact policy-making by creating the need for investigating the establishment of legal protection for AI agency [98]. Public opinion will also affect such policy-making, whose perception of the topic is far from reaching any consensus [99].

3 Research methodology

The research methodology concentrates on three distinct approaches: fieldwork, case studies and legal framework application. First, the fieldwork focusses on understanding the essential data for AI-based PLT processing. The data are called *Proactive Data* (see Definition 9) and are based on our previous research [31]. Second, the basis for selecting case studies is Legalcomplex's list of LegalTech solutions.[28] After validating to what extent Proactive Data applies to the LegalTech solutions, we selected three case studies from

[25] https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

[26] https://www.consilium.europa.eu/nl/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/.

[27] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

[28] https://legalcomplex.com.

Table 1 Categorisation of 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

the LegalTech applications. The aim is to develop Proactive Data explanations. Third, we clarify the AI-Act liability framework to apply to the three case studies in a comparative setting. For the comparative significance, concerning the focus of the AI-Act on risk-based AI, we directed the selection towards *three* categories of case studies (viz. high-risk, mid-risk and low-risk case studies). Even though the AI Act proposal/amendment distinguishes between high-risk and limited-risk AI, for practical research purposes, we have divided limited risk into mid-risk and low-risk to facilitate the creation of more detailed research findings and to show the practical difference between levels of limited risk and their impact. We do not discuss unacceptable AI or Generative AI in the methodology.

3.1 Field work

According to the literature, the most advanced methods for preventing legal problems focus on contract risk management [43]. So, we ground our research on effectivity of *contract automation* and iContracts [7, 26]. iContracts are based on the hybrid approach between human and computer interventions that aim to achieve full automation with self-executing contracts [102]. Owing to their compliance with Hybrid AI principles, contract automation enables state-of-the-art innovations [103].

While investigating contract automation, we observed that two main processes are not given sufficient attention, viz. (1) the *communication process* preceding the drafting of contracts and (2) the *risk analysis* during the drafting of contracts [26]. We subsequently set out to validate our observations [7]. We used as 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 [7].

Surprisingly, to the researchers and the owner of Legalcomplex, there was *no solution* focussing on *both* contract communications and contract risk, which we consider a 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 [26]. 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 [100]. 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) [101]. Fourth, we *validated* the Onassis and Enriched Bow-Tie Ontology structures [31].

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) helps identify measures that will reduce the likelihood of the occurrence of a hazardous event [31]. In summary, the impact of Proactive Data is that (1) it is able to reduce the likelihood of a dispute occurring by improving contract drafting and (2) its quality assessment is feasible via the Logocratic Method (LM) (see Sect. 7) [69].

3.2 Case studies

Given that the development of categorisation criteria is a complex process, we are pleased to report that we were given access to the categorisation used by Legalcomplex. They have been categorising and recording LegalTech solutions for several years, The categorisation is the best one available. Legalcomplex provided us with information of six categories of LegalTech applications listed in Table 1. The six categories of solutions are: 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 assumed that an AI system would be able to advise 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 predictive policing during a riot.

To represent the Proactive Data and their explanations, we will use generated data by ChatGPT. Generated explainable proactive data are produced based on a question that seeks explanation (explanandum) in compliance with the Logocratic Method.

- For the *low-risk* case study, the explanandum is: What insurance should we provide to a client who bought his/her first car [client 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 predictive policing to predict and prevent potential harm to citizens, even if the predictive policing system may consider some of the rioters sufficiently dangerous?

3.3 Liability framework application

The EU is still investigating an appropriate liability regime for regulating AI [73]. From the beginning, the general academic opinion supports a strict liability regime, proposed in a way that does not discourage innovation [72]. Researchers focus on a risk-based approach, whereas the riskiest AI should be strictly liable, with specific uses of AI being prohibited [73]. Indeed, researchers support that having a liability regime for AI will benefit society and the industry.[29]

[29] 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

Table 2 Case study 1: low risk—using Lemonade to purchase car insurance

	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

Question: What insurance should we provide to a client who bought their first car, [client is 27 years old, and (s)he has been caught drinking when (s)he was underage]?

Table 3 Case study 2: mid risk—using OpenAI for creating a construction plan

	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

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?

