Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Semantic Frameworks to Support the EU AI Act's Risk Management and Documentation

Seyedeh Delaram S Golpayegani

Supervised by
Prof. David Lewis

Co-supervised by
Prof. Harshvardhan J. Pandit
Prof. Declan O'Sullivan

A thesis submitted to
Trinity College Dublin, the University of Dublin
in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Computer Science and Statistics

2025

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Singed: *Delaram Golpayegani*

Semantic Frameworks to Support the EU AI Act's Risk Management and Documentation

Delaram Golpayegani

Abstract

The European Union (EU) Artificial Intelligence Act (AI Act), which entered into force on 1 August 2024, stands as a landmark legal regime for development and use of AI, adopting a risk-based approach to govern the potential risks of AI to key areas of concern, including health, safety, and fundamental rights. Under the AI Act, AI systems are subject to a set of regulatory obligations according to the level of risk they pose. Within this risk-based classification, high-risk AI systems need to comply with more rigorous provisions of the Act, which should be addressed by AI providers and deployers.

Translating the AI Act's legal provisions into practical approaches and technical measures for implementing the *essential requirements* needs a range of guidelines, many of which need to be acquired from evidence-based regulatory learning. With the recent enforcement of the Act, such regulatory insights are not yet established, which has created legal uncertainty in regard to compliance with the AI Act. In this context, Regulatory Technology (RegTech) can serve as an enabling force to support the effective implementation and enforcement of the Act, while enhancing legal certainty through regulatory learning.

Focusing on risk management as a central element of the AI Act, this thesis addresses the current lack of RegTech solutions by proposing a compendium of Semantic Web-based artefacts to facilitate compliance with the requirements of the AI Act regarding risk management, risk documentation, and registration of AI systems in a Findable, Accessible, Interoperable, and Reusable (FAIR) manner. To achieve this, specific requirements of the AI Act, related to risk management, documentation, and registration, are analysed. In the current absence of authoritative guidelines and harmonised European standards to guide compliance with the Act, this analysis utilises existing ISO/IEC standards on AI, which are strong candidates for harmonisation and can therefore potentially support the implementation of the AI Act.

As a major contribution, this work proposes a novel compendium of artefacts based on Semantic Web technologies to assist with AI Act compliance

tasks. This compendium is centred around the *AI Risk Ontology (AIRO)*, a foundational ontology for modelling AI risks, and its specialisation the *AI Risk Vocabulary (VAIR)*, which is a taxonomy of concepts provided in AIRO to enable its use in practical applications. Using these two ontologies, this thesis illustrates how open, transparent, traceable, comparable, and interoperable information models of AI use cases, that include information about the system, context of use, and risks, can be created. To further demonstrate the functionality of AIRO and VAIR, this thesis leverages the capabilities offered by the Semantic Web technology stack in rule-checking, querying, expressing policies, and cataloguing information to assist with AI Act compliance tasks in regard to risk management, documentation, and registration.

AIRO and VAIR are novel AI risk ontologies developed based on the AI Act that are explicitly aligned with relevant ISO/IEC standards in anticipation of harmonised standards for the Act. This thesis also implements the first set of open, standardised, and extensible artefacts for determining high-risk AI systems, generating AI and risk documentation, expressing AI use policies, and cataloguing AI systems as required by the AI Act. Complementing this contribution, this thesis introduces *AI Cards* as a documentation framework that provides a holistic view of an AI use case and its risks in both human- and machine-readable formats, aligned with the AI Act, to facilitate communication and sharing of key AI and risk information among various AI stakeholders.

The contributions of this thesis support development of standards-based automated tools to address AI risk management and documentation challenges, particularly those related to compliance with the AI Act. This is especially important for providers and deployers of AI systems in maintaining and sharing AI and risk information in a manageable, transparent, interoperable, and verifiable manner. In addition, this standards-based automation enables the tracking and verification of claims regarding risk management and thereby facilitates conformity assessment tasks for authorities, particularly conformity assessment bodies.

ACKNOWLEDGEMENTS

Completing this work would not have been possible without the support, encouragement, guidance, and friendship of many individuals.

First and foremost, thank you to *Dave Lewis*, my primary supervisor, for the excellent supervision that shaped the perspective of mine as an independent researcher that I am today. Making it this far would not have been possible without the guidance, support, and encouragement I have received from you.

Thank you to *Harshvardhan J. Pandit* for the meticulous reviews, honest feedback, and for supporting me all the way through.

And thank you to *Declan O'Sullivan* for his thoughtful insights on my work, his timely advice, and for being there whenever I needed support.

My research was made possible by the funding of the European Commission, through the PROTECT ITN. Thank you to PROTECT managers: *Jessica Grene* and *Valerie De Moor* for their generous assistance. I would like to thank all PROTECT ESRs, and in particular WP3 members *Rana Saniei*, *Joshua Hovsha*, *Leon Rossmaier*, and *Jana Mišić* for the great teamwork that led to development of the Health AI Risk Taxonomy (HART). I would also like to express my gratitude to *Beatriz Esteves* for the collaboration and support in development of the AI use policy (AIUP) profile.

As part of PROTECT, I was fortunate to undertake three on-site secondments, all of which impacted my work. I wish to express my sincere gratitude to those who made these visits possible.

Thanks to *Víctor Rodríguez Doncel* for his assistance in making my visit to the Ontology Engineering Group (OEG) at Universidad Politécnica de Madrid (UPM) possible and for his guidance on building the tool for determining high-risk AI.

Thanks to *Markus Helfert* for hosting me at Maynooth University and

thanks to *Aphra Kerr* for the insightful discussions that sparked the initial idea of the AI Cards.

Thanks to *Sven Schade* for his guidance and support since we met at SEMIC 2022, especially during my visit to the Joint Research Centre in Ispra. I owe a big thank you to *Isabelle Hupont* and *Cecilia Panigutti* for the fruitful collaboration on development of the AI Cards. Thank you to the JRC Unit T researchers, particularly *Luca Tangi* for the helpful guidance on conducting surveys and *Giovanni Zenga* for all the inspiring conversations.

My PhD was hosted by the *ADAPT Centre*. I am grateful to my colleagues and friends at ADAPT Centre, particularly to *Julio Hernandez* for proofreading some sections of this thesis. I would also like to thank the ADAPT management, commercial, and operation teams for their support and assistance in promoting my research in multiple events and platforms.

Thanks to the anonymous survey participants and reviewers of the papers I submitted to various venues. The invaluable feedback I have received greatly impacted this work.

Thanks to my friends and colleagues in Dublin, Varese, Ispra, Madrid, and Tehran.

Last, but definitely not least, thanks to My parents—*Maman* and *Baba*, for their unwavering love and support.

Thanks to my brother, *Mohammad*, for his creative way of uplifting my mood through sharing tasteful music.

Thanks to my sister, *Reyhaneh*, for her friendship, for all the late-night video calls that we wove everything to anything from law to AI to literature to politics, and for proofreading this thesis.

Funding Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813497 (PROTECT ITN), as part of the ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme through Grant#13/RC/2106_P2.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF LISTINGS

LIST OF ACRONYMS

AI	Artificial Intelligence
AI Act	Artificial Intelligence Act
AICat	AI Catalogue vocabulary
AIRO	AI Risk Ontology
AIUP	AI Use Policy (profile)
CEN	European Committee for Standardisation
CENELEC	European Committee for Electrotechnical Standardisation
DCAT	Data Catalog Vocabulary
DIS	Draft International Standard
DPIA	Data Protection Impact Assessment
DPV	Data Privacy Vocabulary
DQV	Data Quality Vocabulary
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
FRIA	Fundamental Rights Impact Assessment
GDPR	General Data Protection Regulation
IEC	International Electrotechnical Commission
IEEE	The Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
JRC	Joint Research Centre
JTC	Joint Technical Committee
LLM	Large Language Model
LOT	Linked Open Terms
MDR	Medical Devices Regulation
ML	Machine Learning

NGO	Non-Governmental Organisation
NIST	National Institute of Standards and Technology
NLF	New Legislative Framework
NSAI	National Standards Authority of Ireland
ODRL	Open Digital Rights Language
OWL	Web Ontology Language
RDF	Resource Description Framework
RegTech	Regulatory Technology
SC	Subcommittee
SHACL	Shapes Constraint Language
SME	Small and Medium-sized Enterprise
SOTA	State of the Art
SPARQL	SPARQL Query Language for RDF
TC	Technical Committee
VAIR	Vocabulary of AI Risks
W3C	World Wide Web Consortium

CHAPTER

ONE

INTRODUCTION

1.1 Background and Motivation

This era could be marked by the unprecedented advancement in AI capabilities and its seamless integration into personal, political, societal, and environmental spheres that has led to profound changes in daily life, industries, and societies. However, a serious challenge that can overturn this prediction is the dark side of AI, which is unveiled by the, still ever-growing, corpus of evidence on its wide-ranging harms [1, 2], from harms on physical safety to psychological health to human rights and freedoms. It has become evident that use of AI is not ethically-, socially-, and politically-acceptable without risk management [3]. This fact has prompted a global wave of efforts in development of guidelines, policies, standards, and—above all—regulations to manage the potential harms of AI and in turn ensure its trustworthy development and use [4].

The European Union (EU) Artificial Intelligence Act [5] (hereafter the AI Act) is the first horizontal regulatory framework on AI, which was published as an EU law in June 2024 after a three-year-long legislative process (refer to Subsection 1.2.1 for the details of the mandates published over the course of this process). The AI Act takes a risk-based approach, through which AI systems are subjected to a set of regulatory obligations according to level of risk they impose to three key areas of concern: health, safety, and fundamental rights.

The AI Act is criticised as being "lengthy and sometimes opaque" [6], with 180 recitals, 113 Articles, and 13 Annexes. In addition, with multiple references made to other EU regulations and directives, including the 20

Union harmonisation legislation listed in Annex I[1], it requires coordination with other legal acts [7]. As a legal document, it is written in a high-level manner, requiring further interpretations of the requirements [8]. The AI Act follows the New Legislative Framework (NLF) structure [9]—the common EU product-related legal framework adopted in 2008. Therefore, it only defines the *essential requirements* whose interpretation and implementation is expected to be supported by *harmonised standards* (Article 40) [10, 11]—"European standard[s] adopted on the basis of a request made by the Commission for the application of Union harmonisation legislation"[12]. In the implementing decision on a standardisation request [13], the Commission has called upon CEN (European Committee for Standardisation) and CENELEC (European Committee for Electrotechnical Standardisation) to develop the required harmonised standards in relation to the AI Act. With a deadline in April 2025, CEN and CENELEC are delegated to create European standard(s) and/or European standardisation deliverable(s) in 10 areas, including "risk management systems for AI systems". At the time of writing this thesis, development of European standards to address the Commission request is a high priority for CEN-CENELEC Joint Technical Committee (JTC) 21 on AI[2], however these standards are not yet harmonised (see Section 1.2 for the list of relevant standards from ISO/IEC JTC 1 Subcommittee (SC) 42 that are used in this thesis).

Within the risk-based classification for AI systems, the AI Act explicitly identifies two categories: *prohibited AI practices* and *high-risk AI system*. While prohibited systems are prohibited to be placed on the market, put into service, or used within the EU, the high-risk AI systems are allowed as long as they comply with a set of requirements outlined in Chapter III, Section 2 of the Act, including requirements on risk management and documentation. In addition, the Act lays down classification rules for *general purpose AI models* on the basis of their *systemic risk at Union level*, which is not covered within the scope of this thesis (see Subsection 1.2.1 for the justification).

The AI Act applies the high-risk category to AI systems used as products and safety components of products already covered by EU harmonisation legislation (listed in Annex I). Further, it defines specific uses of AI as being high-risk, with a list provided in Annex III, with provisions for the European Commission to modify the list in future amendments.

Within the AI Act, *AI provider* is defined as an entity that develops or

[1]To avoid any confusions, hereafter *Annex* refers to Annex of the AI Act, and *Appendix* refers to the additional material provided in the appendix of this thesis.

[2]Disclaimer: The author has been a member of CEN-CLC/JTC 21 since its establishment in 2021 through membership in the National Standardisation Authority of Ireland (NSAI).

that has an AI system developed and places it on the market or puts the AI system into service under its own name or trademark (Article 3(3)). Although it is not mentioned explicitly, the AI Act (particularly Article 6(4)) implies that the assessment of whether an AI system is prohibited or high-risk should be performed by providers. Moreover, the burden of fulfilling the requirements of high-risk AI systems, in particular implementing risk management systems and documenting the risks, is mainly on the shoulders of the providers of those systems (Article 16). Thus, this thesis mainly addresses the providers' obligations. However, it considers providers' interactions with other relevant actors across the AI value chain[3], including authorities and AI deployers—entities under whose authority an AI system is used (Article 3(4)).

With risk being the yardstick for determining the regulatory requirements that AI systems need to satisfy, it is no surprise that there is a particular emphasis on management of potential harms of high-risk AI systems. Accordingly, *risk management system* provisions (Article 9) play a pivotal role in the implementation of the Act. Compliance with these provisions entails maintaining, querying, and sharing information about the AI system and its associated risks. Documentation of this information is particularly important, given that *technical documentation* is the basis upon which conformity with the AI Act is assessed (Article 11). In connection with sharing information with authorities, information about any high-risk AI system that falls under Annex III should also be registered into the *EU database* (Article 49). Within the obligations for high-risk AI systems, and in general in the AI Act, there is a strong emphasis on *intended purpose* of the system, defined as "use for which an AI system is intended by the provider, including the specific context and conditions of use" (Article 3(12)). Communication of intended purpose is another key point of information sharing with authorities as well as deployers. Against this background, it is argued that while all the requirements of the Act are equally important, risk management system, along with technical documentation and registration, are fundamental to successful compliance with the AI Act, and therefore the primary focus of this thesis is on these provisions.

Given that conformity assessment is based on auditing of information—that is provided within documents—maintaining, querying, and sharing AI and risk information are extremely important and challenging. This is due to the large extent of this information, the rapid pace of changes in AI systems and their incorporating components that should be reflected in documen-

[3]AI value chain refers to the range of activities or actors that create or receive value throughout the AI lifecycle.

tation [14], as well as the complexities of the AI value chain. Although compliance with the AI Act can be facilitated by conformity to harmonised standards, adhering to requirements of multiple standards is still resource-intensive [15]. From the enforcement perspective, in cases where third-party conformity assessment is required, auditing of this substantial information by *conformity assessment bodies* is not a trivial task.

