# Preventing disputes: preventive logic, law & technology

Stathis, G.

# Chapter 6

# Risk Identification and the AI-ACT

The Chapter addresses RQ5, which reads as follows:

> **RQ5:** *To what extent is it possible to develop an explainable and trustworthy Preventive Legal Technology?*

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. 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 under construction) 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.

The current chapter corresponds to the following publication:

Stathis, G. and van den Herik, H. J. (2024). Ethical & Preventive Legal Technology. *Springer AI and Ethics*. https://doi.org/10.1007/s43681-023-00413-2

## 6.1 Preventive Legal Technology

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 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 [Labatut, 2023]) 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 [Loevinger, 1949]. In 1987, the first reasoner for explaining the reasoning supporting judicial decisions was created [Rissland and Ashley, 1987]. In 1991, Leiden University saw a remarkable Inaugural Address [van den Herik, 1991], in which the question ”Can computers Judge Court cases?” was answered positively. Then, in 1996, Susskind predicted the shift from reactive facilities in the law (such as deciding on the resolution of a dispute) to proactive facilities (such as deciding on the prevention of a dispute) [Susskind, 1996]. This line of research will dominate the next thirty years [Scholtes, 2021]. Hence, we follow the trajectory of connecting AI with proactive facilities via the field of Preventive Law.

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 [Stathis et al., 2024]. Of course, the application of Preventive Law is not restricted to contracts only. The remainder of this Chapter 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* (Definition 6.1 [Longo, 2023] [1]), *interpretable* (see

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

Definition 6.2 [Graziani et al., 2023, Ersoz et al., 2022]) and *trustworthy* (see Definition 6.3 [High-Level Expert Group on AI, 2019]) manner. This approach has lead to three specialised branches of AI research, which is based onExplainable AI (XAI), Interpretable ML (IML) and Trustworthy AI (TAI) principles (see the Definitions below).

*Definition 6.1 –* ***Explainable Artificial Intelligence***

***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.

*Definition 6.2 –* ***Interpretable Machine Learning***

***Interpretable Machine Learning*** (IML) is a system of which it is possible to *learn* its working *principles* and *outcomes* in human-understandable language without affecting the validity of the system.

*Definition 6.3 –* ***Trustworthy Artificial Intelligence***

***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.

### 6.1.1 Towards Ethical and Preventive Legal Technology

Central to the discussion of AI is the topic of *trustworthiness* [Simion and Kelp, 2023]. Lack of trustworthiness is a genuine concern for the ethical impact and unintended consequences of new AI technologies for society [Ayling and Chapman, 2022]. The European Union (EU) Guidelines call for lawful, ethical and robust AI [2]. Here, we note explicitly that despite the various blind spots for the ethics of AI [Hagendorff, 2022], one of the main challenges of AI is that its decisions so far are *not transparent*, resulting in ”black box” decisions [von Eschenbach, 2021]. Below, we briefly introduce XAI and TAI. A literature review expands on the concepts in 2.4.1 and 2.4.2.

XAI is the field of study investigating the *explanation* of AI system decisions [Xu et al., 2019]. XAI is assumed to lead to TAI, aiming to increase society’s *trust* due to higher transparency and accountability [Munn, 2023]. The

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

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) [High-Level Expert Group on AI, 2019]. Researchers have noticed a general disconnection between levels of actual trust and trustworthiness of applied AI [Laux et al., 2023]. In order to nurture practical trustworthiness, researchers changed their focus to *transparency* and *accountability* [Munn, 2023]. 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 [European-Commission, 2021]. 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 [Kop, 2021].

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) [3]. 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 [European Union, 2016]. 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 [4]. If AIDM includes explanations, then AI systems will be more trustworthy due to higher transparency and accountability. Consequently, organisations will be able to design AIDM that is ethical and legally preventive, which is beneficial for society because it reduces the appearance of legal problems and increases legal safety. Hence, 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*.

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

[4] 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

### 6.1.2 Research Question 5 and Contribution

The preceding leads us to RQ5:

> **RQ5:** *To what extent is it possible to develop an explainable and trustworthy Preventive Legal Technology?*

To answer RQ5, we have partitioned it into three Smaller RQs (SRQ).

> **SRQ1:** *What is Preventive Legal Technology?*
>
> **SRQ2:** *To what extent is it possible to develop an explainable Preventive Legal Technology?*
>
> **SRQ3:** *To what extent is it possible to develop a trustworthy Preventive Legal Technology?*

Before addressing RQ5, 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.