Table 4 Case study 3: high risk—using Palantir Technologies for applying police force during a riot

	Most serviceable explanation	Less serviceable explanation 1	Less serviceable explanation 2
Risk source	Riot poses an immediate threat to public safety	Concerns about predictive policing' judgement are unwarranted	The riot situation is not as dangerous as it seems; no system needed
Proactive control	Deploy predictive policing system for rapid response	Human intervention is sufficient for handling the situation	Wait for more information about the predictive policing readiness
Hazardous event	Potentially wrongful prosecution of rioter	Rioters' intentions are not as harmful as they appear	Predictive policing judgement may not be harmful, no risk

Question: During a scary, fast-developing riot in the middle of the city centre, should we employ predictive policing to predict and prevent potential harm to citizens, even if the predictive policing system may consider some of the rioters sufficiently dangerous?

The EU started working on a legislative reform investigation in 2015.[30] 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 [104]. The PLD proposes a fault-based liability regime, although, since its establishment in 1985, it has not covered 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 repair the gaps in the PLD and aims to establish a strict liability regime for high-risk AI systems. However, not all academics agreed, and some proposed that different AI systems should adhere to different liability regimes [105]. Also, at this moment (see Sect. 2.4.3), according to some academics, the development of limited-risk AI systems to which no strict liability applies requires compliance with transparency standards. Nevertheless, neither the AI Act nor the PLD provides clear guidelines on handling liability challenges arising from such systems. For the case of Generative AI, a higher level of transparency is required, although some liability challenges will remain unsolved, as we expect. The latest working version of the AI Act in 2023 and following discussions aim at addressing these challenges [85]. Overall, the journey towards an appropriate governance framework for AI is long, and trustworthiness is continuously developing and improving as we go along [106].

Footnote 29 (continued)

European consumers, European Parliament.

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

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

4 Results

The results of our research guided by RQ1, RQ2, RQ3 and the PS are given in this section. First, they highlight that Proactive Data are identifiable in all LegalTech categories and that their explanation can be made feasible, as shown by the three case studies. Second, the results reveal legal and ethical gaps when applying the liability framework of the provisional AI-Act to the case studies. Third, XAI and TAI are quite helpful in answering RQ1, RQ2, RQ3, and the PS.

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[31]. Proactive Data was successfully applied to them. The result was (1) proving that PLT is relevant for all defined LegalTech categories and (2) convincingly validating the relevance of PLT for all LegalTech domains. As stated above, for this article, we selected three case studies, each category representing 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 4 includes Case Study 3, the *high-risk* case study examining the use of Palantir Technologies for applying predictive policing 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 are shown; they include the most serviceably plausible explanation and, after that, two potentially "disqualifying" explanations (called less serviceable). The data generated for the three case studies differ contextually depending on the relevant questions for each case study.

4.2 Legal and ethical gaps

As stated earlier, the provisional AI Act proposes a strict liability regime for high-risk AI systems (Case Study 3). It means that for low-risk (Case Study 1) and mid-risk (Case Study 2) AI systems (limited risk under AI-Act), the AI-Act is *partially applicable* with voluntary compliance standards.

Case Study 2 shows that the reasoning followed by the generated data is different for human experts, who are able to recognise the risk of a lawsuit from a neighbour. What do we learn from this consideration? Even though the AI recommended a proactive control without considering its consequences, a human expert may decide to follow the advice. In this case, if the human follows the advice, then the human is facing a risk of a lawsuit and can hardly put liability on the AI system.

Case Study 1 is a relatively straightforward case. The level of risk is low, and the advice proposed by the generated data complies with the usual direction that a human expert would take. Hence, humans may follow advice without necessarily being concerned with the consequences.

Case Study 3, however, is more complex. If we assume 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

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

to the AI Act, the AI should be held strictly liable, and the official may or may not develop a court defence based on this reasoning. However, in the case of a court defence, applying strict liability may be unfair, because the official essentially interprets the explanations provided by the AI (except if the official is not involved in the final decision-making). Therefore, we are curious to see how Hybrid Intelligence works in the future. Depending on other interpretative explanations of the officer, we are inclined to follow and interpret the other lines of reasoning and compare them to the AI system's. If the officer mindlessly follows the AI's advice and the appearance of wrongful predictive policing 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 the official provides a different explanation, and, eventually, the risk occurs, we can still re-investigate the official's line of reasoning and compare it to the machine's. Arriving at the very essence of this case, in our opinion, we should show more accuracy in assigning liability. Of course, a potential defence might be that an officer may 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.

