Human-AI Teaming Platform for Maintaining and Evolving AI Systems in Manufacturing

D1.3 TEAMING.AI Policies

Deliverable Lead	TIM
Deliverable due date	30/06/2021
Actual submission date	30/06/2021
Version	1.0

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 957402.

Document Control Page	
Title	TEAMING.AI Policies
Editor	TIM
Contributors	SCCH, IDK, TYR, IDEA, WU
Work Package	WP1: Requirements and Prerequisites
Description	D1.3 is a report on the legal and ethical rules and principles that must be taking into account during the project implementation and the design of the solutions. It comprises ethical requirements based on fundamental rights and responsible AI design; as well as legal requirements based on the General Data Protection Regulation and the recent Proposal for a Regulation laying down harmonised rules on artificial intelligence. The report provides relevant recommendations for human centric AI application, and provides an overview of how these will be mapped into TEAMING.AI policies suitable for evaluation during use case execution, thus providing auditable compliance in an innovative manner.
Creation date	1/5/2021
Type	Report
Language	English
Audience	☒ Public ☐ Confidential
Review status	☐ Draft ☐ WP leader accepted ☒ Coordinator accepted
Action requested	☐ to be revised by Partners ☐ for approval by the WP leader ☐ for approval by the Project Coordinator ☒ for acknowledgement by Partners

Document History			
Version	Date	Author(s)	Changes
0.1	9/6/2021	Hans Graux (TIM)	Initial drafting, with general legal and ethics requirements
0.2-0.8	22/6/2021	Hans Graux (TIM)	Updates in relation to the modelling section – initial inputs

0.9	23/6/2021	Gernot Stübl (PROFACTOR), Thomas Hoch (SCCH)	Further updates in relation to the modelling section
0.10	28/06/2021	Hans Graux (TIM)	Finalisation and cleanup for submission
1.0	30/06/2021	Sabine Stockinger (SCCH)	V1.0, PDF generated and submitted to EC portal

List of contents

List of Figures

1 Abstract / Executive Summary

Within TEAMING.AI, a central concern is that all project results are designed, created and tested with European values in mind. The objective of the project is to develop human centric AI applications, in a manner that respects and empowers individuals, and provides them with appropriate safeguards against potential abuses. Moreover, TEAMING.AI aims to innovate by creating auditable compliance and auditable ethics: the legal and ethics requirements must be modelled in such a way that compliance can be verified at any time, including by third parties.

To achieve this objective, Work Package 1 of the project aims to identify the general requirements and prerequisites. Task 1.3 (Modelling of policies) of TEAMING.AI more specifically should carry out a conceptual analysis of relevant human centric AI ethical and legal issues, including those relating to autonomy, transparency, privacy, liability and so forth, so as to formulate guidance for the further project implementation.

As the task name (Modelling of policies) already indicates, the goal is not merely to list relevant requirements on the basis of existing laws and policies, but also to identify how the requirements can be formalised and represented by means of concepts from business process modelling and knowledge graphs, so that compliance can be automatically and continuously evaluated, and to ensure that there is perfect transparency at all times for users of TEAMING.AI solutions on which checks have been applied precisely, and where any potential risks may lie.

With that in mind, this D1.3 defines:

- Specific **ethics** requirements, derived through the application of the principle of Responsible Innovation, on the basis mainly of the European Charter of Fundamental Rights, the EU guidelines on ethics in artificial intelligence; and the Ethics guidelines for trustworthy AI;
- Specific **legal** requirements, derived principally from the General Data Protection Regulation and from the recent Proposal for a Regulation laying down harmonised rules on artificial intelligence, complemented by general product safety regulations;
- A general **methodology** for mapping the resulting ethics and legal requirements into verifiable **policies,** based on standardised XML formats and knowledge graphs.

While these requirements and the approach to modelling are tentative – being formulated in M6 of the project, which is relatively early – they are intended to provide a statement of fundamental requirements, and a flexible and expandable method for the application and verification of these requirements throughout the project's lifecycle and beyond. It is intended that the outputs from this deliverable will be maintained and refined throughout the project, thus contributing to continuous auditable compliance in a manner that improves upon the state of the art, and that contributes to future exploitability of the TEAMING.AI outputs.

2 Introduction

2.1 General introduction

This deliverable is created with the objective of establishing a clear baseline of legal and ethics requirements of the TEAMING.AI project within the first months of its execution, and of providing a methodology for mapping the identified requirements into verifiable policies, thus enabling both compliance and accountability.

While the project is presently still at an early stage, it is crucial to define legal and ethics requirements at the beginning of the project. In this way, all project partners have a clear and common understanding of the project's needs, and a repository of requirements that must be adhered to during the design, implementation and testing work.

The TEAMING.AI project aims to create a human AI teaming framework that integrates the strengths of the flexibility of human intelligence and the scale-up capability of machine intelligence. The general objective is to meet the increased need for flexibility in the maintenance and further evolution of AI systems, driven by the increasing personalization of products and service, as well as tackling the barriers of user acceptance and ethical challenges involved in the collaborative environments where artificial intelligence will be used, in order AI can be considered as "teammate" rather than as a threat.

These priorities are also reflected in the document, which aims to provide a methodology to implement and evaluate the legal and ethical requirements established by these frameworks.

2.2 Description of the document

This deliverable has the double objective of defining legal and ethics requirements, and of providing a general methodology for modelling and evaluation compliance with the requirements.