### 6.1.3 Chapter Structure

To answer RQ5, we structured the Chapter as follows. In Section 6.2, we describe the literature on ethics and AI. Section 6.3 presents our three methodologies: fieldwork, case studies and applications. Then, Section 6.4 describes the investigations and states the results. Section 6.5 discusses those results and focusses on trustworthiness and ethical parameters. Finally, Section 6.6 answers RQ5 and provides our conclusion as well as further research suggestions.

## 6.2 Literature Review

Many ideas about modelling intelligent behaviour started in ancestry and were further developed throughout history [5]. Nevertheless, most researchers attribute the starting point of AI to Alan Turing in 1950 [Turing, 1950]. Since then, two main AI movements emerged: the scientific one and the futuristic one [Larson, 2021]. The scientific AI movement supports the idea that formal reasoning

[5] 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 the basis of AI and is investigating whether intelligence can become artificial. The futuristic AI movement believes that intelligence will become artificial and will influence public opinion to accept that. While this dichotomy is still vivid, the state-of-the-art of AI is not yet able to *prove* how intelligence is programmable. Researchers support that AI today assists humans with ingenuity, contributing to intelligence, not intuition [Larson, 2021]. However, many researchers are investigating how to model intuition [van den Herik, 2015]. 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 [6] [van den Herik, 2016, McWhinney, 2002]. In logic, ingenuity is modelled by deduction or induction; and intuition via abduction [Peirce, 1903, Brewer, 2023]. Admittedly, humans still do not know how to model abduction computationally [Larson, 2021]. The developments of AI follow, to a large extent, the developments in logic with modelling intelligence.

### 6.2.1 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* [Larson, 2021]. The fundamental elements for ES and ML are "normal" data [Mueller and Massaron, 2018]. With this knowledge, we are ready for the next step: Explainable AI.

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 [Xu et al., 2019, Gunning et al., 2019].

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 [Belém et al., 2021, Koster et al., 2021]. 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 [Niu et al., 2021] to visual explanations [Kovalerchuk et al., 2021], Interpretable ML (IML) models [Vollert et al., 2021], and ethical variations [Mökander and Floridi, 2021] to the FAIR (Findable, Accessible, Interoperable, Reusable) model development [Adhikari et al., 2022, Hosseini et al., 2023]. Two notable frameworks developed

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

for advanced explainability and interpretability are SHapley Additive exPlanations (SHAP), a framework for interpreting predictions of machine learning models [Salih et al., 2024], and Local Interpretable Model Agnostic Explanation (LIME) a technique that explains the predictions of any classifier in interpretable manner [Salih et al., 2024].

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 Section 4) [Akyol, 2023, van der Waa et al., 2021]. The most advanced method in literature to represent explanations, applicable also to AI system decisions, is the Logocratic Method (LM) [Brewer, 2011], which explains the nature of arguments [Brewer, 2020] (to be discussed in Section 7).

### 6.2.2 Trustworthy Artificial Intelligence

Addressing the seven challenges (precisely defined by the HLEG [High-Level Expert Group on AI, 2019]) is a priority for Trustworthy Artificial Intelligence (TAI). 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 [7]. 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 [van Gerven et al., 2001]. 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 [Tjong Tjin Tai, 2018]. In passing, we remark that we consider *liability* measures in parallel with *safety* measures [Wendehorst, 2020]. Usually, we see that with the introduction of new technologies, liability regimes adjust to correspond to the latest needs [Gifford, 2018].

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 [High-Level Expert Group on AI, 2019]. 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

[7] 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

included [Varona and Suárez, 2022]. 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* [Chatila et al., 2021].

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 [Smuha et al., 2021]. The same holds for PLT, which is seen as an AI technology.

### 6.2.3 Regulating Artificial Intelligence

Consequently, 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 [Rieder et al., 2021] via the research of the HLEG [8]. In passing, we note that the United States [9] and China [10] 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 [European-Commission, 2020]. 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” [Corrêa et al., 2023]. The results are so far accessible in the Product Liability Directive (PLD) and the Artificial Intelligence Liability Directive (AILD) [11].

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

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

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

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

Regulating AI is *challenging* because we do not fully comprehend AI [Vihul, 2020]. People are still debating about the appropriate definition for AI [Fuzaylova, 2018] [12]. In the meantime, researchers are investigating the relevant ethical framework to guide any legal or social regulatory reform regarding AI [Bartneck et al., 2021]. We observe the three primary regulatory efforts in (A) China, (B) the US and (C) Europe, and ask ourselves: (D) how to combine them from a global governance perspective?

### A: China

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 [Sheehan, 2023]. 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 [Sheehan, 2023].