The gaps we identified are divided into three categories: (1) transparency, (2) accountability, and (3) liability. On a transparency level, we need more visibility over explanations supporting decisions. As for accountability, the lack of explanations makes it challenging to hold specific parties accountable. Then, about liability, it becomes hard to assign liability to responsible parties fairly. All three categories have direct implications for law and ethics and, as a consequence, are the primary legal and ethical gaps observed based on the case studies.

As seen in Table 2, the generated 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.

As for Table 3, the generated 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 generated 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 generated data do not recognise it as an actual risk.

As for Table 4, the proposed proactive control is the deployment of predictive policing for rapid prediction, even when there is a risk of potential wrongful judgement. As given above, one of the less serviceable explanations is waiting for more information about the predictive policing system's readiness, considering that the system can arrive at a wrong judgement. It is a convincing example of generated data because it shows that the official eventually should take the decision-making in conjunction with the advice received from the technological system. If the official faces alternative explanations, yet before deciding, the official should interpret the proposal suggested by the PLT.

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.

4.3 Explainable and trustworthy preventive legal technology

The PS of this research is: *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 provide the answers to RQ1, RQ2, and RQ3 and finally provide an answer to 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.

5 Discussion and implications

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.[32] 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, Sect. 3.1) is that it applies to any risk level. Therefore, it is possible 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.

The case studies validate that generating explainable proactive data is possible, even based on generated 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 generated 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 is 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.

5.2 Ethical implications

The main ethical 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 makes explainability possible. The degree of transparency depends on how an explanation is expressed and accessible. The higher the transparency of the motivations supporting an explanation accompanied by explicit data, the higher 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. Explicit explanations and transparent interpretations that enable accountability support public participation, legal certainty and consistency and can help reflect relevant fundamental rights more easily [77]. As a consequence, Hybrid Intelligence will be enabled [107].

The ethical implications are that more transparency is generated with explainability, which directly impacts accountability. With higher accountability, assigning liability becomes more responsible, thus leading towards LTAI. Our research recognises that liability connects inextricably with the explanatory process supporting AI across its development and application chain. Explanations are applicable on multiple levels but are usually hidden or implicit today. Hence, we highlight the importance of surfacing explanations and the positive ethical impact such surfacing entails.

5.3 Legal implications

Applying the legal framework to the case studies shows that regulating AI technology can be considered a generic approach for applying liability only in 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 risk, we adjust that the quality of explanations deserves the utmost attention. For now, lawyers and legal researchers aim to insert humans in the loop to improve the AI systems' responsibility for explanations. Our results show how shared liability may become possible depending on the distribution of mistakes throughout the explanation chain.

Moreover, they show the extent to which it is possible to provide explanations in light of the AI Act, which provides the right to consumers to launch complaints and receive meaningful explanations.[33]

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

[33] https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

The consideration of robot rights as equal to human rights for establishing a proportional shared liability model can be argued as excessive from the Futuristic AI movement perspective. 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.

The authors opine that current regulatory efforts need to balance social protection and innovation. On the one hand, the second author's opinion (H. J. van den Herik) is that within 50–80 years, robots will outperform human beings in their *quality of thinking*. For this reason, during the European Conference on Artificial Intelligence (ECAI) 2023, he proposed the development of an international treaty similar to that of nuclear weapons [108]. On the other hand, accelerating AI development is crucial, and by adopting an "explanation culture", its effects can be mitigated to a large degree. As long as regulators continue to approach AI development as a black box from the Futuristic AI perspective, innovation will be hampered, and society will not be able to benefit from the positive effects of AI. Still, achieving consensus on an international level is vital to maintaining the focus of AI development in socially positive directions.

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 improves the trustworthiness of PLT while increasing ethical transparency and accountability, directly affecting ethical AI research, LTAI, and AI Legal Liability regulation efforts.

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 need for more explanations across the AI chain, inductive explanations seem complicated today. 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. It is essential considering the rapid adoption of Generative AI technology, the legal implications of which are still being investigated and pose a significant challenge to regulation efforts. Finally, a severe challenge and exciting avenue is investigating the combination of rule-based explainability with statistical explainability models.

7 Further research

The Logocratic Method shows that decision-making is, in essence, based on abductive reasoning, in which explanations play a fundamental role [109]. 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.

So far, the LM has not been applied to rule-based XAI in literature. According to the LM, there is an important distinction between *identifying* and *evaluating* arguments [109]. 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) [107].