With respect to **ethics requirements**, the deliverable builds on the EU's framework for Responsible Research and Innovation (RRI)[1], [2]. As described by the Commission, RRI implies that a normative framework of ethics requirements is first defined, and that the designers of a system – in this case the TEAMING.AI consortium – integrate this framework into their work. In the following sections, we will describe the relevant sources, and the resulting normative framework (i.e. the ethics requirements for the project.

With respect to **legal requirements**, the main concerns are data protection (privacy protection), as well as general product safety and security, taking into account the potential impact that AI applications can have on humans. As such, legal requirements are mainly (but not exclusive) derived from the General Data Protection Regulation (GDPR, [3]), which is the general legal framework in relation to the protection of personal data and informational privacy; and from the recent Proposal for a Regulation laying down harmonised rules on artificial intelligence (AI Regulation [4]).

The outcome is a detailed (though not necessarily exhaustive) list of ethics and legal requirements. Thereafter, an explanation is provided of how the requirements can be applied in practice using **business process modelling**. Described at a high level, this implies a method to translate the requirements into an XML format, that can be integrated into knowledge graphs,

which can in turn be embedded into the general architecture of the TEAMING.AI framework. The result is an innovative, flexible and scalable approach for building auditable compliance into any type of AI application.

2.3 WP and Tasks related with the deliverable

Given the time of submission of this deliverable in month 6, it is clear that the TEAMING.AI consortium aims to continuously evaluate and iteratively develop the outputs in this report, which will be refined and tailored as the project progresses.

Several tasks and deliverables are closely linked to this deliverable:

- D10.1, submitted in M3, was the first ethics deliverable in the TEAMING.AI project. Work Package 10 (Ethics requirements) generally aims to ensure that the project is executed in accordance with the EU's high standards for ethics. D10.1 defined twelve substantive and procedural ethics checks, and provided initial drafts of ethics compliance documents. The requirements identified in D10.1 have been integrated into the legal and ethics requirements of this deliverable D1.3, so that the present report provides an extensive overview of requirements, and D10.1 provides some of the tools that can be used to resolved the requirements.
- Further implementation of ethics requirements and monitoring of compliance will be done through Task 8.5 - Legal and ethical requirements definition. This task, which runs for the full project duration, will further evaluate ethical requirements as they emerge and evolve (building of course on the work of the present deliverable), and will implement the necessary outputs, including more tailored consent forms, impact assessments, risk evaluations, and so forth, including also ethical recommendations for the use of the project's outputs beyond the project duration.

Through these interactions, comprehensive follow-up is required in a logically sequential manner:

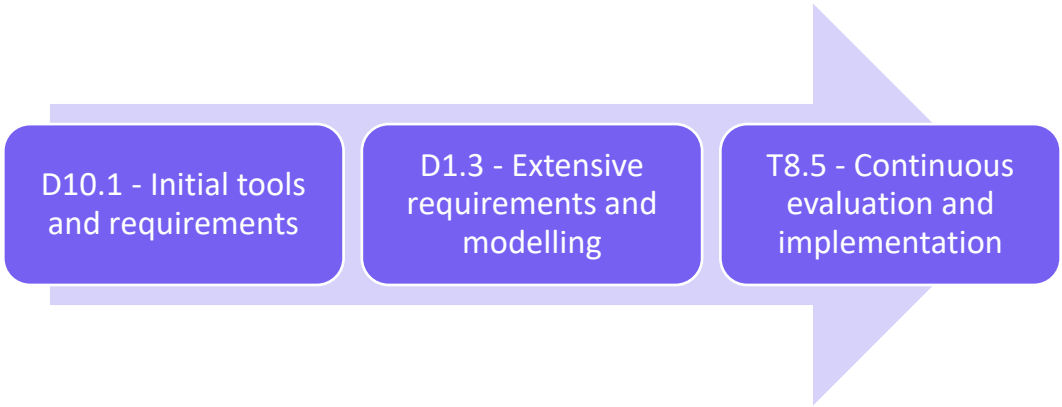

Figure 1: Legal and ethics requirements approach

As such, legal and ethical compliance will be ensured throughout the project. Moreover, due to the modelling approach and the focus on automated auditable compliance, the outputs of TEAMING.AI will be more easily usable after the project's duration.

3 Ethics requirements

3.1 Responsible Innovation as a cross cutting driver

The first pillar of this report deals with **ethics** requirements. With respect to ethics, this report applies the EU's framework for Responsible Research and Innovation (RRI)[1], [2]. As described by the Commission, RRI implies that societal actors (researchers, citizens, policy makers, business, third sector organisations, etc.) work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of society.

Ethics Requirement n°1 – Prior to deployment, **all stakeholders expected to be affected by the AI technology should be consulted**. The purpose of consultation is not to ensure that the AI functions in accordance with everyone's requirements and expectations – that would not be realistic – but rather that the concerns of the stakeholders are known and taken under consideration. In the case of TEAMING.AI, this implies the consultation of affected companies using the AI, and their employees, in order to understand their needs, expectations and concerns.

The objective of the ethics tasks in the TEAMING.AI project is to ensure that the innovation brought about by the project is in line with European ethics and moral values. This is done by applying the theory of Value Sensitive Design, an approach which aims to integrate a wide range of human and moral values into the design of (information) technology.