Complexities of implementation of the AI Act, as well as the cost, time, and human resources it requires, necessitates adoption of Regulatory technology (RegTech) solutions to alleviate compliance tasks [16, 14]. Regarding the desired characteristics of the RegTech solutions, the AI Act gives hints about machine-readability by requiring the information registered into the EU database of high-risk AI systems to be represented in a machine-readable format (Article 71(4)). Additionally, the *AI Office*, which is "the Commission's function of contributing to the implementation, monitoring and supervision of AI systems and general-purpose AI models, and AI governance" (AI Act, Article 3(47)), is obligated to develop an *automated tool* to assist with Fundamental Rights Impact Assessment (FRIA)—an obligation closely related to risk management with a focus on the harmful impacts of AI on fundamental rights (Article 27). Moreover, the mechanism to amend the AI Act, in particular its Annexes, through delegated acts (Article 97) in order to keep it up-to-date with the AI advancements and emergence of new requirements introduced by forthcoming additional authoritative guidelines, e.g. guidelines on AI procurement or the AI Office-issued documents, imply the need for flexible solutions. These entail adoption of FAIR (Findable, Accessible, Interoperable, and Reusable) principles [17] in provision of RegTech solutions for compliance with the AI Act.

Linked Data principles and practices [18] have been widely adopted in publication, integration, and exchange of structured data by governmental institutions, including the European Union [19]. With the *Semantic Web technology* stack providing open and standardised approaches for information retrieval, consistency checking, validating information against a set of rules, expressing policies, and providing catalogues of data, a potential solution to address the needs of the AI Act for open, standardised, and machine-readable solutions is through the use of Linked Data and Semantic Web technologies. Furthermore, Semantic Web has proven to be an effective technology in supporting implementation of open, machine-readable, interoperable, scalable, flexible, and automated RegTech solutions [20], in particular for compliance with the EU digital regulations, including the General Data Protection Regulation (GDPR) [21] (vide [22, 23, 24]) and the Data Governance Act (DGA) [25] (vide [26]), to name a few.

In light of these, this thesis investigates the use of Semantic Web technolo-

gies for supporting the AI Act. As discussed earlier, the AI Act's provisions in regard to risk management, documentation, and registration are central to compliance. Therefore, this thesis focuses on the aforementioned provisions, in particular their *information requirements*—the pieces of information that need to be maintained, documented, and registered. Given that the AI Act is recently finalised, there are no authoritative guidelines, nor comprehensive analysis of the Act in the state of the art, available yet. Therefore, prior to investigating the use of Semantic Web technologies, identification of relevant information requirements through an analysis of the AI Act is essential.

1.2 Research Scope

The EU AI Act is the primary knowledge source used in this work. As mentioned earlier, to facilitate interpretation of the Act, relevant AI standards from the current ISO/IEC standardisation landscape are used as knowledge sources. Narrowing down the scope of the research, this thesis considers a fraction of the AI Act and a few ISO/IEC standards that are of most relevance to the topics under investigation in this thesis, which will be discussed in Subsection 1.2.1 and Subsection 1.2.2. In addition, as risk is a broad concept that can be interpreted from variety of perspectives, the scope of this research in regard to risk will be clarified in Subsection 1.2.3.

1.2.1 Scope Regarding the AI Act

Over the course of undertaking this research, the AI Act was undergoing the EU ordinary legislative process [27]. Following this process, the AI Act was first proposed by the European Commission [28] in April 2021. This proposal had to be approved by both the European Parliament and the Council of the European Union to be passed as EU legislation. At the end of its term in June 2022, the French presidency of the Council published a consolidated version. The Council's General Approach (also known as Common Position) [29] was issued in November 2022 by the Czech presidency. During the first reading of the Act in the European parliament, more than 3000 amendments were tabled by the responsible committees, namely the Committee on the Internal Market and the Committees on Consumer Protection and Civil Liberties, Justice and Home Affairs. The finalised Parliament's mandate, published in June 2023, enabled the entering of the trilogue phase, whereby the Commission, Parliament, and Council negotiated the AI Act behind closed doors to reach a political compromise. The agreed final text was published in July 2024 in the Official Journal of the European Union.

Consequently, since April 2021, this research has evolved over 7 distinct iterations of the AI Act published by EU authorities, as listed in the following in chronological order:

1. The European Commission's proposal [28], published in April 2021,
2. The French Presidency of the Council of the EU's Consolidated text [30], published in June 2022,
3. The Czech Presidency of the Council of the EU's General Approach [29], published in November 2022,
4. The European Parliament's mandate [31], published in June 2023,
5. The provisional agreement [32] resulting from inter-institutional negotiations during the trilogue phase, published in February 2024,
6. The Corrigendum [33], published in April 2024,
7. The final text [5], published in Official Journal of the EU in July 2024.

To deal with multiple mandates, the methodology for this research enabled frequent amendments (see Subsection 1.4.1). The work conducted prior to the publication of the final text were updated to reflect the latest status of the Act. Thus, all the references made to the AI Act in this thesis refer to the final version [5], unless otherwise stated.

The scope of the AI Act covers both uses of AI systems and general-purpose AI models, however, this work focuses on the former. This is due to the fact that the process for governing general-purpose AI models seems separate from product safety regulation that leverages the NLF structure. This implies a divergence between risk management process for high-risk AI systems and general-purpose AI models under the AI Act. In addition, the initial AI Act's proposal [28] did not impose obligations for general-purpose AI models and these were negotiated at the trilogue stage. Inclusion of *general-purpose AI systems*, *foundation models*, and *generative AI* was first proposed by co-legislators, i.e. the Council and the European Parliament, primarily in response to the recent advancement in Large Language Models (LLMs) [34]. As a result of trilogue negotiations on this matter, the final version of the AI Act lays down requirements for general-purpose AI *models*.

As mentioned earlier, the AI Act's regulatory regime is structured around risk, therefore obligations related to the risk-based classification and risk management are central in compliance with the Act's requirements for high-risk AI systems. Demonstrating compliance with these obligations requires

information about the system and its risks to be maintained and shared in the form of technical documentation. Further, a set of information needs to be registered into the EU database, wherein the information should be maintained in a machine-readable manner. Based on these, the scope of this thesis focuses on the obligations of high-risk AI providers in respect to risk management, documentation, and registration. In this thesis, the AI Act is analysed in its entirety, with a strong focus on the articles and annexes listed in Table 1.1.

Table 1.1: List of articles and annexes from the AI Act [5] used as key sources in the thesis

Obligations related to	**Sources from the AI Act**
Risk management	Article 5 - Prohibited AI practices Article 6 - Classification rules for high-risk AI systems Annex I - List of Union harmonisation legislation Annex III - High-risk AI systems referred to in Article 6(2) Article 9 - Risk management system
AI and risk documentation	Article 9 - Risk management system Article 11 - Technical documentation Annex IV - Technical documentation referred to in Article 11(1)
AI and risk information sharing	Article 13 - Transparency and provision of information to deployers Article 49 - Registration Article 71 - EU database for high-risk AI systems listed in Annex III Annex VIII - Information to be submitted upon the registration of high-risk AI systems in accordance with Article 49

1.2.2 Scope Regarding Standards

As previously stated, the AI Act relies on harmonised standards to provide technical support for addressing its essential requirements. The draft standardisation request [35] was communicated to European Standardisation Organisations in December 2022, with the official request [13] published in May 2023. As a response, CEN-CLC/JTC 21 on AI[4] has been occupied with the adoption of existing standards developed by ISO/IEC JTC 1/SC 42 as

[4] https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/

European standards and development of new standards to address requirements of the European Commission in regard to harmonised standards.

From the existing set of standards established by ISO/IEC JTC 1/SC 42, ISO/IEC 22989:2022 - Information technology — Artificial intelligence — Artificial intelligence concepts and terminology [36] is a foundational standard providing the terms and concepts. Given that for compliance with the AI Act, its requirements should be addressed under a quality management system (Article 17), ISO/IEC 42001 - Information technology — Artificial intelligence — Management system [37] is a strong candidate for harmonisation [38], providing normative requirements for establishing, implementing, and maintaining an AI management system— which is a set of elements of an organisation to establish AI policies, objectives, and processes to achieve those objectives (ISO/IEC 42001, 3.4).

Regarding risk management, ISO/IEC 23894 - Artificial intelligence — Guidance on risk management [39], published in February 2023, is a key standard that aims to guide organisations in managing AI risks through integration of risk management tasks into AI development tasks or any activity that incorporate AI. This standard is a specialisation of ISO 31000:2018 - Risk management — Guidelines [40], which is the ISO's generic risk management guideline. In addition, for terminology ISO/IEC 23894 uses ISO Guide 73:2009, Risk management — Vocabulary [41]. Given that the ISO Guide 73 has been withdrawn, this thesis uses ISO 31073:2022 - Risk management — Vocabulary [42], which defines generic risk terminology. In addition to ISO/IEC, activities by national and international standardisation bodies have been undertaken, among which are NIST's AI Risk Management Framework (AI RMF) [2] and IEEE 7000-2021 on process for addressing ethical concerns [43]. While this thesis acknowledges the usefulness of these standards in establishing AI risk management systems, it does not directly use them in relation to the analysis of the AI Act, considering that politically it is unlikely for these standards to be harmonised in the EU.

Regarding documentation, ISO/IEC DIS[5] 12792 - Information technology — Artificial intelligence — Transparency taxonomy of AI systems [44] is under development by ISO/IEC JTC 1/SC 42, as of August 2024. However, the European Commission does not request European standards in relation to technical documentation obligations [13]. Since this standard is not published yet, and more importantly, the use of harmonised standards for addressing technical documentation requirements is uncertain, the scope of this thesis does not include ISO/IEC DIS 12792 [44].

At the time of writing, the European Standardisation Organisations are

[5]Draft International Standard

working on deliverables to prepare a response to the European Commission's standardisation request. Therefore, harmonised standards in relation to the AI Act are not published yet. In this absence, the aforementioned ISO/IEC standards (listed in Table 1.2), that are contenders for harmonisation, are used in this work. It, however, should be noted that these standards have been found insufficient in meeting the requirements of the AI Act [38]. This thesis does not aim to investigate the extent of alignment or sufficiency of these standards for compliance with the provisions of the AI Act, it rather investigates how these existing standards can assist in implementation of the requirements.

Table 1.2: List of standards used as key sources to assist with interpretation of the AI Act in the thesis

Obligations related to	**Standards**
Risk management	ISO/IEC 23894:2023 - Artificial intelligence — Guidance on risk management [39] ISO 31000:2018 - Risk management — Guidelines [40] ISO 31073:2022 - Risk management — Vocabulary [42]
AI and risk documentation and sharing	ISO/IEC 22989:2022 - Information technology — Artificial intelligence — Artificial intelligence concepts and terminology [36] ISO/IEC 42001:2023 - Information technology — Artificial intelligence — Management system [37]

1.2.3 Scope Regarding AI Risks

Regarding the scope of **AI risk**, this thesis follows the AI Act's approach in taking a domain-agnostic and horizontal viewpoint to develop generic semantic models, which can be further expanded for creating new models that address particularities of specific domains. In addition, the scope of the AI Act in respect to risk—that is risk to 3 key areas of health, safety, and fundamental rights—is prioritised in this work. This scope, however, is not very well aligned with ISO/IEC 23894, whose focus is mainly on organisational risks, with limited reference to risk to individuals and society (see point 6.4.3.2 on assessment of consequences in ISO/IEC 23894) [38]. It should be noted that with the upcoming enforcement of the AI Act, AI providers and deployers will face regulatory risks related to non-compliance. Additionally,

for providers and deployers of high-risk AI systems, risks to health, safety, and fundamental rights would be translated into reputational risks and ultimately would be regarded as organisational risks. With these being said, ISO/IEC 23894 in this thesis is used as a guiding document for identification of risk concepts, their relations, and information elements that should be documented throughout the risk management process.

Regarding **AI risk management**, the scope of this research does not include risk management processes. It rather is focused on providing a model that enables AI and risk specifications, which result from planning or performing risk management processes. Methods and approaches for performing AI risk management, for example, statistical methods to measure robustness of a neural network [45] and approaches for bias testing [46], are beyond the scope of this work, given that this thesis focuses on modelling information requirements, rather than providing methods for risk management.

1.3 Research Question

The research question this thesis investigates is:

To what extent can Semantic Web technologies facilitate compliance with the EU AI Act's risk management, documentation, and registration requirements?

1.3.1 Terminology Used in the Thesis

Within the context of the research question and this thesis, the following definitions are used:

Semantic Web Technologies

Semantic Web technologies refer to a combination of methods and tools arising out of the field of Semantic Web [47], many of which are based on the use of World Wide Web Consortium (W3C) standards and languages. In this work, the key standards and languages used are: Resource Description Framework (RDF) [48], OWL 2 Web Ontology Language [49], Simple Knowledge Organization System (SKOS) [50], SPARQL query language [51] for information retrieval, Shapes Constraint Language (SHACL) [52] for rule-checking, the Open Digital Rights Language (ODRL) [53] for expressing policies, and Data Catalog Vocabulary (DCAT) [54] for cataloguing.

AI Risk

The AI Act, Article 3(2) defines risk as:

> *"'risk' means the combination of the probability of an occurrence of harm and the severity of that harm"*

This definition was introduced by the European Parliament mandate wherein *significant risk* is also defined as "a risk that is significant as a result of the combination of its severity, intensity, probability of occurrence, and duration of its effects, and its the ability to affect an individual, a plurality of persons or to affect a particular group of persons". However, this definition is not included in the final text and therefore the final text does not provide any definitions of risk in the context of AI. Similarly, ISO/IEC 23894 does not introduce any modified definitions for AI risks and relies on ISO 31000's generic definition of risk, that is:

> *"effect of uncertainty on objectives*
> *Note 1 to entry: An effect is a deviation from the expected. It can be positive, negative or both, and can address, create or result in opportunities and threats.*
> *Note 2 to entry: Objectives can have different aspects and categories, and can be applied at different levels.*
> *Note 3 to entry: Risk is usually expressed in terms of risk sources, potential events, their consequences and their likelihood."* (ISO 31000 [40], 3.1)

As both definitions do not consider risk in the AI context, within this thesis *AI risk* is conceptualised as ***the state of uncertainty associated with an AI system, that has the potential to cause harms and is expressed in terms of risk sources, consequences, impacts, likelihood, and severity***. The definition only includes harms (negative effects), given the absence of risk-benefit analysis in the AI Act [55]. In regard to how risk is expressed, while likelihood and severity are explicitly mentioned in the Act's definition of risk, they are not sufficient in reflecting the notion of risk adopted within the Act, which seeks to safeguard individuals from negative *impacts* of AI on health, safety, and fundamental right. This definition distinguishes between the effects of risk on systems and operations (*consequence*) and the effects on individual and groups, which fall under the scope of 'harm' in the AI Act (*impact*). Aiming to promote transparent and accountable risk management practices, the AI Act, particularly its data governance requirements, hints on the need to "nip risk in the bud" by linking risk to its sources.