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 [29]. Eventually, the systematic evaluation of AI explanations and interpretations will facilitate the evaluation of underlying values, principles and laws, contributing to greater trustworthiness [110]. Studying how the LM contributes to XAI will help develop an "explanation culture" that can contribute towards more TAI.

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

Author Contributions GS is the principal author. JH acted in the double role of co-worker and supervisor. The authors thank Raymond Blyd (Legalcomplex) for the pleasant and fruitful cooperation.

Funding There is no funding for the research.

Data Availability The data supporting all tables are directly available by reading this article and are not accessible in any other repository.

Declarations

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. Minn. Law Rev. **33**, 455 (1949)
3. Rissland, E.L., Ashley, K.D.: HYPO: A Precedent-Based Legal Reasoner. Department of Computer and Information Science, University of Massachusetts, Connecticut (1987)
4. Herik, H.J. van den: Can computers judge court cases? (*Kunnen Computers* Recht*spreken?*). Gouda Quint, Inaugural address Leiden University, 21st June 1991, Arnhem (1991)
5. Susskind, R.E.: The Future of Law: Facing the Challenges of Information Technology. Oxford University Press, Oxford (1996)
6. Evans, P.: Ari: The Life and Times of Aristotle Socrates Onassis. Summit Books, New York (1986)
7. Stathis, G., Trantas, A., Biagioni, G., Graaf, K.A. de, Adriaanse, J.A.A., Herik, H.J. van den: Designing an intelligent contract with communications and risk data. Springer Nature: Recent Trends on Agents and Artificial Intelligence (Submitted) (2023)
8. Papinianus: The Lawyer. Hestia Publishers & Booksellers, Athens, Greece (2003) - Translated from Greek, see: Παπινιανός, Ο δικηγόρος, Βιβλιοπωλείον της Εστίας, Αθήνα, Ελλάδα (2003)
9. Simion, M., Kelp, C.: Trustworthy artificial intelligence. Asian J. Philos. **2**(1), 8 (2023)
10. Ayling, J., Chapman, A.: Putting AI ethics to work: are the tools fit for purpose? AI Ethics **2**(3), 405–429 (2022)
11. Hagendorff, T.: Blind spots in AI ethics. AI Ethics **2**(4), 851–867 (2022)
12. Eschenbach, W.J. von: Transparency and the black box problem: why we do not trust AI. Philos. Technol. **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. Springer (2019)
14. Munn, L.: The uselessness of AI ethics. AI 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)
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 Commission: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised 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. Stanford-Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021 (2021)
19. European Union: Article 22. Off. J. Eur. Union **59**(4), 46 (2016) (**Regulation (EU) 2016/679 (GDPR)**)
20. Bradford, A.: Digital empires: the global battle to regulate technology. Oxford University Press, Oxford (2023)
21. Susskind, R.: Online courts and the future of justice. Oxford University Press, Oxford (2019)
22. Katsh, M.E., Rabinovich-Einy, O.: Digital justice: technology and the internet of disputes. Oxford University Press, Oxford (2017)
23. Barton, T.D.: Preventive law and problem solving: lawyering for the future. Vandeplas Publishing, Florida (2009)
24. Brown, L.M.: Preventive Law. Prentice Hall, New York (1950)
25. International Organization for Standardization: ISO 31022:20220 Risk management: Guidelines for the management of legal risk (2020)
26. Stathis, G., Trantas, A., Biagioni, G., Herik, H.J. van den, 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)
27. Sander, F.E., Rozdeiczer, L.: Selecting an appropriate dispute resolution procedure: detailed analysis and simplified solution. In: Moffitt, M.L., Bordone, R.C. (eds.) The Handbook of Dispute Resolution, pp. 386–406. Wiley (2005)
28. Longo, L.: Explainable Artificial Intelligence: First World Conference, xAI. Springer, Lisbon, Portugal, July 26–28, 2023, Proceedings, Part II (2023)
29. 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. Artif. Intell. Rev. **56**(4), 3473–3504 (2023)
30. Ersoz, B., Sagiroglu, S., Bulbul, H.I.: Methods of explainable artificial intelligence, trustworthy artificial intelligence and interpretable machine learning in renewable energy. Int. J. Smart Grid ijSmartGrid **6**(4), 136–143 (2022)
31. Stathis, G., Biagioni, G., Graaf, K.A. de, Trantas, A., Herik, H.J. van den.: The value of proactive data for intelligent contracts. World Conference on Smart Trends in Systems, Security and Sustainability, Springer LNNS (2023)
32. Dauer, E.A.: Four principles for a theory of preventive law. In: Haapio, H. (ed.) A Proactive Approach to Contracting and Law, pp. 13–33. Turku University of Applied Sciences (2008)
33. 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)
34. Wexler, D.: Therapeutic jurisprudence: an overview. TM Cooley L. Rev. **17**, 125 (2000)