In other words, Value Sensitive Design implies that **a normative framework is defined**, and that the designers of a system – in this case the TEAMING.AI consortium – integrate this framework into their work, thus recognising that systems are rarely ethically neutral, and that human well-being, human dignity, justice, welfare, and human rights can be served by integrating them into technological design.

As a first step, it is important to determine the relevant sources of ethical norms. Within the EU, the European Charter of Fundamental Rights[6] provides the legal underpinning of the ethics protections for European citizens. The Charter applies a structure of six value domains:

- **Dignity**, notably individuals' right to be secure in their physical and mental integrity.
- **Freedoms**, comprising the rights to data protection and privacy, but also intellectual freedoms (education, expression, thought, religion and information) and social freedoms (assembly, marriage, asylum and property);
- **Equality**, including non-discrimination and rights of minorities and of societally more vulnerable parties;
- **Solidarity**, covering workers' rights and labour rights, social security, collective bargaining, health care and environmental protection;

- **Citizens' rights**, such as the right to vote, to proper administration, access to documents and freedom of movement;
- **Justice**, including access to fair trial and effective remedy, and the right to defence.

These fundamental rights remain relatively abstract. For that reason, two other and more specific authoritative European sources are taken as the baseline for determining ethics requirements for TEAMING AI: the EU guidelines on ethics in artificial intelligence [4]; and the Ethics guidelines for trustworthy AI [5]. Both will be discussed in greater detail below, including the resulting ethics requirements.

3.2 Ethics guidelines for trustworthy AI

The Ethics guidelines for trustworthy AI are the outcome of a series of discussions and consultations by the High-Level Expert Group on AI, an expert group that was established under the auspices of the European Commission, with the specific mandate of clarifying ethics ramifications and requirements for the use of AI In the European Union. A first draft of the Guidelines was published in December 2018; and the final result was published in April 2019.

3.2.1 Principles of the Ethics guidelines

The Guidelines build on the fundamental requirements that trustworthy AI should be:

(1) **lawful** - respecting all applicable laws and regulations

(2) **ethical** - respecting ethical principles and values

(3) **robust** - both from a technical perspective while taking into account its social environment

All three of these elements are addressed in this deliverable:

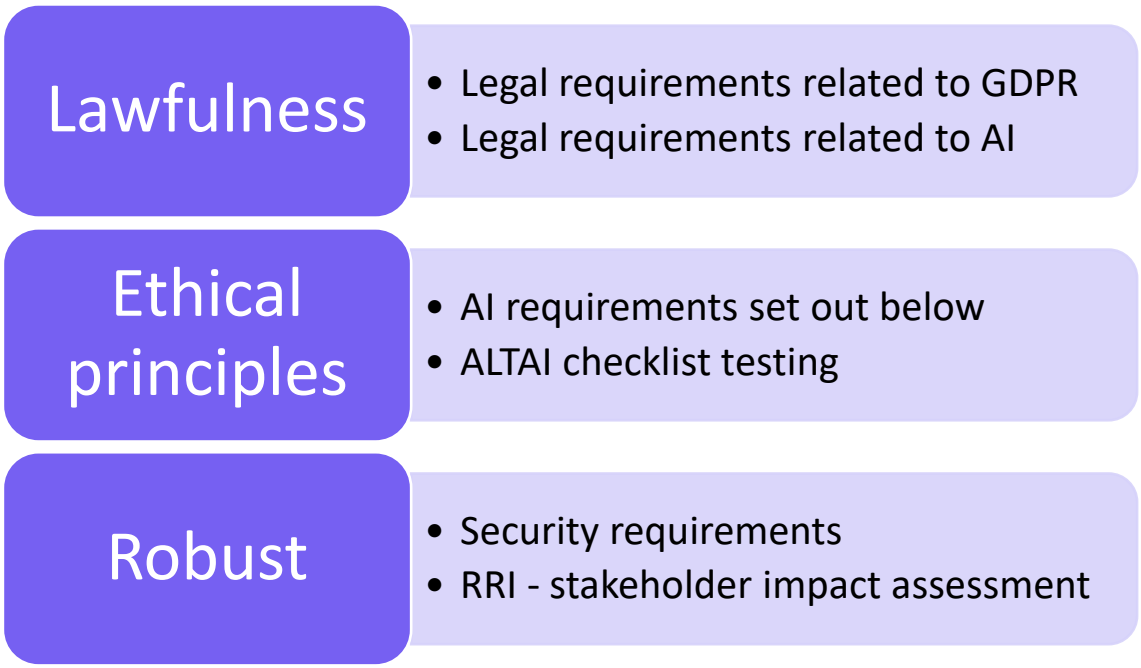

Figure 2: Mapping the Guidelines to D1.3

The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy, which are described as follows:

- ***Human agency and oversight****: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches*
- ***Technical robustness and safety****: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.*
- ***Privacy and data governance****: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.*
- ***Transparency****: the data, system and AI business models should be transparent. Traceability mechanisms can help achieving this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.*
- ***Diversity, non-discrimination and fairness****: Unfair bias must be avoided, as it could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.*

- ***Societal and environmental well-being****: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.*
- ***Accountability****: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate an accessible redress should be ensured.*

These requirements can be concretised into the following requirements for TEAMING.AI (noting that privacy, data protection requirements and accountability will be addressed in section 4 below):

Ethics Requirement n°2 – Prior to deployment and at any time during use, **a human contact person must be available**. This person must be able to explain the underlying logic and intended functioning of the AI application, and provide assistance in case of doubts on the proper functioning of the AI components. **Contact information of that person must be made available to affected persons.**

Ethics Requirement n°3 – Prior to deployment, **a safety risk management plan must be available**. This plan must outline known risk and their expected impacts, outline mitigation measures taken, and identify appropriate plans to address security problems**.** Given TEAMING.AI's objectives, no significant risks to health or wellbeing should exist.