AI Risk Management

Within the AI Act, the traces of AI risk management appears in (i) the AI system assessment to situate the system in the risk-based classification (Articles 5 and 6), (ii) the risk management system, which "shall be understood as a continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic review and updating." (Article 9), and (iii) fundamental rights impact assessments (FRIAs), mandated for AI deployers (Article 27). In this thesis, the first two activities are considered as part of AI risk management that should be performed by AI providers. FRIA is considered as a separate but highly-related activity that is expected to be performed by AI deployers, and therefore is beyond the scope of risk management that should be conducted by providers.

Information Requirements of the AI Act

Information requirements are defined in connection with the binding legal obligations imposed by the AI Act. These requirements are interpreted from the AI Act and specify the information that needs to be maintained, queried, or shared to comply with the Act. As will be shown in this thesis, this information is related to technical details of the system, the context within which the AI system is used or intended to be used, and risk management system implemented in regard to the system.

Information Documentation

Information documentation refers to providing communicable material, in any format and media.

Information Registration

Information registration refers to sharing information with a third-party, particularly authorities, to index an AI system into a repository or database. In this thesis, *AI authority* refers to bodies involved in assessment, monitoring, or investigation of AI systems or AI-related incidents, which includes the *AI Office*, *conformity assessment bodies*, and *notified bodies*.

AI Use Case

AI use case refers to a given use of an AI system in a specific context. Examples of AI use case that are mentioned in this thesis are: an AI-based student proctoring system, a facial recognition system used for identification

of drivers, and an individual risk assessment system for predicting risk of gender violence.

1.3.2 Research Objectives

Research Objective 1 (RO1)

The first step towards addressing the research question concerns analysis of the AI Act's obligations that are relevant to the research question (mentioned in Subsection 1.2.1). On the basis of this information *requirements* and the *applications* of Semantic Web-based artefacts are specified. As discussed in Subsection 1.2.2, relevant ISO/IEC standards are used to assist with the analysis.

RO1. Analyse the EU AI Act to identify the information requirements for:
(a) determining an AI system's risk category as per the AI Act's risk-based classification,
(b) expressing intended purpose of an AI system,
(c) generating risk management documentation and technical documentation, and
(d) registering high-risk AI systems into the EU database.

Research Objective 2 (RO2)

Fulfilment of RO1 results in identification of information that should be maintained for compliance with risk management, documentation, and registration requirements of the AI Act. Compliance checking, whether it is performed by the provider (self-assessment) or by conformity assessment bodies (third-party assessment), involves sharing and querying this information. As argued earlier, Semantic Web technologies have been successfully adopted for RegTech solutions, in particular for compliance with the EU digital regulations. Aiming to investigate the use of Semantic Web technologies in compliance with the AI Act, this thesis provides a semantic model of the information requirements in a FAIR [17] manner. This is pursued by developing an OWL 2 ontology that models key AI and risks concepts and relations. Using OWL 2 enables automated reasoning and facilitates data integration across multiple resources. This provides the second research objective as:

RO2. Design a FAIR ontology that enables modelling of AI use cases in a way that the information requirements, identified from the AI Act (RO1), are addressed.

Research Objective 3 (RO3)

Following the development of the ontology, Semantic Web technologies, including W3C well-established methods and standards, are employed to implement FAIR approaches and mechanisms to assist with AI Act compliance tasks. This includes utilising Semantic Web capabilities for validating a use case against the rules for high-risk AI systems, retrieving information to generate documentation, expressing AI use policies, and cataloguing AI systems. Therefore, the third objective is defined as follows:

RO3. Implement and evaluate FAIR approaches based on the ontologies developed in RO2 and by utilising Semantic Web technologies to assist with
(a) determining high-risk AI systems as per Annex III,
(b) expressing intended purpose of AI systems,
(c) generating risk management documentation and technical documentation,
(d) cataloguing AI systems for registering into the EU database.

Research Objective 4 (RO4)

Compliance with the AI Act requires risk management system documentation and technical documentation to be extensive. The extensiveness, as well as confidentiality concerns, hinder information sharing within the AI value chain. This also impedes comparison of multiple AI systems, which is needed for adopting suitable AI solutions, particularly in AI procurement, for investigating the effectiveness of risk management practices by citizens and civil organisations, and for effective cooperation and exchange of information among market surveillance authorities in different sectors or different EU Member States. With a semantic model of the information requirements (RO1), that is developed leveraging the ontology (RO2), and the approaches and mechanism implemented using Semantic Web technologies (RO3), providing customised views, that can be configured to respect confidentiality concerns, is straightforward to implement. Establishing a summarised view is valuable in addressing the struggles in the sharing and collation of multiple AI systems and their risk models as well as in communication of AI and risk information with a wide range of stakeholders. This leads to the fourth research objective as:

RO4. Create and evaluate a documentation framework based on the information requirements, identified in RO1, and the artefacts developed in RO2 and RO3 for providing a holistic view of AI use cases and their risks in a summarised manner, as per the documentation requirements of the AI Act.

1.4 Research Methodology

1.4.1 Research Methodology and Technical Approaches

The overall research methodology, shown in Figure 1.1, consists of three key steps: (1) analysis of the AI Act (RO1), (2) ontology development (RO2), (3) development of compliance mechanisms and tools (RO3 and RO4). Considering the AI Act was under development over the course of this research, an iterative methodology was used to ensure the amendments made to the Act were reflected in the developed ontology and tools.

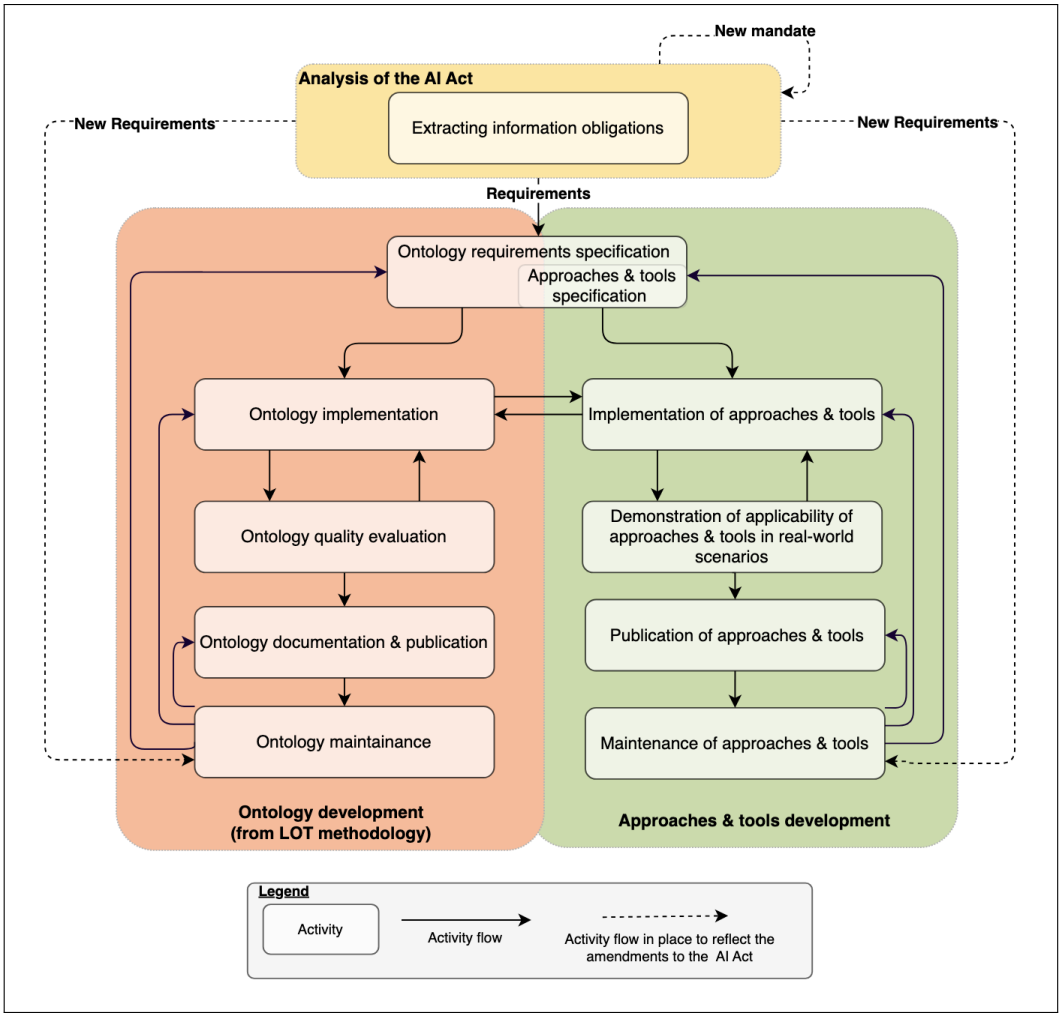


Figure 1.1: The overall research methodology

Analysis of the AI Act

The first step in addressing the research question was to fulfil RO1 by gathering information from the AI Act and related standards. The analysis led to the following:

- A codified set of concepts to determine whether an AI system is high-risk as per Annex III or prohibited as per Articles 5 and 6 to address RO1(a) (see Section 3.2),
- An analysis of the description of intended purpose to address RO1(b) (see Section 3.3),
- A detailed analysis of the AI risk management system and technical documentation requirements to address RO1(c) (refer to Section 3.4),
- An analysis of the registration obligations to address RO1(d) (see Section 3.5).

Ontology Development

For ontology development (RO2), the *Linked Open Terms (LOT) methodology*, which is a lightweight iterative methodology for ontology development proposed by Poveda-Villalón et al. [56], was used. The rationale behind selecting the LOT methodology is as follows: first, LOT's iterative workflow makes the methodology a great fit for this research to deal with the changes made to the AI Act throughout the ordinary legislative procedure. This also makes LOT suitable for future updates and revisions of the ontology with the upcoming publication of supporting material related to implementation and enforcement of the Act, including harmonised standards (Article 40), the AI Office-issued guidelines and templates, European Commission's implementing and delegated acts (Article 97), and national and industry-specific policies. New insights also will emerge from the AI risk learning mechanisms established in the AI Act through regulatory sandboxes (Article 57) and real-world testing environments (Article 60), which can impact the ontologies developed in this work. The second reason for using LOT is that its alignment with industrial development enhances industry usage and extension of the ontology. Finally, utilising LOT, compared to other ontology development methodologies, is more straightforward due to its comprehensive guidelines on ontology development and evaluation processes, which are accompanied by tool recommendations.

Following the LOT methodology and based on the requirements identified in the previous step (i.e. analysis of the AI Act to fulfil RO1), the concepts

and relations were identified and modelled as an OWL 2 ontology. The **AI Risk Ontology (AIRO)** includes the key concepts and relations for modelling AI use cases and their risks. Following the design of the Data Privacy Vocabulary (DPV) [57] in developing information models from legal sources in a modular and manageable manner, the instances of core AIRO concepts were modelled within the **Vocabulary of AI Risks (VAIR)**. Both AIRO and VAIR will be described in Chapter 4.

To ensure quality of AIRO and VAIR, Semantic Web best practices and guidelines including W3C Best Practice Recipes for Publishing RDF Vocabularies [58] and Data on the Web Best Best Practices [59], OOPS! common ontology pitfalls [60], and FAIR best practices [61, 62] were followed. The ontologies have been made available online on GitHub under permissive licenses, permitting further reuse and enhancement by the community. Documentation of AIRO and VAIR was created using ReSpec[6], a HTML template for W3C specifications, and published under the CC-BY-4.0 license[7] using W3ID permanent identifiers[8] at `https://w3id.org/airo` and `https://w3id.org/vair`, respectively.

The applicability of AIRO and VAIR in modelling AI use cases and their risks will be demonstrated using one synthetic use case (Proctify) and two following incidents indexed in the AI, Algorithmic and Automation Incidents and Controversies (AIAAIC) repository[9]—an open-access dataset of more than 1500 AI incidents and issues covered by the media: Uber's Real Time ID (incident ID = AIAAIC0756) and VioGén (incident ID = AIAAIC0848). The use cases are listed in the following and will be detailed in Section 4.4:

- *Proctify* is a synthetic AI-based student proctoring system, described in consultation with Joint Research Centre (JRC) researchers, based on [63, 64]. This use case was designed to demonstrate the full potential of AIRO and VAIR as well as the Semantic Web artefacts proposed in this thesis, without being limited to the information provided by third-parties.

- *Uber's Real Time ID (RTID)* was used as a facial recognition identification system to ensure that the driver's account is not used by anyone other than the registered Uber driver. This incident was chosen due to its use of the prevalently-used facial recognition technology. Also, it was chosen to include an example of risk to fundamental rights.

[6] https://respec.org/docs/
[7] https://creativecommons.org/licenses/by/4.0/
[8] https://w3id.org/
[9] https://www.aiaaic.org/aiaaic-repository

- *VioGén Domestic Violence System* [65] is currently being used by the Spanish law enforcement agencies to assess the risk of being assaulted again for a victim of gender violence, upon which the victim's eligibility for police protection is determined. This case was selected as an example of risk to both rights and safety.

Development of Compliance Mechanisms and Tools

To address RO3, the following standardised Semantic Web technologies, summarised in Figure 1.2, were used:

- To achieve RO3(a), semantic rule-checking process was defined through utilising Shapes Constraint Language (SHACL) [52], which is the standard for validating RDF graphs. In this, Annex III high-risk AI rules were modelled using SHACL shapes to enable expressing high-risk AI conditions in a machine-readable manner. Based on this, a web application was developed and published under the MIT license[10] (see Section 5.1).

- For describing AI intended purposes (RO3(b)), the **AI Use Policy (AIUP)** profile was developed by extending the W3C's recommendation on Open Digital Rights Language (ODRL) [53]. This provides a technical solution for declaring intended purposes of AI systems as AI use policies in an open, machine-readable, and interoperable format based on the evolving requirements of the AI value chain, particularly the obligations of the EU AI Act (see Section 5.2).