35. 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)
36. Lessig, L., Nesson, C., Zittrain, J.: Open code-open content-open law: building a digital commons. Beckman Center for Internet and Society Harvard Law School, Cambridge (1999)
37. Corrales, M., Fenwick, M., Haapio, H.: Digital technologies, legal design and the future of the legal profession. In: Corrales, M., Fenwick, M., Haapio, H. (eds.) Legal Tech, Smart Contracts and Blockchain, pp. 1–15. Springer (2019)
38. Corrales, M., Fenwick, M., Haapio, H., Vermeulen, E.P.: Tomorrow's lawyer today? Platform-driven LegalTech, smart contracts & the new world of legal design. J. Internet Law **22**(10), 3–12 (2019)
39. Iversen, J.: Legal Risk Management. Forlaget Thomson Gad-Jura, Copenhagen (2004)
40. Mahler, T., Bing, J.: Contractual risk management in an ICT context: searching for a possible interface between legal methods and risk analysis. Scand. Stud. Law **49**, 339–357 (2006)
41. Barton, T.D.: Preventive Law: A Methodology for Preventing Problems. National Centre for Preventive Law, San Diego (2002)
42. Barton, T.D.: Thinking Preventively and Proactively. Stockholm Institute for Scandinavian Law, Stockholm (1957)
43. Haapio, H., Siedel, G.J.: A Short Guide to Contract Risk. Routledge, Oxfordshire (2013)
44. Esayas, S., Mahler, T.: Modelling compliance risk: a structured approach. Artif. Intell. Law **23**(3), 271–300 (2015)
45. Barton, T.D.: Re-designing law and lawering for the information age. Notre Dame J. Law Ethics Pub. Policy **30**, 1 (2016)
46. Barton, T.D., Berger-Walliser, G., Haapio, H.: Contracting for innovation and innovating contracts: an overview and introduction to the special issue. J. Strateg. Contract. Negot. **2**(1–2), 3–9 (2016)
47. Whalen, R.: Computational Legal Studies: The Promise and Challenge of Data-Driven Research. Edward Elgar Publishing, Cheltenham (2020)
48. Stathis, G.: The shock of legal tech: no one ignorant of technology should read this. Legal Bus World (7), 54–59 (2018)
49. Ashley, K.D.: Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press, Cambridge (2017)
50. Fitzgerald, B.: Cyberlaw: International Library of Essays in Law and Legal Theory. Second Series. Ashgate Publishing Group, Britannica Academic, Encyclopedia Britannica, London (2021)
51. De Franceschi, A., Schulze, R.: Digital Revolution-New Challenges for Law: Data Protection, Artificial Intelligence, Smart Products, Blockchain Technology and Virtual Currencies. Carl Heinrich Beck & Nomos, Munich (2019)
52. Turing, A.M.: Computing machinery and intelligence. Mind **49**(433), 460 (1950)
53. Larson, E.J.: The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do. Harvard University Press, Cambridge (2021)
54. Herik, H.J. van den: Computers and intuition. ICGA J. **38**(4), 195–208 (2015)
55. Herik, H.J. van den: Intuition is Programmable. Valedictory Address, Tilburg University, Tilburg (2016)
56. McWhinney, W.: Grammars of Engagement. Human and Organizational (2002)
57. Peirce, C.S.: Harvard Lectures on Pragmatism. In: Collected Papers, vol. 5, pp. 188–189. Belknap Press of Harvard University Press, Cambridge (1903)
58. 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, Carl Heinrich Beck, München (2023)
59. Mueller, J., Massaron, L.: Artificial Intelligence for Dummies. For Dummies, Newark (2018)
60. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.-Z.: XAI—Explainable artificial intelligence. Sci. Robot. **4**(37), 7120 (2019)
61. Belém, C., Balayan, V., Saleiro, P., Bizarro, P.: Weakly supervised multi-task learning for concept-based explainability. arXiv preprint http://arxiv.org/abs/2104.12459 (2021)
62. Niu, Z., Zhong, G., Yu, H.: A review on the attention mechanism of deep learning. Neurocomputing **452**, 48–62 (2021)