Ethics Requirement n°4 – Prior to deployment and at any time during use, **accessible information must be available in writing, and explained in face to face contact, on the use, underlying logic and intended impact of the AI technology**. The communication should be in accessible and easy to understand language.

Ethics Requirement n°5 – Prior to deployment and at any time during use, **it should be assessed that the AI application has no negative impacts in terms of discrimination**. The assessment should be done both ex ante (prior to deployment) and ex post (afterwards, to assess whether any discrimination has unintentionally occurred. In particular, the position of less abled persons in the workplace should be considered in TEAMING.AI

3.2.2 Ethics Guidelines and self-assessment as a methodology – ALTAI testing

The High Level Expert Group recognised the challenges of applying these relatively generic requirements in a broad and fast moving field such as AI. To facilitate the application and improve user friendliness of the Guidelines, in 2020 it developed and released a self-assessment tool, the Assessment List for Trustworthy Artificial Intelligence (ALTAI)[1].

ALTAI is available both as a downloadable list[2], and as a web based tool[3]. While it contains significantly greater details on the aforementioned core ethics requirements, it contains no independent substantive requirements that are not a part of the five cross-cutting requirements above.

None the less, it will be required in TEAMING.AI to complete the ALTAI self-assessment prior to initiating the Use Cases, as a methodological tool to ensure that the aforementioned requirements are correctly interpreted:

Ethics Requirement n°6 – Prior to deployment of AI technologies in TEAMING.AI, **the ALTAI self-assessment tool should be completed**. Any action points revealed by the tool should be documented and addressed.

[1] See https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment
[2] See https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342
[3] See https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence

4 Legal requirements

Next to the ethics requirements, TEAMING.AI's solutions obviously must also be capable of satisfying the EU's regulatory requirements. These relate principally to two separate vectors: on the one hand the EU's rules on data protection and privacy (incorporated into the GDPR); and on the other hand general safety and market legislation (incorporated into product safety legislation, and envisaged to be impacted by the newly proposed AI Regulation).

Both of these vectors will be briefly analysed below, with specific legal requirements again being derived from the general rules.

4.1 Privacy and data protection – the General Data Protection Regulation

The General Data Protection Regulation[3] is the EU's principal framework for the protection of personal data, i.e. any data that can be used to identify a specific natural person. Personal data is a broad term under EU data protection law; it comprises not only directly identifiable information (such as names, addresses, contact information, video or audio recordings), but also indirectly identifiable information (such as pseudonymous information where data can only be linked to a specific semantically meaningless number). The GDPR comprises particular protections against profiling and automated decision making, which makes it particularly relevant for projects such as TEAMING.AI.

As stated in the GDPR, this implies compliance with seven key principles:

- **lawfulness, fairness and transparency** – meaning that a legal basis for any data processing (including but not limited to consent) must be available, and that the persons concerned must be appropriately informed of how their data will be used;
- **purpose limitation** – meaning that data must be collected for specific purposes, and may thereafter only be used for compatible purposes;
- **data minimisation** – meaning that data collected and used in the project must be as minimal as possible, taking into account the intended purposes;
- **accuracy** – meaning that measures must be taken to ensure the quality and accuracy of the data, and that measures must be available to detect and remedy problems;
- **storage limitation** – meaning that data may only be retained for as long as necessary given the intended purposes, and that it must thereafter be deleted or anonymised;
- **integrity and confidentiality** – meaning that data must be protected by appropriate technical and organisational measures to ensure its confidentiality, integrity and availability;
- **accountability** – meaning that responsible entities must be identified, and that appropriate controls (such as logs) are available to ensure that any problems can be attributed to the correct entity[11].

Based on these general principles, and taking into account the guidance on the application of the GDPR to AI technologies (notably the EDPS Opinion on the European Commission's White Paper on Artificial Intelligence – A European approach to excellence and trust[12]), a series of legal requirements can be derived to ensure that TEAMING.AI follows the objectives of 'data protection by design' and 'data protection by default':

Legal Requirement n°1 – Any processing of personal data must have a **legal basis** recognised under the GDPR. If personal data is to be processed in the context of an employment relationship, the legal basis should not be consent, since it cannot be ensured that consent is lawfully given (and therefore it would be invalid).

Legal Requirement n°2 – Any processing of personal data must be clearly **disclosed** to the affected persons in accordance with the requirements of the GDPR. This applies both to the collection of data, and to the application of AI analytics. Given the likelihood of **profiling and automated decision making,** this includes disclosure of the underlying reasoning.

Legal Requirement n°3 – Data collection for the purposes of AI analytics must be **limited** to what is strictly necessary for piloting purposes.

Legal Requirement n°4 – If personal data is collected specifically in the context of the pilots, the data may only be used for piloting purposes. Personal data may not be used by third party applications without the user's consent.

Legal Requirement n°5 – Users must be able to **access, correct (if applicable) or delete** their data. They must be able to **withdraw** from the pilot at all times, and this right should be communicated to them.