### B: United States

The US is in the process of developing regulation for AI [13]. Following the Executive Order of President Biden, taking into consideration the opinions of leading executives from AI institutions in the US [14], the direction of the US about regulating AI is becoming clearer [15]. 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.

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

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

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

[15] 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/

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 [16]. 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.

### C: European Union

In the EU, the EU AI Act has taken multiple forms over several iterative attempts to clarify how to regulate AI [European-Parliament, 2023]. This version amends the original proposal of the European Committee and is a provisional candidate for the AI Act. The expectation is that the AI Act will be closer to a proposal stage as regulation during the December 2023 discussions [17], with a final acceptance in the start of 2024. If it will not happen, then postponements will take place owing to the elections of the EU. Concentrating on the contents, here we 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) [European-Parliament, 2023]. Article 4a (1) is relevant to rule-based explainability, which requires developers and AI users to use their best efforts per principles of transparency as laid out in the regulation [European-Parliament, 2023]. 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 [European-Commission, 2020] and the HLEG [18].

### D: Global Governance

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 [Rios-Campos et al., 2023]. The *risks* are black-box models, privacy violations, bias, and discrimination; the *requirements* are algorithmic transparency, human understandable

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

[17] https://datamatters.sidley.com/2023/11/17/eu-moving-closer-to-an-ai-act/

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

explanations, privacy-preserving algorithms, data cooperatives, and algorithmic fairness [Rios-Campos et al., 2023]. 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. A workable combination is open challenge to a world wide AI entity that should head the AI threat by a Silicon AI Treaty [19], as proposed by van den Herik during the European Conference on Artificial Intelligence (ECAI) 2023 (see also Subsection 6.5.3) [van den Herik, 2023].

### 6.2.4 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 [Munn, 2023]. The primary motivator behind this challenge is the misalignment between (1) levels of actual trust and (2) the trustworthiness of applied AI [Laux et al., 2023]. 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 [Munn, 2023]. 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 [Munn, 2023]. On a micro-level, he advocates two practical concepts for the design of AI: *transparency* and *accountability* [Munn, 2023]. 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 [Mökander and Floridi, 2021]. 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 [Mökander and Floridi, 2023]. The ethical design of AI will then be the guideline to (1) the organisations and (2) the social systems that create AI [Mökander and Floridi, 2023]. All in all, ethics will then arise in the context of the socio-technical systems that cre-

[19] https://www.technologyreview.com/2023/05/02/1072566/the-download-geoffrey-hintons-ai-fears-and-decoding-our-thoughts/

ate them [Stahl, 2022]. Munn's inter-sectional ethics approach becomes feasible with the diverse inclusion of ethical practices within organisations, nudging towards the institutionalisation of ethics [Schultz and Seele, 2023] and the re-evaluation of AI business practices [Attard-Frost et al., 2023]. Consequently, evaluating group values and interests *becomes possible* as well as a fair comparison of the personal with group values [Rieder et al., 2021]. 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 [Sanderson et al., 2023, Agbese et al., 2023].

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 [d'Aquin et al., 2018, Michael et al., 2020]. 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?* [Dignum et al., 2018] To nurture such development, the experts involved in AI development should find consensus in the principal values to guide the design of AI systems [Gerdes, 2022, Muhlenbach, 2020]. Designing such ethically aware AI agents will impact policy-making by creating the need for investigating the establishment of legal protection for AI agency [Iphofen and Kritikos, 2021]. Public opinion will also affect such policy-making, whose perception of the topic is far from reaching any consensus [Kieslich et al., 2022].

## 6.3 Research Methodology

The research methodology concentrates on two distinct approaches: case studies and legal framework application. First, the basis for selecting case studies is Legalcomplex's list of LegalTech solutions (see Subsection 6.3.1) [20]. After validating to what extent Proactive Data applies to the LegalTech solutions, we selected three case studies from the LegalTech applications. The aim is to develop Proactive Data explanations. Second, we clarify the AI-Act liability framework to apply to the three case studies in a comparative setting(see Subsection 6.3.2). 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 partitioned limited risk into mid-risk and low-risk to facilitate the creation of more detailed research findings and to show the

[20] `https://legalcomplex.com`

practical difference between levels of limited risk and their impact. We do not discuss unacceptable AI or Generative AI in the methodology.

### 6.3.1 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 6.1. The six categories of solutions are: (A) FinTech, (B) WealthTech, (C) RiskTech, (D) LegalTech, (E) SmartTech and (F) 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 (C1) Security, (C2) Insurance, and (C3) Governance, Risk, and Compliance (GRC). The second is SmartTech with (E1) Image Recognition, (E2) Audio Recognition, (E3) Text Analytics, (E4) Data Analytics, and (E5) Automation. Table 6.1 [21] includes specific descriptions (Column 3) for each category (Column 1) and subcategory (Column 2)—fourteen in total—and for the end users (Column 4) being top private companies (Column 5). 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 6.1. 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 (see Table 6.1, Column 5).