- To address RO3(c), information featured in the AI and risk documentation was retrieved by executing SPARQL queries [51] over a machine-readable specification of a given AI use case (see Section 5.3).

- To address RO3(d), **AICat** (AI Catalogue vocabulary) was developed as an extension of the Data Catalog Vocabulary (DCAT) [54], which is the W3C standard for modelling datasets as well as general resources within a catalogue. AICat supports the data governance requirements to build and maintain registries of AI systems such as the EU high-risk AI database (see Section 5.4).

[10] https://opensource.org/license/MIT

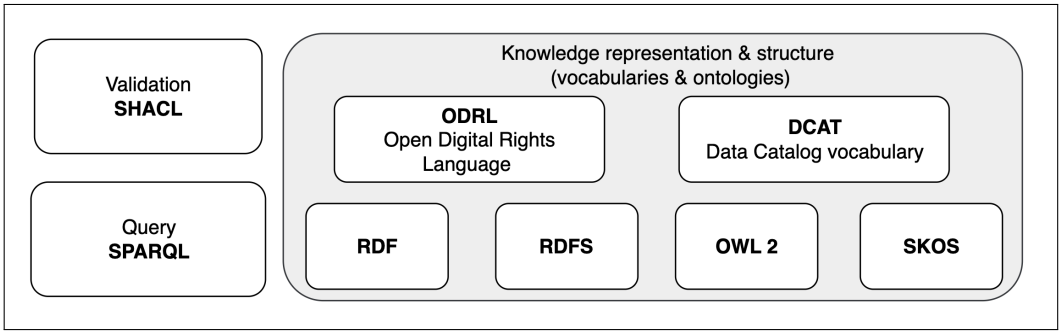


Figure 1.2: Technologies used in implementation of the thesis artefacts

Addressing RO4, a documentation framework called **AI Cards**, has been proposed for documentation of AI use cases in both human- and machine-readable representations. As a framework, the AI Cards is characterised as a coherent and adaptable structure that (i) provides a summarised overview of AI and risk information within a logically organised visual specification and (ii) encodes this information in a machine-readable format using the ontologies developed in this work, i.e. AIRO and VAIR.

A summary of key artefacts proposed in this thesis and how they relate to each other is depicted in Figure 1.3.

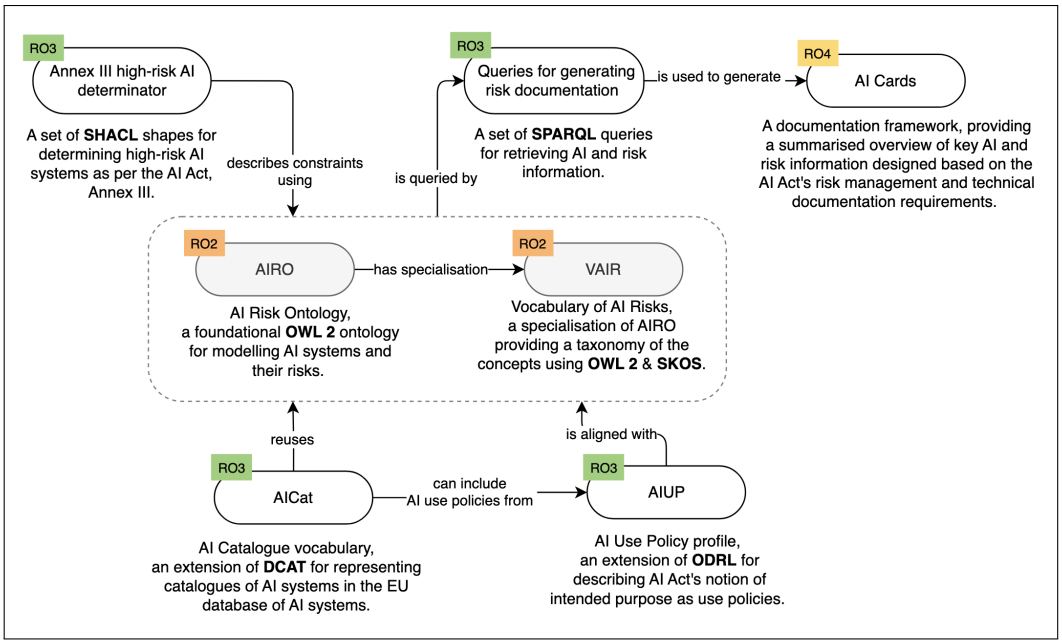


Figure 1.3: Key artefacts presented in this thesis and their relations

1.4.2 Evaluation Strategy

The evaluation of the research presented in this thesis takes two overall dimensions. The first dimension consists of validation of the AI Act analysis and evaluation of the usefulness of the AI Cards framework. The second dimension focuses on evaluation of Semantic Web-based artefacts from a technical perspective.

The best practice for building conceptual information models from legal text requires collaboration of multidisciplinary teams [66, 57]. Following the best practice, EU policymakers involved in development of the AI Act, experts in digital law and AI standardisation, and experts in legal data modelling were consulted in an iterative manner at different stages to ensure the AI Act analysis provided in this thesis (Chapter 3) is valid. Further, the author participated in ISO/IEC JTC 1/SC 42 and CEN-CLC/JTC 21 meetings to gain insights regarding related standards. It should be mentioned that publication of future guidelines, harmonised standards, and case laws allow further validation of this work in the future (for the details of the validation see Section 7.2).

The AI Cards framework was designed and validated through consultation with experts from European Commission's Joint Research Centre (JRC), who provide research services to the European Commission in regard to the AI Act and other EU digital regulations (refer to Subsection 7.4.1). Further validation of the AI Cards was performed through an anonymous online survey to assess its usefulness (see Subsection 7.4.3). An overview of methods used for validation of non-Semantic Web artefacts is shown in Table 1.3.

Table 1.3: Overview of the approaches used for validating analysis of the AI Act and the AI Cards framework

Approach ↓ Artefacts ⟶	The AI Act analysis	The AI Cards framework
Consultation with experts	✓	✓
User study	✗	✓
Comparison with SOTA	✓	✓

The Semantic Web-based artefacts developed in the thesis were assessed based on the following criteria:

- Their sufficiency in addressing the identified requirements from the AI Act (functional requirements),
- Their quality (non-functional requirements),
- Their applicability, which is shown through proof-of-concept implementation,
- Their distinguishing features in comparison with the state of the art.

An overview of approaches used for evaluation of the Semantic Web-based artefacts is presented in Table 1.4.

Table 1.4: Overview of the approaches used for evaluation of Semantic Web-based artefacts

Approach ↓ Artefacts ⟶	AIRO	VAIR	High-risk AI (SHACL)	AI Cards (SPARQL)	AIUP	AICat
Fulfilment of information requirements	✓	✓	✓	✓	✓	✓
Logical consistency	✓	✓	✓	✓	✓	✓
Following W3C best practices	✓	✓	✓	✓	✓	✓
Following FAIR principles	✓	✓	✓	✓	✓	✓
Demonstration of applicability	✓	✓	✓	✓	✓	✓
Comparison with SOTA	✓	✓	✓	✓	✓	✓
Peer-reviewed publication	✓	✓	✓	✓	✓	✗
Reproducibility (open-access resources)	✓	✓	✓	✓	✓	✓

1.5 Contributions

1.5.1 Major Contribution: a Set of FAIR Artefacts to Assist with Compliance with the EU AI Act

By providing a compendium of Semantic Web-based artefacts to assist with the compliance with the AI Act, this thesis provides **novel FAIR RegTech solutions for compliance with the AI Act's requirements** regarding risk management, documentation, and registration. This compendium consists of 2 ontologies (AIRO and VAIR), SHACL shapes for determining high-risk AI systems, SPARQL queries for retrieving information for documentation generation, an ODRL profile (AIUP) for describing AI intended purposes as policies, and an extension of DCAT (AICat) for describing AI systems in a catalogue. All these resources made available online under permissive licenses to allow their free use and further enhancement by the community (see Table 1.5 for the links to resources).

Table 1.5: Links to open resources

Artefact	Link to open resources
AIRO	https://w3id.org/airo
VAIR	https://w3id.org/vair
SHACL shapes for determining Annex III high-risk AI	https://github.com/DelaramGlp/airo/blob/main/high-risk-shacl/shapes-final.ttl
Web application for determining Annex III high-risk AI	https://github.com/DelaramGlp/highrisk_app https://regtech.adaptcentre.ie/highrisk
SPARQL queries for documentation (AI Cards) generation	https://github.com/DelaramGlp/aicards/tree/main/sparql_queries
AIUP	https://w3id.org/aiup
AICat	https://w3id.org/aicat

1.5.2 Major Contribution: The AI Cards Framework

The AI Cards, as a major contribution of this thesis, positions itself as a documentation framework developed based on the provisions of the AI Act. Although there is a considerable body of work on AI documentation, AI Cards is a **novel holistic framework for documenting a *given use* of an AI system based on the AI Act**'s risk management systems and technical documentation requirements in two complementary formats: a visual human-readable representation and a machine-readable specification.

While the visual representation provides a summary of key information, the machine-readable representation enables the provision of information with a higher level of detail, facilitates consistency checking, and further allows automation in generation of the AI Cards (and other documentation).

1.5.3 Minor Contribution: Analysis of the AI Act

The final text of the AI Act was published in July 2024. The AI Act establishes mechanisms for providing further guidelines, templates, and tools to assist with the implementation and enforcement, through publication of harmonised standards, AI Office-issued guidelines and codes of conduct, and Commission-issued implementing and delegated acts. At the time of writing, none of these are available and there is also no case laws available to assist with the interpretation of the AI Act. Further, there is a lack of academic resources as well as comprehensive openly-accessible expert analysis of the AI Act; the latter is mainly due to the fact that this is a competitive advantage for the consultancy firms and Tech companies. Therefore, the minor contribution of this thesis include an analysis of the EU AI Act's provisions in regard to classification rules for AI systems, AI risk management, technical documentation, and registration. Within this, the 5-concept structure proposed for determining high-risk AI systems as per Annex III (Subsection 3.2.1) has gained considerable traction (see Subsection 8.3.1).

1.5.4 Other Contributions

Contribution to the EU's Horizon 2020 PROTECT Innovative Training Network (ITN)

This research has been funded and carried out as part of the PROTECT (Protecting Personal Data Amidst Big Data Innovation) project—a Horizon 2020 multidisciplinary Innovative Training Network (ITN) with researchers from fields of knowledge engineering, technology ethics, and data protection law. As a result of collaboration with Early Stage Researchers (ESRs) across the PROTECT ITN, an extension of AIRO for representing risks of using AI in the heath domain was developed. The Health AI Risk Taxonomy (HART) [67] was populated based on real-world incidents indexed in the AIAAIC repository in a collaborative manner within PROTECT's Work Package 3 (WP3). The author, in collaboration with another knowledge engineering researcher, were responsible for implementation of HART. The taxonomy is published under the CC-BY-4.0 license at https://w3id.org/hart.

Contribution to the DPVCG and AIAAIC

The W3C Data Privacy Vocabularies and Controls Community Group[11] (DPVCG) works towards development of semantic models for expressing information related to personal data processing based on legal requirements, in particular the GDPR. The ontologies proposed in this thesis, i.e. AIRO and VAIR, contributed to expansion of the scope of DPV to include concepts for representing use of technologies, including AI, and supporting implementation of the AI Act. In the development of the newly-published DPV 2.0 [57], the author has contributed to the core DPV specification[12], the EU-AIAct[13] and TECH[14] extensions by proposing more than 150 concepts.

The author also has been involved in the development of the AI, algorithmic and automation harms taxonomy, which is being developed by the AIAAIC working group [68]. Establishing the core objectives in regard to machine-readability was mainly the result of the author's contribution.

Contribution to AI Standardisation and Policies

The author has been participating in ISO JTC 1/SC 42 and CEN-CLC/JTC 21 as a nominated Irish expert through membership in the National Standards Authority of Ireland (NSAI) since January 2021. In addition to participation in plenary meetings, a systematic mapping of the Assessment List on Trustworthy AI (ALTAI) [69] with the then-under-development ISO/IEC 42001 was submitted to the CEN-CLC/JTC 21 as an input for future standardisation activities. Further, the AI Cards framework was indexed as an official contribution to the Irish AI committee in NSAI (NSAI/TC 2/SC 18).

In respect to contribution to AI policies, the author contributed to the ADAPT Centre's response submitted to the public consultation on implementation of the AI Act held by the Irish Department of Enterprise, Trade and Employment (DETE). During a secondment at the European Commission's JRC, the author collaborated with experts involved in EU policymaking activities and discussed the potential of the analysis and the RegTech solutions provided in this thesis as an input to development of future policies and guidelines related to the AI Act.

[11] https://www.w3.org/groups/cg/dpvcg/
[12] https://w3id.org/dpv/2.0
[13] https://w3id.org/dpv/legal/eu/aiact
[14] https://w3id.org/dpv/tech

1.5.5 Publications

The following peer-reviewed papers have been published in association with this research:

1. "**AIRO: An Ontology for Representing AI Risks Based on the Proposed EU AI Act and ISO Risk Management Standards**" [70]

   ***Delaram Golpayegani**, Harshvardhan J. Pandit, and Dave Lewis*

   The 18th International Conference on Semantic Systems (SEMANTiCS), 2022

   Citation count[15] *= 14*

   This publication presents an initial version of AIRO as an ontology for modelling AI use cases and their risks as per the Commission's proposed AI Act [28] and key standards from the ISO 31000 series, i.e. ISO 31000 [40] and ISO 31073 [42]. The paper provides an analysis of Annex III high-risk AI applications and a high-level list of information that needs to be featured in technical documentation. It further discusses use of SHACL for describing conditions that make an AI system high-risk under the proposed AI Act and use of SPARQL for generating technical documentation.

2. "**To Be High-Risk, or Not To Be—Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards**" [71]

   ***Delaram Golpayegani**, Harshvardhan J. Pandit, and Dave Lewis*

   The sixth annual ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), 2023

   Citation count = 14

   This paper discusses the criteria for high-risk AI systems, as defined in the Council's General Approach [29], and proposes 5 concepts whose combination makes an AI system high-risk. It further proposes VAIR and shows how SHACL can be used to automate reasoning needed for high-risk AI determination. The paper also provides an analysis of the scope of standardisation activities within ISO and CEN-CENELEC in regard to high-risk AI requirements.