Legal Requirement n°6 – Whenever AI services are used in a pilot, the processes and outcomes must be **logged and monitored**, at a minimum by the pilot service provider, in order to proactively detect any problems that may occur, and to avoid any adverse effects on the user.

Legal Requirement n°7 – Personal data collected for the purposes of TEAMING.AI must automatically be **deleted or anonymised** at the end of the TEAMING.AI project, except where retention is necessary to show compliance with legal requirements. .

Legal Requirement n°8 – Personal data processed withing TEAMING.AI infrastructure must be protected with appropriate **access controls or effective encryption** in order to protect the data against unlawful access. Any personal data sent between TEAMING.AI components through a network must be protected against unlawful interception through effective encryption.

Legal Requirement n°9 – Any personal data processing in the context of pilots must be supervised by a duly qualified data protection officer (**DPO**) meeting the requirements of the GDPR. The contact information of the DPO must be made available to the user of any pilots.

Legal Requirement n°10 – Any personal data processing in the context of pilots must be preceded by a data protection impact assessment (**DPIA**) created in the context of TEAMING.AI. Any piloting constraints (other than those referenced in this deliverable) must be disclosed in the DPIA and adhered to..

Legal Requirement n°11 – Any **personal data sent to a third country** using TEAMING.AI components or services must satisfy the transfer requirements from the GDPR. Given the piloting objectives, an explicit consent MAY be used as the legal basis for third country transfers.

Legal Requirement n°12 – Personal data processing in the context of pilots must **not relate to minors, or to persons who are legally impaired, nor may it comprise special categories of data** (notably data concerning health), unless this has been verified and approved by a competent DPO.

4.2 Product safety and bringing an AI product to the market – the proposed AI Regulation

Beyond data protection, a cross cutting legal question is also whether an AI product or service can be safely brough to the market, i.e. which responsibilities are incumbent on the developers of the AI logic itself, and what the responsibilities are of the manufacturers of products and services incorporating the AI itself.

This has traditionally been a complex question. A 2020 study from the European on Artificial Intelligence and Civil Liability [13] observed that the EU has robust product safety legislation in place through the Product Liability Directive 85/374/EEC[14] and related frameworks; but that the application of this framework in an AI context is challenging due to its focus on "products", and the ambiguity as to whether AIs qualify as products under European law.

In order to alleviate this risk to some extent, the European Commission published a new Proposal for a Regulation laying down harmonised rules on artificial intelligence[10] (the AI Regulation). The proposed Regulation puts forward rules to enhance transparency and minimise the risks to safety and fundamental rights. These rules must be applied before AI systems can be brought to the European market.

The Regulation generally focuses on so–called 'high-risk' AI use cases, i.e. where the risks that the AI systems pose are particularly high, as determined by specific criteria included in an Annex of the Regulation. Whether an AI system is classified as high-risk depends on its intended purpose of the system and on the severity of the possible harm and the probability of its occurrence. High-risk systems include:

1. ***Biometric identification and categorisation of natural persons****: (a) AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons;*
2. ***Management and operation of critical infrastructure****: (a) AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.*
3. ***Education and vocational training****: (a) AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions; (b) AI systems intended to be used for the purpose of assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions.*
4. ***Employment, workers management and access to self-employment****: (a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests; (b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behaviour of persons in such relationships.*
5. ***Access to and enjoyment of essential private services and public services and benefits****: (a) AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services; (b)*

AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use; (c) AI systems intended to be used to dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid.

6. ***Law enforcement**: (a) AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences; (b) AI systems intended to be used by law enforcement authorities as polygraphs and similar tools or to detect the emotional state of a natural person;(c) AI systems intended to be used by law enforcement authorities to detect deep fakes as referred to in article 52(3); (d) AI systems intended to be used by law enforcement authorities for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences; (e) AI systems intended to be used by law enforcement authorities for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups; (f) AI systems intended to be used by law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences; (g) AI systems intended to be used for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data.*
7. ***Migration, asylum and border control management**: (a) AI systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person; (b) AI systems intended to be used by competent public authorities to assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State; (c) AI systems intended to be used by competent public authorities for the verification of the authenticity of travel documents and supporting documentation of natural persons and detect non-authentic documents by checking their security features; (d) AI systems intended to assist competent public authorities for the examination of applications for asylum, visa and residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status.*
8. ***Administration of justice and democratic processes**: (a) AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.*

TEAMING.AI could be qualified as a high-risk initiative, due to its application in the employment relationship (point 4 in the list above), and notably "*AI intended to be used for […] task allocation and for monitoring and evaluating performance and behaviour of persons in such relationships.*

This does not imply that TEAMING.AI outcomes would be unlawful, of course. However, the proposal does provide that high-risk AI systems need to respect a set of specifically designed requirements, which include the use of high-quality datasets, the establishment of appropriate documentation to enhance traceability, the sharing of adequate information with the user, the design and implementation of appropriate human oversight measures, and the achievement of the highest standards in terms of robustness, safety, cybersecurity and accuracy.

High-risk AI systems must be assessed for conformity with these requirements before being placed on the market or put into service. To smoothly integrate with existing legal frameworks the proposal takes account, where relevant, of the sectorial rules for safety, ensuring coherence between the legal acts and simplification for economic operators.