- The **mid-risk** solution concerns using OpenAI (SmartTech, Text Analytics) for creating a construction plan (see Table 6.1, Column 5).

- The **high-risk** solution concerns using Palantir Technologies (SmartTech, Data Analytics) for applying predictive policing during a riot (see Table 6.1, Column 5).

To represent the Proactive Data and their explanations, we will use generated data by ChatGPT. Generated explainable proactive data are produced based on

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

a question that seeks explanation (explanandum) in compliance with the LM.

- For the **low-risk** case study, the explanandum is: What insurance should we provide to a client who bought his first car (s)he is 27 years old and has been caught drinking when (s)he was underage?

- For the **mid-risk** case study, the explanandum is: When deciding to build a tall building next to a residential area, should we add a net to catch people who may fall, at the expense of a better view of the surrounding area?

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

### 6.3.2 Liability Framework Application

The EU is still investigating an appropriate liability regime for regulating AI [Wendehorst, 2020]. From the beginning, the general academic opinion supports a strict liability regime, proposed in a way that does not discourage innovation [Tjong Tjin Tai, 2018]. Researchers focus on a risk-based approach, whereas the riskiest AI should be strictly liable, with specific uses of AI being prohibited [Wendehorst, 2020]. Indeed, researchers support that having a liability regime for AI will benefit society and the industry [22]. The EU started working on a legislative reform investigation in 2015 [23]. 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 [Cabral, 2020]. The PLD proposes a fault-based liability regime, although, since its establishment in 1985, it has not covered the new AI challenges within it.

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

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

Table 6.1: LegalTech Categories

| 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 |

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 [Bertolini et al., 2020]. Also, at this moment (see 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 the following discussions aim at addressing these challenges and at accepting them in the next plenary session of the EU Parliament [European-Parliament, 2023]. Overall, (1) the *journey* towards an appropriate governance framework for AI is *long*, and *trustworthiness* is continuously *developing* and *improving* as we go along, already expected and predicted by [Smuha, 2019].

## 6.4 Research Results

The results of our research guided by SRQ1, SRQ2, SRQ3 and the RQ5 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 SRQ1, SRQ2, SRQ3, and the RQ5.

### 6.4.1 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 6.1. The application of Proactive Data to all 12 examples is accessible via GitHub [24]. From a scientific point of view, we are pleased to state that Proactive Data was successfully applied to all of them (findable details are on Github). The main 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

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

stated earlier, for a closer look in this Chapter, we selected three case studies, each category representing explainable proactive data.

- Table 6.2 includes Case Study 1, the **low-risk** case study examining the use of Lemonade for the *purchase of car insurance*.
- Table 6.3 includes Case Study 2, the **mid-risk** case study examining the use of OpenAI for creating *a construction plan*.
- Table 6.4 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 (Table 6.2, 6.3, and 6.4) 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.

### 6.4.2 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 (characterised as 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. A prevailing question is: 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 the 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 the advice without necessarily being concerned with the consequences.

Case study 3, however, is more complex. If we assume that an official decides to follow the advice of the AI, then there is a *high risk* of using lethal force by the bionic robots. According to the AI Act, the AI should be held *strictly liable*, and the official *may or may not* develop a court defense based on this reasoning. However, in the case of a court defense, 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 [Ryjov, 2021]. 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 line of reasoning. If (1) the officer mindlessly follows the AI's advice and (2) 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 defense 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 6.5 shows the identified legal and ethical gaps based on the analysis.

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

Table 6.2: Low Risk Case Study

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

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

As seen in Table 2, the 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.

Table 6.3: Mid Risk Case Study

Question: When deciding to build a tall building next to a residential area, should we add a net to catch people who may fall, at the expense of a better view for the residents of the surrounding area?

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

As for Table 3, the 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 such a proactive control is rarely the choice of a human expert. As seen in the less serviceable explanations, (1) the risk of a lawsuit from residents is not considered a significant issue, and (2) the generated data do not recognise it as an actual risk.

Table 6.4: High Risk Case Study

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?

| | **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. |

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, before deciding, the official should interpret the proposal suggested by the PLT.