63. Kovalerchuk, B., Ahmad, M.A., Teredesai, A.: Survey of explainable machine learning with visual and granular methods beyond quasi-explanations. In: Pedrycz, W., Chen, S.M. (eds.) Interpretable artificial intelligence: a perspective of granular computing, Studies in Computational Intelligence, vol. 937, pp. 217–267. Springer (2021)
64. 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. IEEE (2021)
65. Mökander, J., Floridi, L.: Ethics-based auditing to develop trustworthy AI. Minds Mach. **31**(2), 323–327 (2021)
66. Adhikari, A., Wenink, E., van der 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)
67. Akyol, S.: Rule-based explainable artificial intelligence. In: Pioneer and Contemporary Studies in Engineering, pp. 305–326 (2023)
68. Waa, J. van der , Nieuwburg, E., Cremers, A., Neerincx, M.: Evaluating XAI: A comparison of rule-based and example-based explanations. Artif. Intell. **291**, 103404 (2021)
69. Brewer, S.: Logocratic method and the analysis of arguments in evidence. Law Probab. Risk **10**(3), 175–202 (2011)
70. Brewer, S.: Interactive virtue and vice in systems of arguments: a logocratic analysis. Artif. Intell. Law **28**, 151–179 (2020)
71. Gerven, W. van, Droshout, D., Lever, J.F., Larouche, P.: Tort Law: Ius Commune Casebooks for the Common Law of Europe. Hart Publishing, Oxford (2001)
72. Tjong Tjin Tai, E.: Liability for (semi) autonomous systems: robots and algorithms. In: Mak, V., Tjong Tjin Tai, E., Berlee, A. (eds.): Research Handbook on Data Science and Law, pp. 55–82. Edward Elgar, Cheltenham (2018)
73. Wendehorst, C.: Strict liability for AI and other emerging technologies. J. Eur. Tort Law **11**(2), 150–180 (2020)
74. Gifford, D.G.: Technological triggers to tort revolutions: steam locomotives, autonomous vehicles, and accident compensation. J. Tort Law **11**(1), 71–143 (2018)
75. Varona, D., Suárez, J.L.: Discrimination, bias, fairness, and trustworthy AI. Appl. Sci. **12**(12), 5826 (2022)
76. Chatila, R., Dignum, V., Fisher, M., Giannotti, F., Morik, K., Russell, S., Yeung, K.: Trustworthy AI. In: Braunschweig, B., Ghallab, M.: Reflections on Artificial Intelligence for Humanity, pp. 13–39. Springer (2021)
77. 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. SSRN 3899991 (2021)
78. Rieder, G., Simon, J., Wong, P.-H.: Mapping the stony road toward trustworthy AI: expectations, problems, conundrums. In: Pellilo, M., Scantamburlo, T. (eds.) Machines We Trust: Perspectives on Dependable AI, pp. 3–27. MIT Press (2021)
79. European Commission: White paper: On Artificial Intelligence - A European approach to excellence and trust. Eur Lex, European Union (2020)
80. Corrêa, N.K., Galvão, C., Santos, J.W., Del Pino, C., Pinto, E.P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., et al.: Worldwide AI ethics: a review of 200 guidelines and recommendations for AI governance. Patterns **4**(10), 100857 (2023)
81. Vihul, L.: International Legal Regulation of Autonomous Technologies. Centre for International Governance Innovation, Waterloo (2020)
82. Fuzaylova, E.: War torts, autonomous weapon systems, and liability: why a limited strict liability tort regime should be implemented. Cardozo Law Rev. **40**, 1327 (2018)
83. Bartneck, C., Lütge, C., Wagner, A., Welsh, S.: An Introduction to Ethics in Robotics and AI. Springer, Cham (2021)
84. Sheehan, M.: China's AI Regulations And How They Get Made. Carnegie Endowment for International Peace (2023)
85. European Parliament: Artificial Intelligence Act: Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Eur Lex, European Union (2023)
86. Rios-Campos, C., Vega, S.M.Z., Tejada-Castro, M.I., Viteri, J.D.C.L., Zambrano, E.O.G., Gamarra, J.M.B., Núñez, J.B., Vara, F.E.O.: Ethics of artificial intelligence. S. Fl. J. Dev. **4**(4), 1715–1729 (2023)
87. Mökander, J., Floridi, L.: Operationalising AI governance through ethics-based auditing: an industry case study. AI Ethics **3**(2), 451–468 (2023)