Finally, the proposed Regulation lays down a ban on a limited set of uses of AI that contravene European Union values or violate fundamental rights. The prohibition covers AI systems that distort a person's behaviour through subliminal techniques or by exploiting specific vulnerabilities in ways that cause or are likely to cause physical or psychological harm. It also covers general purpose social scoring of AI systems by public authorities. None of these use cases apply to EU, obviously.

For the specific case of remote biometric identification systems (e.g. facial recognition tools to check passers-by in public spaces), the proposed regulation establishes a stricter approach. The real-time use for law enforcement purposes would in principle be prohibited in publicly accessible spaces, unless when exceptionally authorised by law29. Any authorisation is subject to specific safeguards. In addition, all AI systems intended to be used for remote biometric identification of natural persons must undergo an ex ante conformity assessment procedure by a notified body to check compliance with the requirements for high-risk AI systems, and will be subject to stricter logging and human oversight requirements.

Other (non-high risk) uses of AI systems are only subject to minimal transparency requirements, for example in the case of chatbots, emotion recognition systems or deep fakes.

Compiling these requirements, the following additional legal requirements in terms of authorization, safety and product quality standards should be applied in TEAMING.AI, building on the assumption that TEAMING.AI should be designed to be able to deal with the implications of a classification as a potentially high risk use case under the AI Regulation:

Legal Requirement n°13 – A **risk management system** must be established, implemented, documented and maintained in relation to TEAMING.AI's systems, consisting of a continuous iterative process run throughout the entire lifecycle of TEAMING.AI's systems, requiring regular systematic updating.

Legal Requirement n°14 –TEAMING.AI's systems must be **tested** for the purposes of identifying the most appropriate risk management measures. Testing shall ensure that high-risk AI systems perform consistently for their intended purpose.

Legal Requirement n°15 –TEAMING.AI's systems must be **training with data sets covered by appropriate data governance and management practices**. Those practices shall concern in particular (a) the relevant design choices; (b) data collection; (c) relevant data preparation processing operations, such as annotation, labelling, cleaning, enrichment and aggregation; (d) the formulation of relevant assumptions, notably with respect to the information that the data are supposed to measure and represent; (e) a prior assessment of the availability, quantity and suitability of the data sets that are needed; (f) examination in view of possible biases; (g) the identification of any possible data gaps or shortcomings.

Legal Requirement n°16 – (likely out of scope) **technical documentation** (complying with Annex IV of the AI Regulation) must be drawn up before that system is placed on the market or put into service and shall be kept up-to date.

Legal Requirement n°17 – **automatic recording of events** ('logs') must be enabled at all times while TEAMING AI's systems are operating.

Legal Requirement n°18 – (likely out of scope) **a conformity declaration** (complying with Annex V of the AI Regulation) must be drawn up before that system is placed on the market or put into service and shall be kept up-to date, and **CE marking must be applied**

Legal Requirement n°19 – (likely out of scope) **the system must be registered in the EU database of high risk applications** before that system is placed on the market or put into service and shall be kept up-to date.

5 Casting requirements into verifiable policies

5.1 General structure of requirements

The sections above have provided an overview of the principal legal and ethics requirements for TEAMING.AI. These can be graphically summarised as follows:

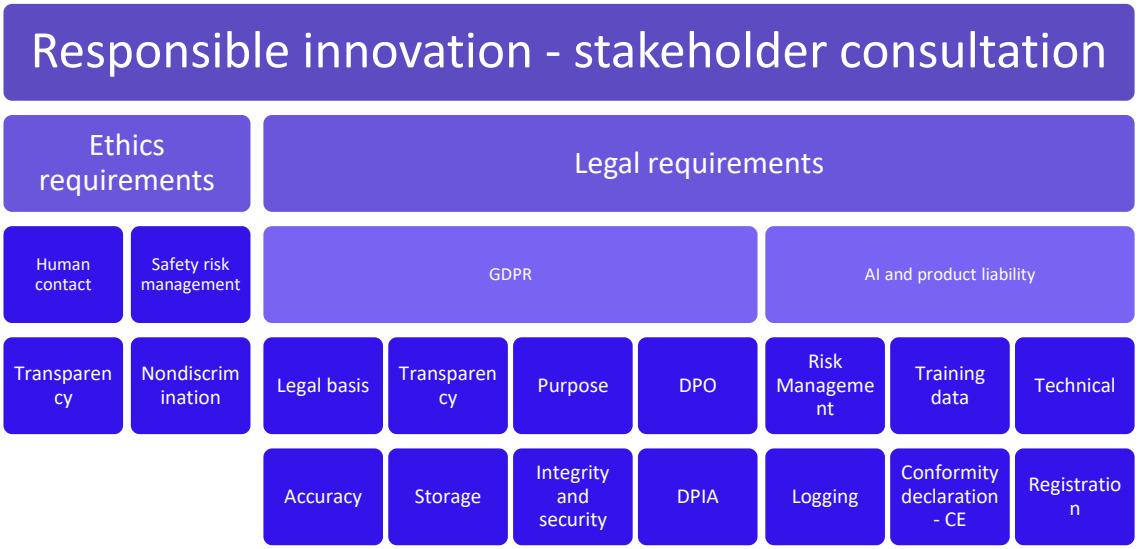

Figure 3: Overview of legal ethics requirements

The majority of these requirements are suitable for modelling and for automated verification, since it is possible to indicate simply whether the requirement has been satisfied, and to link to an appropriate resource to corroborate claimed compliance. In other words, the template above could be recast into a simple legal knowledge graph, describing a specific TEAMING.AI use case in a structured manner in terms of its legal and ethics compliance, using a series of predefined concepts. In this way, the requirements of this deliverable could be made suitable for automated evaluation and assessment.