[15] According to Google Scholar, as of September 2024

3. "**AI Cards: Towards an Applied Framework for Machine-Readable AI and Risk Documentation Inspired by the EU AI Act**" [72]

   ***Delaram Golpayegani**, Isabelle Hupont, Cecilia Panigutti, Harshvardhan J Pandit, Sven Schade, Declan O'Sullivan, Dave Lewis*

   The Annual Privacy Forum (APF), 2024

   This publication is focused on risk management system and technical documentation requirements of the AI Act. Based on an analysis of the aforementioned requirements, the paper presents the AI Cards framework as a novel approach for documenting a given use of an AI system, with a discussion on how its human- and machine-readable representations can be a valuable input for future EU policymaking and standardisation efforts. This publication is the result of collaboration with JRC.

4. **"AIUP: an ODRL Profile for Expressing AI Use Policies to Support the EU AI Act"** [73]

   ***Delaram Golpayegani**, Beatriz Esteves, Harshvardhan J. Pandit, and Dave Lewis*

   Posters, Demos, Workshops, and Tutorials of the 20th International Conference on Semantic Systems (SEMANTiCS-PDWT), 2024

   This publication discusses the complexities in defining intended purpose of an AI system, as regulated by the AI Act, and puts forward the idea of expressing this complex concept as AI use policies. As a technical solution for declaring AI use policies in an open, machine-readable, and interoperable format, the paper extends ODRL through an AI Use Policy (AIUP) profile.

5. "**Comparison and Analysis of 3 Key AI Documents: EU's Proposed AI Act, Assessment List for Trustworthy AI (ALTAI), and ISO/IEC 42001 AI Management System**" [74]

   ***Delaram Golpayegani**, Harshvardhan J. Pandit, and Dave Lewis*

   30th Irish Conference on Artificial Intelligence and Cognitive Science (AICS), 2022

   Citation count = 6

   This publication provides an initial comparison of (the draft of) ISO/IEC 42001 AI management system standard with the ALTAI [69] and the

proposed AI Act [28] using an upper-level ontology for semantic interoperability between trustworthy AI documents [75] with a focus on activities to find the potential alignment between these 3 key documents. Part of this work was conducted prior to the publication of the AI Act proposal.

6. "**Towards a Taxonomy of AI Risks in the Health Domain**" [67]

   ***Delaram Golpayegani**, Joshua Hovsha, Leon WS Rossmaier, Rana Saniei, Jana Mišić*

   Fourth International Conference on Transdisciplinary AI (TransAI), 2022

   Citation count = 2

   This publication provides discussions on risks of using AI in the health domain from legal, ethical, and societal perspectives. Following from the discussions, it proposes HART—a taxonomy, based on AIRO, that mirrors the risks caused by the use of AI in the health sector according to a variety of different real-world incidents. This publication is the result of the collaboration within PROTECT's WP3.

7. "**Data Privacy Vocabulary (DPV)–Version 2**"

   *Harshvardhan J. Pandit, Beatriz Esteves, Georg P. Krog, Paul Ryan, **Delaram Golpayegani**, Julian Flake* [57]

   Accepted to be published in the proceedings of the 23rd International Semantic Web Conference (ISWC), 2024

   Citation count = 7

   This publication describes DPV 2.0 and includes the contributions of the author to DPV's AI Act and technology extensions.

1.6 Thesis Structure

The thesis roadmap is shown in Figure 1.4 and its structure is presented in the following. Chapter 2 reviews the state of the art according to four themes related to the research objectives. Chapter 3 delves into the Act's requirements. Building on the requirements identified, Chapter 4 presents AIRO and VAIR and illustrates how they can be applied in modelling AI use cases and their risks as knowledge graphs. Chapter 5 provides the implementation of Semantic Web-based approaches and mechanisms for determining

Annex III high-risk AI systems, generating documentation, expressing intended purposes, and cataloguing AI systems. It further shows applicability of the approaches through proof-of-concept implementation. Chapter 6 proposes the AI Cards, as a novel framework for documentation of AI use cases in two complementary human- and machine-readable representations. Chapter 7 demonstrates evaluation of the work presented in the thesis. Chapter 8 concludes the thesis with a discussion on how the thesis addresses the research question and objectives and outlines potential directions for future work. The supplementary material is presented in the appendices.

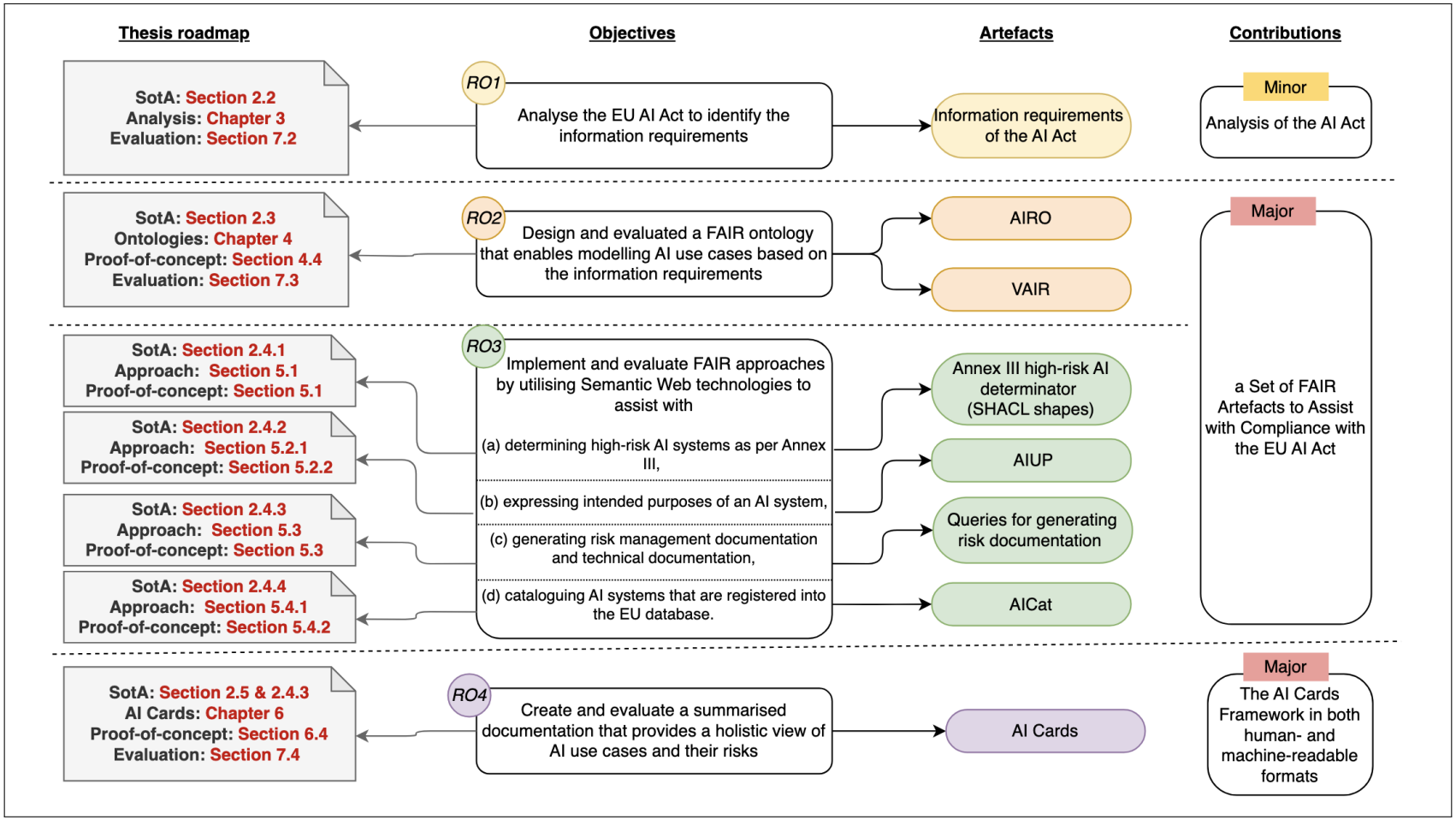


Figure 1.4: Thesis roadmap

CHAPTER

TWO

STATE OF THE ART

2.1 State of the Art Review Methodology

The state of the art analysis was conducted in regard to the four following themes:

1. In relation to RO1, the emerging body of work on the analysis of the EU AI Act, specifically its classification rules for prohibited and high-risk AI systems and its requirements regarding risk management, documentation, and registration, is reviewed in Section 2.2.

2. In relation to RO2, the emerging body work on AI risk taxonomies and ontologies is surveyed in Section 2.3.

3. In relation to RO3, existing Semantic Web-based approaches for implementation or enforcement of the EU AI Act are reviewed in Section 2.4.

4. In relation to RO4, existing documentation approaches for documenting AI use cases and their risks are analysed in Section 2.5.

Given that the AI Act was undergoing the EU's legislative procedure during the time this research was conducted, the state of the art has been updated frequently to reflect the latest state of affairs. The search was performed by querying scientific search portals, including Google Scholar[1], Scopus[2], IEEE Xplore[3], and ACM digital library[4], in addition to AI-based search

[1] https://scholar.google.com/
[2] https://www.scopus.com/
[3] https://ieeexplore.ieee.org
[4] https://dl.acm.org/

platforms such as Semantic Scholar[5] and SciSpace[6]. It should be noted that the AI-based engines were *solely* used for search purposes and therefore the AI-generated content was neither analysed nor used in this thesis. General search engines, e.g. DuckDuckGo[7], and social media platforms, including LinkedIn[8] and Twitter[9] (now known as X) were used for monitoring the research environment that rapidly progressed with the legislative process. The ongoing formation of the AI Act and the discourse on the EU digital policies were followed in forums including Euractiv[10]. In addition, formation of AI-related standards in CEN-CLC/JTC 21 and ISO/IEC JTC 1/SC 42 was monitored through membership in the Irish AI mirror committee (NSAI/TC 02/SC 18). The OECD's[11] catalogue of trustworthy AI tools[12] was also used in the search for existing tools related to the AI Act.

2.2 Analysis of the AI Act

Since the publication of the AI Act's proposal by the European Commission in April 2021, the research community, international organisations, and industrial actors have been exploring the new avenues it opened by analysing the AI Act's contents that needed clarifications, identifying potential gaps, or giving critique. While some of the previously-expressed comments and opinions are no longer applicable to the final version of the Act, they were helpful in shaping contributions of this thesis in regard to the AI Act's analysis (will be presented in Chapter 3). Among the work that reviewed the AI Act, the following have had significant impact on this thesis's view of the Act: the first and highly-cited analysis of the Act by Veale and Borgesius [8], the review of the Act by Members of the Robotics and AI Law Society (RAILS) [76], Smuha et al.'s response to the proposed AI Act [77], Mazzini and Scalzo's detailed explanation of the Act [10], and the recently-published work of Fernández-Llorca on the AI Act's terminology [34].

To give a direction to the review, the remainder of this section examines the studies that analyse those requirements of the Act that are relevant to this thesis. However, those comments and opinions published prior to the release of the final text, that are no longer applicable, are excluded.

[5] https://www.semanticscholar.org/
[6] https://typeset.io/
[7] https://duckduckgo.com/
[8] https://www.linkedin.com/
[9] https://twitter.com/
[10] https://www.euractiv.com/
[11] The Organization for Economic Cooperation and Development
[12] https://oecd.ai/en/catalogue/tools

2.2.1 The AI Act's Risk-Based Classification Rules

Existing studies on the AI Act's categories of **prohibited AI practices (Article 5)** are primarily focused on providing clarifications, in particular clarity regarding the definition and scope of *subliminal techniques*. Some of the notable studies are Neuwirth's analysis of prohibited categories stated in the commission's proposal [6], Bermúdez et al.'s effort to provide a definition for subliminal techniques [78], Franklin et al.'s proposed definitions for subliminal, purposefully manipulative and deceptive techniques [79], Bulgakova's analysis of the prohibition on the use of subliminal techniques [80], and Leiser's comparative analysis of prohibited uses that deploy manipulative techniques (Article 5(1a)) in 3 mandates of the Act, namely Commission's proposal, European Parliament's amendments, and Council's Common Position [81].

The body of work on the criteria for prohibited systems is mainly focused on clarification of the wording of the AI Act's text. Since these studies are not authoritative, there is still uncertainty regarding the prohibited AI legal regime. In addition, none of these studies establish a holistic view of the prohibited categories, nor do they analyse the individual concepts of AI use cases that make them prohibited.

On the matter of **high-risk AI systems (Article 6)**, some studies investigated specific types of AI systems that are perceived as high-risk under the Act. For example, Schwemer et al. have examined the implications of high-risk systems under the administration of justice and democratic processes, outlined in the Commission's proposal [82]. Dijck provides a discussion on whether individual risk assessment falls under Annex III high-risk categories [83]. Similar to the related work on prohibited systems, none of these studies specifically identify the individual concepts of use cases to determine when a use case falls under the high-risk category or may become high-risk through changes.

Closest to the approach of this thesis in regard to Annex III high-risk AI is the work of Hupont and Gómez [84]. This study identifies the use case information requirements for assessing an AI system's risk category according to the AI Act, which are: intended use, user, targeted persons, context of use, application areas, reasonably foreseeable misuse, inputs, and outputs. However, the study does not elaborate on how to determine the risk category using this information.

2.2.2 Risk Management and Technical Documentation Requirements

Article 9 - Risk Management System

The only available comprehensive legal analysis of the AI Act's risk management provisions is provided in the work of Schuett [85]. The analysis covers the role, purpose, and application of risk management in the Act in addition to the requirements outlined in Article 9 - *Risk management system*. However, it does not delve into the details of risk management system documentation.

Since the focus of this work is on documentation and sharing of risk information, rather than the approaches for AI risk management per se, analysis of the AI risk management approaches is considered out of the scope. Despite this, this thesis acknowledges some novel emerging approaches for compliance with AI risk management obligations, including AIRMan (AI Risk Management System) [86], Key AI Risk Indicators (KAIRI) framework [87], Novelli et al's AI risk assessment model that integrates the AI Act's risk-based approach with the risk framework developed by the Intergovernmental Panel on Climate Change (IPCC) [88], the quality model for safety-critical systems proposed by Kelly et al. [89], the Trustworthy Assurance Process [90], and the approach proposed by Novelli et al. to automate business process compliance for the AI Act's requirements regarding fundamental rights impact assessments [91].

Article 11 - Technical Documentation

There are some studies looking into the overall role of documentation within the AI Act. For instance, Panigutti et al. explore the role of the AI Act's transparency and documentation requirements in addressing the opacity of high-risk AI systems [63]. Gyevnara et al. also discuss compliance-oriented transparency required to satisfy the AI Act's requirements in regard to risk and quality management systems [92].