88. Stahl, B.C.: From computer ethics and the ethics of AI towards an ethics of digital ecosystems. AI Ethics **2**(1), 65–77 (2022)
89. Schultz, M.D., Seele, P.: Towards AI ethics' institutionalization: knowledge bridges from business ethics to advance organizational AI ethics. AI Ethics **3**(1), 99–111 (2023)
90. Attard-Frost, B., De los Ríos, A., Walters, D.R.: The ethics of AI business practices: a review of 47 AI ethics guidelines. AI Ethics **3**(2), 389–406 (2023)
91. Sanderson, C., Douglas, D., Lu, Q., Schleiger, E., Whittle, J., Lacey, J., Newnham, G., Hajkowicz, S., Robinson, C., Hansen, D.: AI ethics principles in practice: perspectives of designers and developers. IEEE Transactions on Technology and Society **4**(2) (2023)
92. Agbese, M., Mohanani, R., Khan, A., Abrahamsson, P.: Implementing AI ethics: making sense of the ethical requirements. In: Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering, pp. 62–71 (2023)
93. d'Aquin, M., Troullinou, P., O'Connor, N.E., Cullen, A., Faller, G., Holden, L.: Towards an "ethics by design" methodology for AI research projects. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 54–59 (2018)
94. Michael, K., Abbas, R., Roussos, G., Scornavacca, E., Fosso-Wamba, S.: Ethics in AI and autonomous system applications design. IEEE Trans. Technol. Soc. **1**(3), 114–127. IEEE (2020)
95. Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L., Génova, G., Haim, G., Kließ, M.S., Lopez-Sanchez, M., et al.: Ethics by design: necessity or curse? In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 60–66 (2018)
96. Gerdes, A.: A participatory data-centric approach to AI ethics by design. Appl. Artif. Intell. **36**(1), 2009222 (2022)
97. Muhlenbach, F.: A methodology for ethics-by-design AI systems: dealing with human value conflicts. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1310–1315. IEEE (2020)
98. Iphofen, R., Kritikos, M.: Regulating artificial intelligence and robotics: ethics by design in a digital society. Contemp. Soc. Sci. **16**(2), 170–184 (2021)
99. Kieslich, K., Keller, B., Starke, C.: Artificial intelligence ethics by design: evaluating public perception on the importance of ethical design principles of artificial intelligence. Big Data Soc. **9**(1), 20539517221092956 (2022)
100. Stathis, G., Biagioni, G., Trantas, A., Herik, H.J. van den, 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)
101. Stathis, G., Biagioni, G., Trantas, A., Herik, H.J. van den: Risk visualisation for trustworthy intelligent contracts. In the Proceedings of the 21st International Industrial Simulation Conference (ISC), EUROSIS-ETI, pp. 53–57 (2023)
102. Mason, J.: Intelligent contracts and the construction industry. J. Leg. Aff. Disput. Resolut. Eng. Constr. **9**(3), 04517012 (2017)
103. 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 (2020)
104. Cabral, T.S.: Liability and artificial intelligence in the EU: assessing the adequacy of the current product liability directive. Maastricht J. Eur. Comp. Law **27**(5), 615–635 (2020)
105. Bertolini, A., et al.: Artificial Intelligence and Civil Liability. European Parliamentary Research Service, Brussels, European Union (2020)
106. Smuha, N.A.: The EU approach to ethics guidelines for trustworthy artificial intelligence. Comput. Law Rev. Int. **20**(4), 97–106 (2019)
107. 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)
108. Herik, H.J. van den: History and Future of AI. In: European Conference on Artificial Intelligence, Krakòw, Poland, Invited Lecture, October 3 (2023)
109. Brewer, S.: First Among Equals: Abduction in Legal Argument from a Logocratic Point of View. Oxford Jurisprudence Discussion Group, University of OxfordSchool of Law, Oxford (2022)
110. Winikoff, M., Sidorenko, G., Dignum, V., Dignum, F.: Why bad coffee? Explaining BDI agent behaviour with valuings. Artif. Intell. **300**, 103554 (2021)