5.2 From requirements to policies: general approach

The legal and ethics requirements summarily stated above represent a typical human readable rendition of the main compliance concepts. One of the objectives of TEAMING.AI is however to advance the state of the art by providing auditable compliance. Specifically, the requirements should be rendered and integrated into the TEAMING.AI architectural framework in a way that allows compliance to be continuously and automatically verifiable.

In practical terms, this would ideally mean that a user of the TEAMING.AI framework – such as e.g. a company using a TEAMING.AI driven AI application, or independent verifiers such as auditors or even labour unions – would be able at any time to assess precisely which controls have been applied to the application, and precisely in which way the requirements have been satisfied. A trivial visual rendition of auditable compliance would then be a table that would be structured as follows:

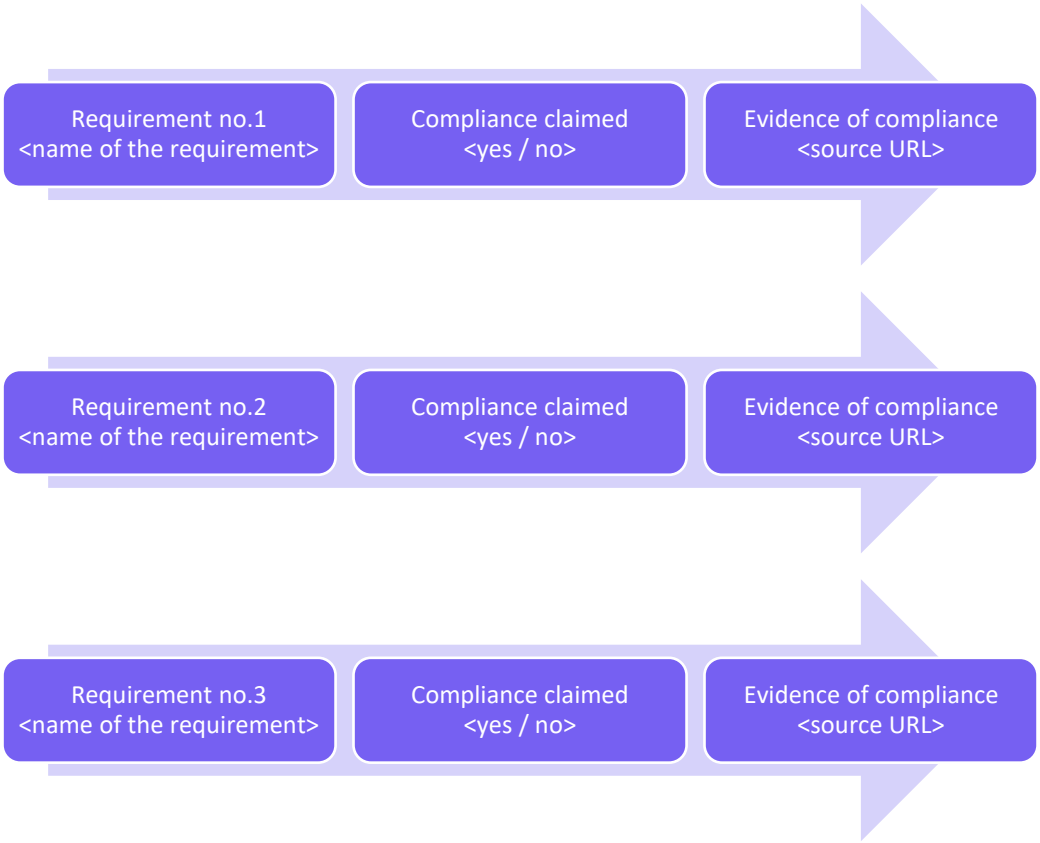

Figure 4: Modelling approach

Thus, every requirement has a unique name, corresponding to the blue boxes of Figure 3. For every requirement, it is indicated whether the application claims to be compliant with the requirement. Noncompliance is not necessarily problematic, since compliance may not be mandatory – e.g. not every AI application would require CE marking, or require the supervision of a data protection officer. Since compliance claims are binary (compliance is claimed, or it is not claimed), this can be verified automatically.

If compliance is claimed, then a URL is included to a resource that shows how the requirement is complied with in practice. The URL can point to a fully structured resource (e.g. a standardised

signed XML file), or to an unstructured resource (such as a web page containing a description of a transparency policy), depending on the requirement and on whether a standardised proof of compliance exists.

Using this approach, any interested third party could verify quickly and automatically whether the requirements that are integrated into this deliverable are claimed to be complied with. If they want further corroboration, the evidence can be obtained as well. Since the evidence is not necessarily structured or standardised (it may be prose text), it is also not necessarily automatically verifiable. Human interpretation will then be required in all situations where a relying party is not able or willing to rely on the compliance statement that can be integrated into a TEAMING.AI application.

However, this state of play can evolve over time: when new evidences are standardised, then this model can integrate those standards easily. For instance, standards on the quality of training data are presently unavailable, so that evidence of compliance with this requirement cannot be rendered in an automatically verifiable way – policies describing the measures taken to ensure the quality of training data will need to be written out in a human readable format. If, however, in the future a standard is adopted for this requirement, the evidence could be rendered in the form of a compliance certificate, in which case not just compliance, but also proof of compliance becomes automatically auditable.