Specifically regarding information elements that are required to be documented as per Article 11 (*Technical documentation*) and Annex IV (*Technical documentation referred to in Article 11(1)*), the only available study is the work of Hupont et al. [93] that identifies 20 information elements that should be featured in documentation of AI systems and their constituting datasets. Nevertheless, it fails to be fully comprehensive in terms of covering both technical and risk management system documentation requirements. Mustroph and Rinderle-Ma [7] discuss requirements of technical documentation and risk management system to be fit within an AI quality management

system. The analysis lacks depth in regard to the information elements that should be included in risk and technical documentation.

2.2.3 Intended Purpose of AI Systems

Within the Act, the legal term of *intended purpose* is a key concept in identification of high-risk AI systems, as acknowledged by European Commission's researchers [34, 84, 64]. In the context of general-purpose AI systems, Fernández-Llorca et al. argue that "intended purpose has been incorrectly understood as specific purpose". Reiterating the AI Act's definition, which is "the use for which an AI system is intended by the provider", the authors highlight that intended purpose covers both specific and general purposes [34]. Within the Use Case Cards [64], an AI Act-inspired documentation approach that will be discussed later in Subsection 2.5.2, intended purpose is defined using the combination of 3 concepts: context of use, scope, and the Sustainable Development Goals (SDGs) to which the use case contributes.

2.2.4 Registration Requirements

The AI Act registration regime has been touched upon in broader analyses of the Act, including in [76, 8]. However, the eligibility criteria and the information requirements for registration under the AI Act have not been examined in the literature yet.

2.3 AI Risk Taxonomies and Ontologies

With the risk of AI in the spotlight, there has been a surge in studies that identify, analyse, and classify AI harms. Consequently, there exist several taxonomies of AI risks that have been developed adopting various perspectives, from technical to social to regulatory, and appeal to specific domains, e.g. health, or particular AI technologies, such as LLMs. By offering an overview of the existing work, this section recognises the lack of formal taxonomies and ontologies for expressing AI systems and their risks.

2.3.1 Generic Taxonomies for AI and its Risks

With the rise of incidents caused by AI, there have been multiple initiatives focused on reporting these incidents in repositories. To serve a diverse

set of stakeholders, from AI developers to policy makers to the general public, such repositories are required to provide incidents in a structured and annotated format. Therefore, many of existing taxonomies for describing AI and its risks have grown out of the need to annotate AI incidents. This subsection reviews the most prominent taxonomies derived from or associated with AI incident repositories.

The **OECD framework for classification of AI systems** [94], published in 2022, is a tool for assessing potential risks and benefits of AI use cases by considering five high-level dimensions: people & planet, economic context, data & input, AI model, and task & output, as depicted in Figure 2.1. The framework incorporates taxonomies for its risk assessment criteria. AI Incidents Monitor (AIM)[13] is an OECD initiative aiming to track *actual* AI incidents by automatically annotating and classifying them using machine learning models according to their country, industry, related AI principle, type of harm, severity of harm, and affected stakeholders. Developing a common framework for reporting AI incidents is on the OECD's agenda for future work[14].

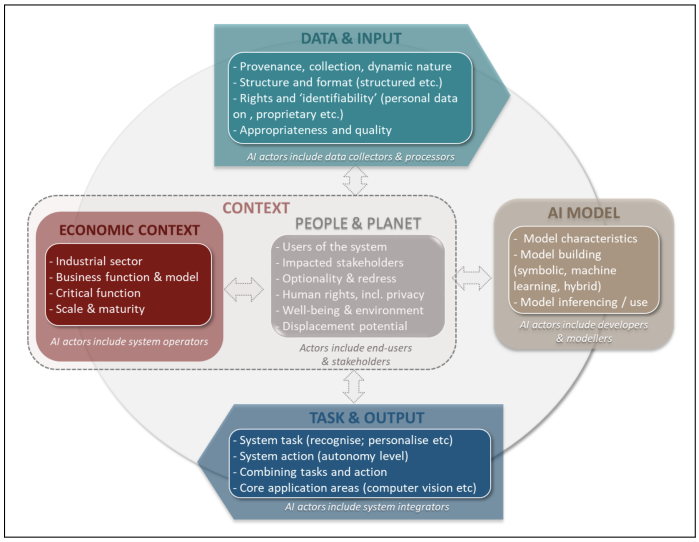


Figure 2.1: OECD's framework for classification of AI systems [94]

The **AI, Algorithmic, and Automation Incidents and Controversies (AIAAIC)** repository[15], which is an open-access dataset of more than 1500 AI incidents covered by the media, uses a taxonomy for annotating the incidents. The set of concepts for incident annotation includes

[13] https://oecd.ai/en/incidents
[14] https://oecd.ai/en/incidents-methodology
[15] https://www.aiaaic.org/aiaaic-repository

categories of sectors, technologies, purposes, and impacts of AI on individuals, society, environment, and providers[16]. In a recent project, an extension of this taxonomy that focuses only on harms is being developed by the AIAAIC community. The **AI, algorithmic and automation harms taxonomy** [68], which has been made available in 2024, classifies AI harms into 9 top-level categories and 69 sub-categories, illustrated in Figure 2.2. Although machine-readability is stated as an objectives of the AIAAIC taxonomy, no such representation is provided at the time of writing[17].

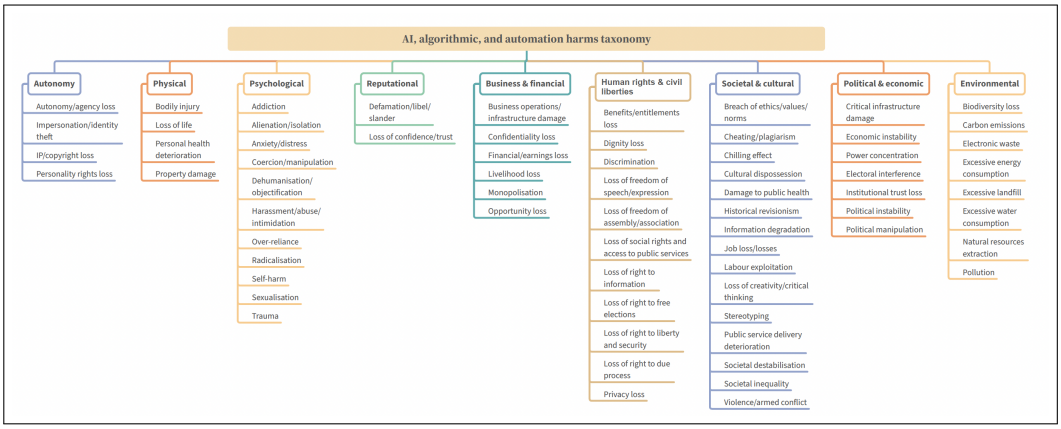


Figure 2.2: The AIAAIC's harm taxonomy [68]

The Center for Security and Emerging Technology (CSET) AI harm framework [95], published in 2023, is a generic customisable framework that has been developed based upon annotating AI incidents through a collaborative process. In its classification of harms, at the top-level it considers the following types of harms: *tangible* and *intangible*, and distinguishes between the *actual* and *reasonable potential* harm, referring to them as harm event and harm issue, respectively. Enabling annotation of incidents, the framework also includes a component to describe harmed entities and requires the harm to be linked to an AI system and the impacts on the harmed entity. The CSET AI Harm Framework was customised for annotating **AI incident database (AIID)**[18] [96] with a fine-grained, yet not formal, schema provided in [97].

In addition to the customised CSET taxonomy, **Goals, Methods, and Failures (GMF) taxonomy** [98], published in 2023, has been developed

[16] https://www.aiaaic.org/aiaaic-repository/classifications-and-definitions, visited on 01/08/2024

[17] The author has been involved in the development of the harms taxonomy and advocated the use of machine-readable formats for representation.

[18] https://incidentdatabase.ai/

based on AIID and used for its annotation. The taxonomy, as its name implies, considers three aspects of an AI incident: the AI system's goals, the technologies used within the system, and technical causes of the incident. The taxonomy has been further populated in a bottom-up approach through manual annotation of AIID incidents.

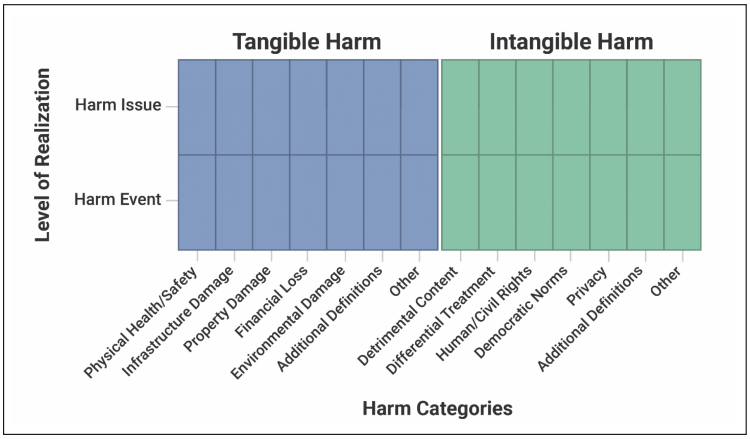
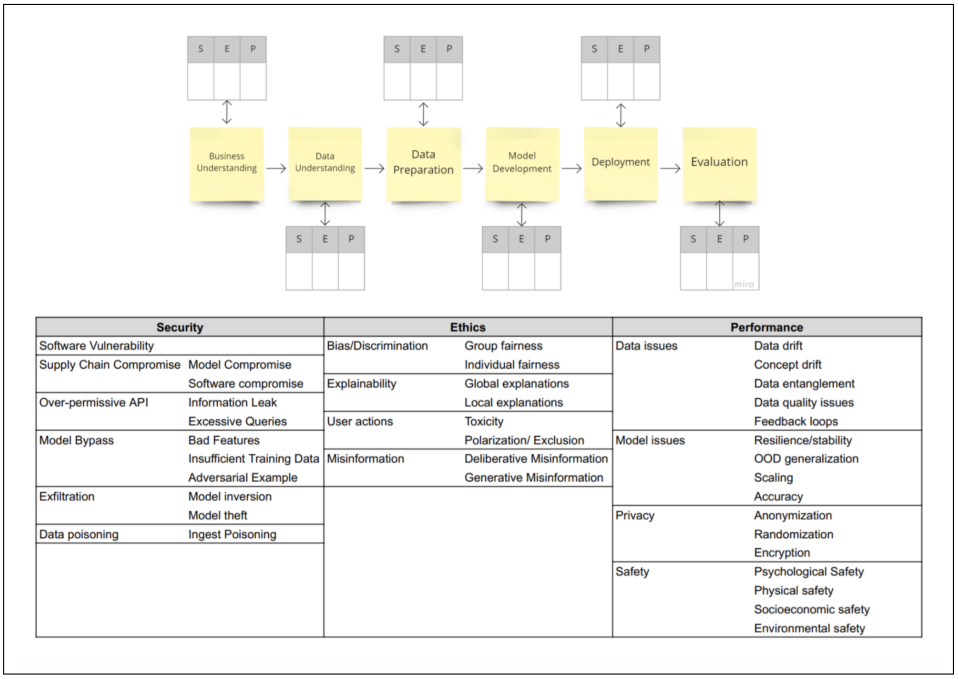


Figure 2.3: An overview of CSET's AI harms framework [95]

AITopics[19] [99] is the AAAI's (Association for the Advancement of Artificial Intelligence) corpus of AI-related news stories, research articles, conferences, and journals. The scope of AITopics is not limited to AI incidents and therefore it indexes all types of AI-related news articles as well as scientific papers. Discovery, categorisation (determining the main focus), and summarisation of AI news featured in AITopics have been automated [100].

The **taxonomy of AI Vulnerability Database (AVID)**[20] [101] considers 3 dimension of risks, which are security, ethics, and performance, across the machine learning (ML) development lifecycle, as shown in Figure 2.4. The notable advantage of the AVID taxonomy is the JSON schema, providing better interoperability and extensibility compared to the other taxonomies discussed in this subsection. The AVID data model is also provided as a part of a Python toolkit[21], enabling the creation of reports using the schema.

[19] https://aitopics.org/
[20] https://avidml.org/database/
[21] https://docs.avidml.org/developer-tools/python-sdk



Security		Ethics		Performance	
Software Vulnerability		Bias/Discrimination	Group fairness	Data issues	Data drift
Supply Chain Compromise	Model Compromise		Individual fairness		Concept drift
	Software compromise	Explainability	Global explanations		Data entanglement
Over-permissive API	Information Leak		Local explanations		Data quality issues
	Excessive Queries	User actions	Toxicity		Feedback loops
Model Bypass	Bad Features		Polarization/ Exclusion	Model issues	Resilience/stability
	Insufficient Training Data	Misinformation	Deliberative Misinformation		OOD generalization
	Adversarial Example		Generative Misinformation		Scaling
Exfiltration	Model inversion				Accuracy
	Model theft			Privacy	Anonymization
Data poisoning	Ingest Poisoning				Randomization
					Encryption
				Safety	Psychological Safety
					Physical safety
					Socioeconomic safety
					Environmental safety

Figure 2.4: The two-aspect AVID taxonomy [101]

Table 2.1 provides a comparison of the reviewed taxonomies based on the following criteria:

1. What is the scope of the taxonomy in respect to AI?
2. Based on which resources the taxonomy has been developed?
3. What is the taxonomy development approach (e.g. top-down or bottom-up) and what is the level of human involvement and automation (e.g. automated, semi-automated, manual)?
4. In which formats the taxonomy is made accessible?
5. What are the *key* concepts related to the technical aspects of AI?
6. What are the *key* concepts related to the use of AI?
7. What are the *key* concepts related to AI risks?

In this comparison, the AIAAIC taxonomy represents the harm taxonomy in combination with its classifications for incident annotation. Also, CSET's AI harm taxonomy for AIID [97] is considered in this comparative analysis, given its high level of detail.

As shown in Table 2.1, multiple terminologies and structures have been used for AI incident annotation. This discrepancy is not necessarily a drawback, on the contrary it can be viewed as an indicator of the presence of diverse viewpoints required for addressing AI harms. However, the lack of structured data formats is a significant barrier to comparison of incidents and issues across different repositories. Further, lack of such specifications makes it difficult to align and integrate the state of the art AI risk taxonomies to create comprehensive taxonomies.