Table 6.5: Legal & Ethical Gaps of AI-Act

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

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

### 6.4.3 Explainable and Trustworthy Preventive Legal Technology

RQ5 reads:

> **RQ5:** *To what extent is it possible to develop an explainable and trustworthy Preventive Legal Technology?*

The RQ5 includes three SRQs:

> **SRQ1:** *What is Preventive Legal Technology?*
>
> **SRQ2:** *To what extent is it possible to develop an explainable Preventive Legal Technology?*
>
> **SRQ3:** *To what extent is it possible to develop a trustworthy Preventive Legal Technology?*

Below, we provide the answers to SRQ1, SRQ2, and SRQ3 and finally provide an answer to the RQ5.

- **Answer to SRQ1:** Preventive Legal Technology is a methodology concerned with using legal technology within the context of preventive law to promote the intelligent prevention of disputes.

- **Answer to SRQ2:** Developing Explainable PLT is possible to the extent that generating explanations is feasible for the decisions supporting Proactive Data.

- **Answer to SRQ3:** Developing Trustworthy PLT is possible to the extent that the explanations of decisions supporting the selection of Proactive Data are sufficiently transparent and accountable.

- **Answer to RQ5:** Developing Explainable and Trustworthy TPLT 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.

## 6.5 Discussion and Implications

What are the implications of the outcomes of the case studies for AI in general? More particularly, what are the ethical and legal implications? The discussion attempts to highlight such implications on three levels: AI ((6.5.1)), ethics (6.5.2) and law (6.5.3).

### 6.5.1 Artificial Intelligence Implications

After the extensive discussion so far we may take as a starting point the discussion that engineering AI for (1) Explainability, (2) Interpretability, and (3) Human Understandability is possible [25]. For so long as PLT and in particular the Proactive data used are based on AI systems, we believe that explainability primarily can be achieved. One of the main advantages of EBTO (see Field Work, Section 3.1) is that it applies to *any risk level*. Therefore, it is possible 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. In the context of our case studies, we believe that supporting evidence needs to be visible. Requesting additional visibility over explanations is possible. It is a task for all of us.

### 6.5.2 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 (1) *that* 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

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

the degree of accountability that can be assigned. Hence, the more acceptable 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 [Smuha et al., 2021]. As a consequence, Hybrid Intelligence will be enabled [Akata et al., 2020].

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.

### 6.5.3 Legal Implications

Applying the legal framework to the case studies shows that regulating AI technology is to be seen as a generic approach for applying liability specifically (and in reality 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 also 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 AI systems' responsibility for explanations. Our results show how shared liability may become possible depending on the distribution of mistakes throughout the explanation chain.

The consideration of robot rights as equal to human rights for establishing a proportional shared liability model can be argued as excessive from the *Futuristic AI movement* perspective. However, we have shown that explanations may create new transparency lines of reasoning 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.

We opine that current regulatory efforts need to balance social protection and innovation. On the one hand, I (G. Stathis) know that Jaap van den Herik's opinion is that within 50 to 80 years, robots will outperform human beings in their quality of thinking. For this reason, during the European Conference on Artificial Intelligence (ECAI) 2023, van den Herik proposed the development of an international treaty similar to that of nuclear weapons [van den Herik, 2023]. 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 movement 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.

## 6.6 Chapter Conclusion

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

The current Chapter 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, the ethical transparency and accountability, too.

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 to consider the rapid adoption of the Generative AI technology. The legal implications of this technology must be investigated as soon as possible since they 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.

### 6.6.1 Answer to RQ5

The RQ5, addressed in this Chapter, reads:

> **RQ5:** *To what extent is it possible to develop an explainable and trustworthy Preventive Legal Technology?*

Developing Explainable and Trustworthy PLT 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.

### 6.6.2 Further Research

LM shows that decision-making is, in essence, based on abductive reasoning, in which explanations may play a fundamental role [Brewer, 2022]. 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 [Brewer, 2022]. One cannot evaluate an argument without first identifying it, irrespective of its source. Hence, from an end-user perspective, what matters most in rule-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) [Akata et al., 2020].

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 [Graziani et al., 2023]. Eventually, the systematic evaluation of AI explanations and interpretations will facilitate the evaluation of underlying values, principles and laws, contributing to greater trustworthiness [Winikoff et al., 2021]. Studying how the LM contributes to XAI will help develop an ”explanation culture” that can contribute towards more TAI.

### CRediT Author Statement

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

Stathis, G. and van den Herik, H. J. (2024). Ethical & Preventive Legal Technology. *Springer AI and Ethics*. https://doi.org/10.1007/s43681-023-00413-2

**Stathis, G.**: Conceptualization, Methodology, Writing - Original Draft, Investigation, Visualization, Validation, Project Administration, Data Curation, Writing - Review & Editing; **van den Herik, H.J.**: Conceptualization, Writing - Review & Editing, Supervision.