5.3 Casting legal and ethics requirements into a knowledge graph

Continuing this approach, the essential method for modelling compliance in a verifiable way consists of rendering the aforementioned compliance policy as part of the Teaming.AI knowledge graph (KG). Every requirement is modelled as a node in the KG that has links to a set of claims (usually questions) that the application needs to comply with. Every claim itself is a node in the KG that describes the contextual information of the claim (e.g. whether compliance is fulfilled, at what date, based on which version of the data, ...) and also provides a URL to the evidence of the claimed compliance.

In this manner, an AI application can indicate which requirements it claims to satisfy and how. Where evidences are structured and standardised – which will be rarely the case initially – validation can also be standardised. E.g. if a DPO can be identified on the basis of a standardised certification (which is presently not the case), then the evidence can link to that certification, thus providing automated validation of the evidence as well. Or in the case of model quality, the evidence can be automatically generated from the training protocol and linked to the claim with the relevant context information.

For the avoidance of doubt: this approach always enables auditable compliance, since the approach always allows requirements to be identified along with the relevant evidences. The audit is however not fully automatic, since (in the current state of play) human intervention for the interpretation of evidences is normally required. The principal goal of the ethics framework in Teaming.AI is to integrate these necessary human interventions seamlessly into the teaming workflow such that we can assure that every change in the system complies with the requirements. We want to achieve this

- through carefully selection of claims and requirements that need to be evaluated after specific system changes (e.g. after an application of a new prediction model),
- by integrating the assessment into the human-machine interface (HMI) such that the evidences can be collected thoroughly (e.g. with the relevant context),
- by automatising the generation of evidence if possible and to display the collected information in structured way such that the cognitive effort of the human intervention can be minimized,
- and by training all the team members to familiarize them with the concept of trustworthy AI and to encourage thoughtful reflection of the data usage within Teaming.AI.

6 Conclusions

As this deliverable shows, it is possible to:

- Model legal and ethics requirements for AI applications in a standardised and structured way. The project team recognises that the requirements can be made more or less fine grained, depending on the needs of a use case. The current approach represents a reasonable balance between detail and feasibility in the context of the TEAMING.AI project.
- Render the legal and ethics requirements in a verifiable way, by defining requirement names, compliance claims, and applicable evidences. This approach allows a relying party to verify compliance claims, and to define its policy requirements in a way that allows automated verification of whether a TEAMING.AI application respects its policies.
- Translate the legal and ethics requirements into knowledge graphs, thus allowing flexible, scalable and automated verification of compliance.

This approach advances the state of the art, and can provide an important building block for creating auditable compliance in any AI driven initiative that aims to comply with European legal and ethical norms.

7 Bibliography

[1] Responsible Research and Innovation Guidelines; see https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

[2] Value Sensitive Design and Information Systems, BATYA FRIEDMAN, PETER H. KAHN, JR., AND ALAN BORNING; see https://vsdesign.org/publications/pdf/non-scan-vsd-and-information-systems.pdf

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance); see https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

[4] 2019 Ethics guidelines for trustworthy AI from the European Commission's High-Level Expert Group on AI; see https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[5] 2020 European framework on ethical aspects of artificial intelligence, robotics and related technologies, Study by the European Parliament; see https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU%282020%29654179

[6] EU Charter of Fundamental Rights, see https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_en

[7] Guidelines on Data Protection Officers ('DPOs') (wp243rev.01) from the Article 29 Working Party, as endorsed by the European Data Protection Board, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048

[8] Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679 https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

[9] CNIL Privacy Impact Assessment (PIA) tools; see https://www.cnil.fr/en/privacy-impact-assessment-pia

[10] 2021 Proposal for a Regulation laying down harmonised rules on artificial intelligence; see https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

[11] 2020 Global Privacy Assembly Adopted Resolution On Accountability In The Development And Use Of Artificial Intelligence; see https://edps.europa.eu/sites/default/files/publication/final_gpa_resolution_on_accountability_in_the_development_and_use_of_ai_en.pdf

[12] EDPS Opinion on the European Commission's White Paper on Artificial Intelligence – A European approach to excellence and trust; see https://edps.europa.eu/sites/default/files/publication/20-06-19_opinion_ai_white_paper_en.pdf

[13] 2020 Study from the European Parliament - JURI Committee on Artificial Intelligence and Civil Liability; see https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU%282020%29621926_EN.pdf

[14] Product Liability Directive 85/374/EEC; see https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3Al32012

8 Annex I - Ethics guidelines for trustworthy AI

According to the Guidelines, trustworthy AI should be:

(1) lawful - respecting all applicable laws and regulations

(2) ethical - respecting ethical principles and values

(3) robust - both from a technical perspective while taking into account its social environment

The Guidelines put forward a set of 7 key requirements that AI systems should meet in order to be deemed trustworthy. A specific assessment list aims to help verify the application of each of the key requirements:

- Human agency and oversight: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches
- Technical Robustness and safety: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.
- Privacy and data governance: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.
- Transparency: the data, system and AI business models should be transparent. Traceability mechanisms can help achieving this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.
- Diversity, non-discrimination and fairness: Unfair bias must be avoided, as it could could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.
- Societal and environmental well-being: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.
- Accountability: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate an accessible redress should be ensured.