Table 2.1: Comparison of existing generic AI risk taxonomies

Taxonomy	Scope	Main resource	Development approach	Format	Technology concepts	AI use concepts	AI risk concepts
OECD [94]	AI	Related research	Unknown	Unstructured	Application area, AI system task	User, Industrial sector, Business function	Impacted stakeholders, Impact, Redress
AIAAIC [68]	AI	News articles	Bottom-up, collaborative	CSV	Technology	Sector, Purpose	Harm
CSET [97]	AI	News articles	Bottom-up, collaborative, manual annotation	Unstructured schema	AI functionality & techniques, Involved technology	Incident domain, Environmental & temporal characteristics	Harm Assessment, Harmed entity
GMF [98]	AI	News articles	Bottom-up, manual annotation	Unstructured	Method and technology	Goal	Failure cause
AITopics [99]	AI	Online resources	Bottom-up, automated discovery & annotation	Unstructured	Technology	Industry	None
AVID [101]	ML	Unknown	Unknown	JSON schema	Lifecycle stage	None	Effect

2.3.2 Specific AI Risk Taxonomies

A highly active area of research is currently on taxonomies of AI risks with a scope limited to the type of AI system, e.g. LLMs, or the domain of use, e.g. health. The common approach in dissemination of these taxonomies is through research papers, without any standardised or structured format. Given that this thesis pursues development of horizontal ontologies, without aiming to be exhaustive, this subsection only refers to some promising specific risk taxonomies, which are: Weidinger et al.'s taxonomy of ethical and social risks of LLMs [102], Tanaka et al.'s taxonomy for risks of generative AI systems [103], Lee et al.'s taxonomy of AI privacy risks [104], the Open Loop's taxonomy of potential harms associated with machine learning applications and automated decision-making systems [105], NIST's taxonomy for adversarial machine learning [106] and categories of AI Bias [107], Steimers and Schneider's work on taxonomy of risk sources that impact AI trustworthiness [108], and Roselli et al.'s work on classification of AI bias [109].

2.3.3 Ontologies Related to AI Risks

At the time of writing, there is an absence of ontologies that specifically provide a holistic model of "AI risks". However, there are state of the art ontologies that address trustworthy AI issues. It should be noted that these ontologies were reviewed as a part of the ontology development process (Section 4.1) to identify ontologies that have the potential to be reused.

The **EA-Ontology**[22] [110] has been designed for ethical assessment of emerging technologies, including AI, based on existing approaches for ethical analysis. Within its module for modelling *levels of ethical analysis*, the concept of `ateo:Risk` enables modelling risks identified during the ethical assessment and further linking them to ethical issues. The compound concept of `Use Of Technological Artefact` also allows modelling the context of use. Moreover, the EA-ontology provides a taxonomy of ethical issues by providing 4 top-level categories: `Harm And Risk`, `Justice`, `Right`, `Well-being And Common Good`. However, the classification of ethical issues seems disjoint from the taxonomy, since no relations link the issues to other concepts within the ontology.

The **Ai System use Case Explanation oNTology (ASCENT)** [111] is a framework for describing explainable AI measures. The framework includes 3 aspects: (i) technical information about the system that potentially impact explain AI solutions, (ii) information regarding the context of use that is related to explainability requirements, and (iii) the characteristics of

[22] https://protect.oeg.fi.upm.es/eaontology/eaontology_widoco/index-en.html

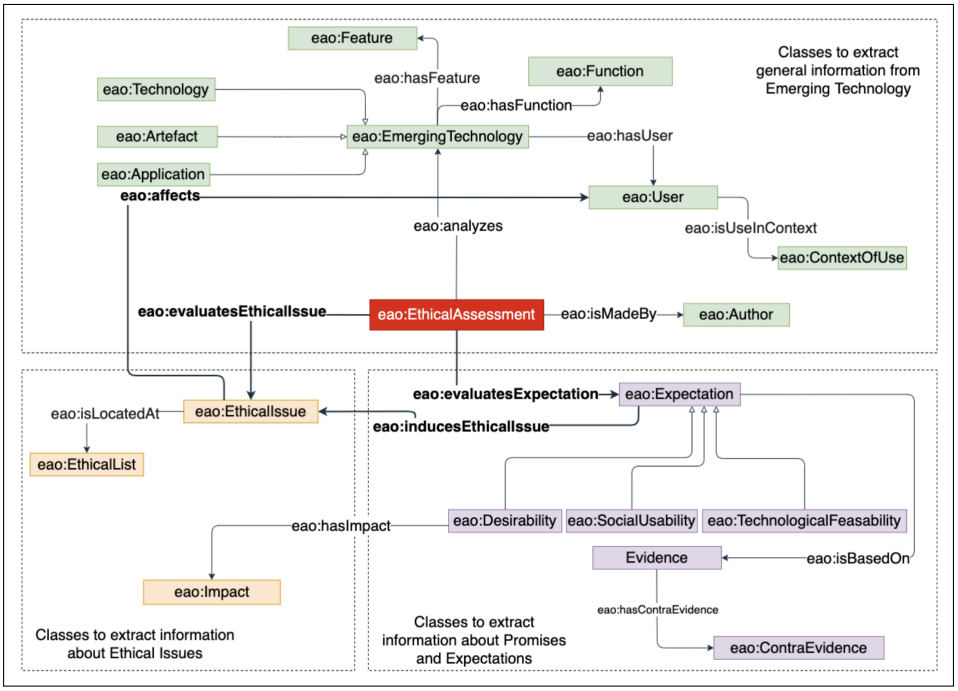


Figure 2.5: An overview of EA-ontology [110]

the explanation measures. Within the use case module, a `Risk Level` (high risk or low risk) is assigned to a task for which the AI system is used, as an indicator of the significance of consequences. The implementation of the ontology is provided as an open resource. However, definitions for concepts are not available in the model.

ExplainableMLOntology [112] is an ontology aiming to enhance explainability of ML models through a process-oriented viewpoint. The ontology focuses on expressing processes used in the ML development and explanation and includes 3 modules for representing general ML, ML classification, and explanation.

FIDES ontology[23] [113] is a minimal semantic model of accountability-related information about statistical ML models. FIDES supports expressing ML models and their data across different lifecycle phases, including development, deployment, and execution. The ontology is accompanied by an automated tool for generating a knowledge graph from the information provided regarding an ML model.

[23] https://w3id.org/fides

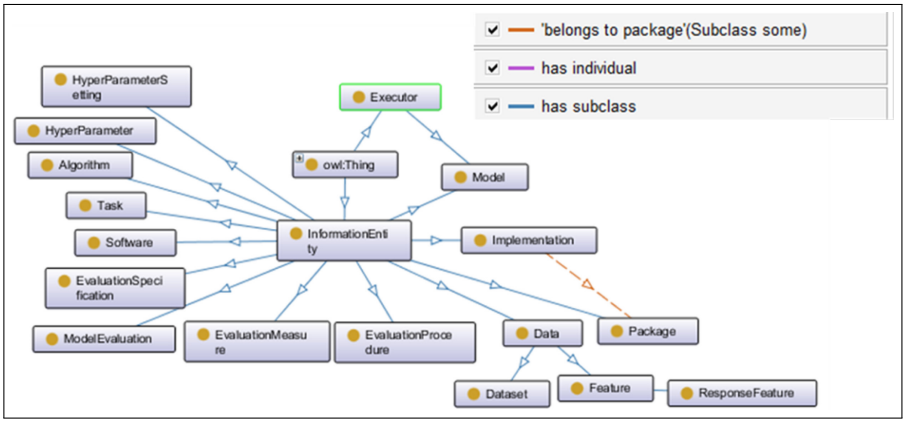


Figure 2.6: An overview of FIDES ontology [113]

The **Doc-BiasO Ontology** [114] has been proposed for expressing and documenting biases across ML/AI workflows. Inspired by the contribution presented in this thesis, Doc-BiasO comprises a high-level ontology describing the core concepts and relations (shown in Figure 2.7), with a controlled vocabulary extending the ontology (which is underdevelopment). Doc-BiasO Ontology reuses contributions of this thesis (AIRO and VAIR). But as the implementation and documentation of the ontology is not published online yet, the extent of reuse is not clear. Moreover, the application of the ontology in generation of bias documentation is not demonstrated.

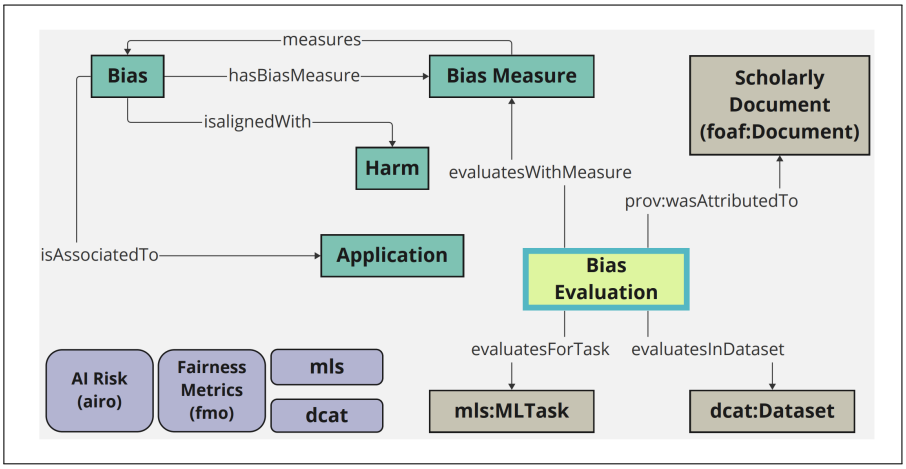


Figure 2.7: An overview of Doc-BiasO [114]

The **Artificial Intelligence Ontology (AIO)**[24] [115] has been developed using a hybrid approach which consists of manual ontology development and automated processes empowered by the capabilities of LLMs to assist with the population of the ontology. AIO consists of 444 concepts related to AI, grouped into 6 modules: networks, layers, functions, LLMs, preprocessing, and bias. Despite being called an ontology, it lacks object properties to model the relations between concepts.

The **ontology for standardising trustworthy AI** [75] is an ontology for representing and mapping trustworthy AI concepts from different emerging AI standards. The ontology provides a way to express activities related to manifestation of AI trustworthiness. It also provides a way to depict the influence of entities, activities, and agents on AI trustworthiness, and captures the role of stakeholders in disclosing and exhibiting trustworthiness of AI through its characteristics. The work uses activities from (the draft of) ISO/IEC 42001 on AI management systems, wherein risk management is a key activity. As claimed by the authors, the ontology has the potential to be extended to represent AI risks and treatments from relevant standards. Enhancing this ontology, the **Trustworthy AI Requirements Ontology (TAIR)**[25] [116] has been developed for mapping AI regulations and standards. As shown in Figure 2.8, `Requirement` is the central element in TAIR, which can be linked to the entities, activities, and agents involved in its implementation. It further supports modelling the relationships between requirements extracted from different AI documents. An example of such relations is `statisfiedBy`, which enables aligning AI regulations and standards through the obligations and requirements they impose. This is particularly important in showing how high-level obligations of the EU AI Act can be addressed by satisfying requirements from standards or common specifications. This specific application of TAIR is showcased through modelling and mapping of sets of concepts and requirements from the EU AI Act and ISO/IEC 42001[26].

The **Ethical AI principles ontology (AIPO)** [117] is a semantic model for trustworthy AI principles emerged from different guidelines and policies. The ontology models principles as `skos:Concept`, which enables relating and matching principles through leveraging SKOS relations. Within AIPO, trustworthy AI documents are modelled as `dcat:Resource`, allowing modelling the details of the document from which the principles are extracted.

[24] https://bioportal.bioontology.org/ontologies/AIO

[25] https://tair.adaptcentre.ie/documentation/tair_documentation.html

[26] The author has been involved in development of TAIR and has provided inputs that facilitated identification of the concepts and requirements, that are in the scope of this thesis.

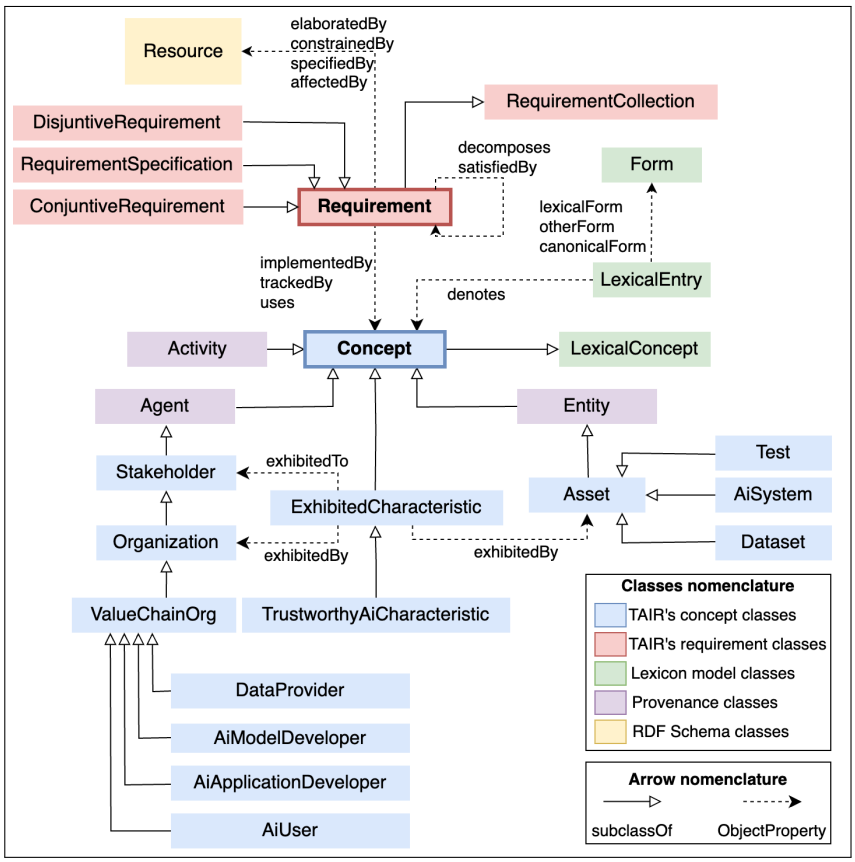


Figure 2.8: An overview of TAIR ontology [116]

Table 2.2 provides a comparison of the aforementioned ontologies. Within this table, and the subsequent ones, a black circle (•) indicates presence and a blank circle (○) signifies absence. Additionally, the pink bullets provide links to the corresponding encoding or documentation.

Table 2.2: Comparison of ontologies related to AI risks (a black circle (•) indicates presence and a blank circle (○) indicates absence)

Ontology	Scope	AI technology	AI use	AI risk	Open-source encoding	Documentation	Peer-reviewed	Proof-of-concept
EA-Ontology [110]	Technology ethics assessment	○	•	Ethical risks	•	•	•	•
ASCENT [111]	AI explainability	○	•	Measures related to explainable AI	•	○	○	•
ExplainableML Ontology [112]	AI explainability	•	○	Measures related to explainable AI	•	○	•	•
FIDES [113]	Accountability of statistical ML models	•	○	○	•	•	•	•
Doc-BiasO Ontology [114]	AI Bias documentation	•	•	AI bias & its measures	○	○	•	○
AIO [115]	AI terminology	•	•	Types of bias	•	•	○	○
TAIR [116]	Alignment of trustworthy AI requirements	○	○	○	•	•	•	•
AIPO [117]	Alignment of trustworthy AI principles	○	○	○	•	○	○	•

2.3.4 Generic Risk Ontologies

Risk management is a vast and rich field of knowledge and there exist several ontologies to facilitate risk management in different areas, ranging across construction to software development to health. Some examples are Masso et al.'s SRMO (Software Risk Management Ontology) [118], Hayes' ontology for modelling risk associated with online disclosure of personal information [119], McKenna et al.'s work on the Access Knowledge Risk (ARK) platform which employs SKOS data models to enable risk analysis, risk evidence collection, and risk data integration in socio-technical systems [120].

In alignment with the scope of this research and with the intention to identify resources for reuse in ontology development (Section 4.1), in the following generic risk ontologies that use international standards resources or the AI Act are mentioned. The common ontology of value and risk [121] describes risk by associating it to the concept of value. A formal ontology for ISO/IEC 27005 standard on information security risk management is also proposed in [122]. In terms of conceptual modelling, both ontologies are useful resources, however they have been presented without any implementations. The Data Privacy Vocabulary (DPV 2.0) [57] provides a model of fundamental risk concepts based on ISO 31000 series of standards on risk management, which is accompanied by a minimal set of risk assessment concepts in its RISK extension[27]. While it provides an open model for expressing risks, it is not aligned with the requirements of the AI Act yet[28]. This, however, is expected given the recent publication of the AI Act.

2.4 Approaches Related to Implementation and Enforcement of the AI Act

2.4.1 Approaches for Determining Risk Level as per the AI Act's Classification

There is an absence of approaches and tools for assessing risk category under the AI Act, which is justifiable given that the Act published recently in July 2024. Similarly, within the regulatory landscape, there are no authoritative guidelines available on this matter, as of August 2024. Despite this absence, tools to assist with determining the risk category under the AI Act have been emerging. Although these are online tools that neither utilise

[27] https://w3id.org/dpv/risk

[28] The author has been actively involved in development of DPV and its risk extension.

Semantic Web technologies nor are they accompanied by peer-reviewed papers, their review is necessary to capture the current state of affairs. The following reviews some of these tools with a focus on the criteria they establish for determining prohibited (Article 5) and high-risk AI systems (Annex I and III).

Future of Life Institute (FLI)'s EU AI Act Compliance Checker[29] is a tool for identifying the risk level and obligations associated with an AI system under the AI Act. Concerning the prohibited systems, the tool only relies on the function selected by the user; examples of functions listed in the tool are biometric categorisation, exploiting vulnerabilities, social scoring, and subliminal techniques, manipulation, and deception. For determining high-risk systems, including both Annex I and Annex III, the tool is primarily based on the domain in which the system is used. Additionally, to identify the non-high-risk systems that meet Annex III high-risk, the tool asks the user to indicate whether "the AI system pose a significant risk of harm to the health, safety or fundamental rights of any person". The tool's output represents the risk category, without referring to specific clauses on the basis of which the system is classified as prohibited or high-risk. The tool also outlines related obligations according to the type of the AI Act actor, e.g. provider, as selected by the user.

Holistic AI's **EU AI Act Risk Calculator**[30] includes lists of Article 5's prohibited clauses and Annex III's high-level areas for user to select from. The output indicate only the risk level without any additional information. The user must provide contact information to access the assessment.

EIT community—the initiative of European Institute of Innovation and Technology co-funded by the EU— provides an online **EU AI Act Conformity Check**[31]. For prohibited AI systems, the tool presents each clause of Article 5 as a question. For Annex I high-risk AI, a list of regulated domains is provided from which the user should select. For Annex III high-risk AI, the user first needs to select a domain and then a list of respective objectives, stating the points under that domain in Annex III, is shown. The output indicates the risk category and a document that includes guiding materials. The output is only shown when the user provides contact information.

The source code of none of the aforementioned tools is made publicly available. Further, none of them signal adoption of Semantic Web technologies. More importantly, they only repeat the text of the AI Act without breaking down the risk categories into conditions expressed using fine-grained

[29] https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/

[30] https://www.euaiact.com/eu-ai-act-risk-calculator

[31] https://statworx.typeform.com/to/Y9MdIGDP?typeform-source=ai.eitcommunity.eu

concepts.

A closely related work to the contributions of this thesis is Hanif et al.'s **Decision-Tree-based framework** [123, 124] which aims to help individuals, from different backgrounds, classify AI systems as per the AI Act. This static framework employs a decision tree including 20 questions for determining the risk category—that can be unacceptable, high, limited, and minimal. These questions fall under 4 overall themes: protected values, objective/intention, domain, and use-case/technology.

Table 2.3: Comparison of existing risk classifiers for the AI Act

Work	Prohibited	Annex I high-risk	Annex III high-risk	Open source	Web app	Personal info needed
Future of Life Institute (FLI)'s Compliance Checker	•	•	•	○	•	○
Holistic AI's Risk Calculator	•	•	•	○	•	•
EIT community's Conformity Check	•	•	•	○	•	•
Decision-Tree-based framework [123, 124]	•	•	•	•	○	N/A

2.4.2 Semantic Modelling of Policies Related to Legal Compliance

For cooperation of actors across the AI value chain, the AI Act relies on contractual arrangements (see for example Recital (90)). As will be discussed in Section 3.3, this thesis proposes declaring intended purpose as use policies. In terms of declaring policies, within the landscape of Semantic Web technologies, the Open Digital Rights Language (ODRL) [125, 53] is a key W3C recommendation that provides a formal semantic language for declaring policies. The ODRL information model (Figure 2.9), along with its standardised vocabulary, enables the expression of policies concerning actions over assets. Extension of ODRL is enabled through its Profile Mechanism.

In the context of the AI Act, there is no research investigating open models for expressing use policies or contractual terms and agreements. However, of relevance to RO3(b) is the existing research on the use of ODRL for compliance with the EU GDPR [21]. In this context, ODRL has been used

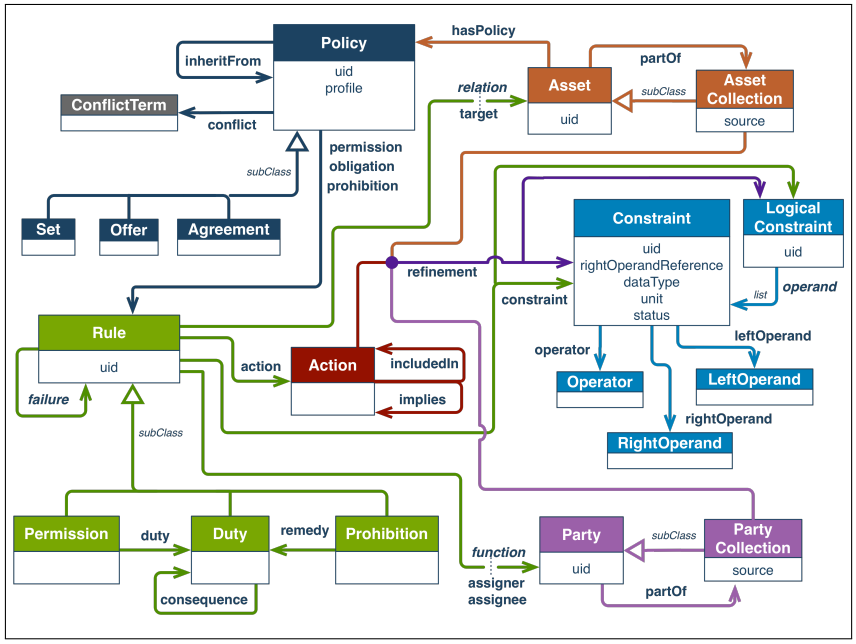


Figure 2.9: ODRL information model [125]

in particular for automated checking of consent permissions [126], expressing legal obligations [127], modelling the obligations in terms of permissions and prohibitions regarding executing business processes [128], and expressing privacy policies in the digital fashion domain [129]. In the context of data governance, ODRL has been extended for expressing policies related to access control to data stored in Solid Pods [126] and utilised for modelling policies associated with responsible use of genomics data [130], expressing contracts for research data governance [131], and expressing data spaces' usage and access control policies [132].

2.4.3 Machine-Readable AI and Risk Documentation

Generally, AI documentation approaches are acknowledged as instruments for improving transparency, and in turn enhancing trustworthiness. However, there has been little attention to the barriers to the generation, maintenance, assessment, and exchange of conventional text-based documentation. Providing machine-readable specifications is an idea taken up by some recent work to support adaptable and interoperable documentation. This subsection aims to review **machine-readable** AI, data, or model documentation frameworks. Later in Section 2.5, a review of existing documentation approaches that can be used in addressing the requirements of

the AI Act will be presented.

Open DataSheets[32] [133] provides a metadata framework for representing information elements included in Datasheets for Datasets [134] and further linking them to responsible AI principles. Similarly, **Model Card Report Ontology (MCRO)**[33] [135] offers the metadata for the content of Model Card reports. **Linked Model and Data Cards (LMDC)** [136] presents a schema for integration of Model and Data Cards in a data space to provide a holistic view of a model or an AI service. The AI usage Card [137], that will be discussed in Subsection 2.5.2, has been made available in JSON format, however no semantic model is presented.

Concerning documentation requirements of the EU digital regulations, **DPV**[34] [138], which is developed by W3C Data Privacy Vocabularies and Controls Community Group (DPVCG), has been applied to generate documentation required for compliance with the EU GDPR, such as Data Protection Impact Assessment (DPIA) [139], data breach reports [140], and Register of Processing Activities (ROPA) [141].

Table 2.4: Comparison of existing approaches for machine-readable documentation of AI, data, or models

Work	Scope	License	Peer-review
Open DataSheets [133]	Dataset	MIT License	○
MCRO [135]	Model	Attribution 3.0 Unported	•
LMDC [136]	Dataset & model	Apache License 2.0	•
DPV [138] for GDPR documentation	DPIA & ROPA & data breach	W3C Software and Document license	•

2.4.4 Machine-Readable Catalogues of AI, Model, and Datasets

With the proliferation of AI models, systems, and use cases, open AI repositories and commercial marketplaces have been created to facilitate the discovery and sharing of resources [142]. This subsection investigates the literature to identify studies related to RO3(c), which addresses registering

[32] https://github.com/microsoft/opendatasheets-framework
[33] https://github.com/UTHealth-Ontology/MCRO
[34] https://w3id.org/dpv/

and sharing metadata about AI systems and their risks, in particular within the EU database of high-risk AI systems, using Semantic Web technologies.

Currently, there are a few well-known repositories that leverage metadata for describing resources. The **Hugging Face Hub**[35] is a centralised repository of open-source models and datasets, wherein each model or dataset is accompanied by metadata describing it. This enables discovery, sharing, and filtering of resources available on Hugging Face's Model and Data Hubs through the use of open JSON-based metadata. The Hub contains a repository of Spaces, i.e. ML demo apps, which, unlike the Model and Data Hub, do not support the inclusion of documented information and structured metadata. Similarly, **Kaggle** provides repositories of Datasets[36] and Models[37], where datasets, models, and generative AI applications are indexed and documented using detailed Data and Model Cards. Using the Kaggle repository, datasets and models can be published, shared, tagged, searched, and sorted. Compared to Hugging Face Data Hub which supports indexing only open-source resources, Kaggle Datasets allows for sharing metadata about both proprietary and publicly available datasets. The **AI-on-Demand (AIoD) platform**[38] is a European-funded project that serves as a community-driven hub for cataloguing AI-related solutions and components that contribute to the European ecosystem of AI excellence and trust. AIoD's asset catalogue[39] covers a wide range of resources including datasets, libraries, ML models, AI services, tools, use cases, and even tutorials. AIoD also provides JSON-based metadata for describing resources[40].

Croissant [143] is a framework developed by MLCommons—a non-profit open AI engineering consortium that enables expressing metadata for datasets with a focus on information that is essential in machine learning workflows. The Croissant vocabulary[41] is an extension of `schema.org/Dataset` vocabulary for metadata of ML datasets, which is expressed in JSON-LD. The Croissant framework is supported by a user-friendly tool to assist non-technical users in creation and modification of metadata. Although it is not a dataset repository, it has been integrated with existing data repositories, including HuggingFace, adding a layer of metadata.

While the information in the aforementioned registries is mostly presented in semi-structured formats such as JSON, none of them follow standardised

[35] https://huggingface.co/docs/hub/index
[36] https://www.kaggle.com/datasets
[37] https://www.kaggle.com/models
[38] https://aiod.eu/
[39] https://www.ai4europe.eu/research/ai-catalog
[40] https://api.aiod.eu/redoc
[41] https://docs.mlcommons.org/croissant/docs/croissant-spec.html

approaches for data sharing or cataloguing.

In regard to standardised approaches, the Data Catalog Vocabulary (DCAT) [54]—the W3C's recommended vocabulary for publishing data catalogues—and particularly its application profile for data portals in Europe (DCAT-AP) [144] have been adopted by the European Commission to promote open, standardised, and interoperable data sharing, prominently in the European Data Portal (EDP)[42], which is the central point of access to open data provided by the EU's public agencies [145]. Recently, **MLDCAT-AP** [146] has been proposed as an extension of DCAT-AP for including information about machine learning models in data catalogues. One of the distinguishing features of MLDCAT-AP is inclusion of information about *risks* associated with ML models.

Of relevance to the contributions of this thesis is the **Data Processing Catalogue (DPCat)** [141], which is an extension of DCAT and DCAT-AP that enables representing, maintaining, and exchanging ROPA-related information in the form of datasets and catalogues. DPCat further enables creating documentation to address the GDPR's ROPA requirements.

Table 2.5 shows a comparison of existing approaches for cataloguing AI, models, and datasets. Currently, providing metadata, typically in JSON format, regarding datasets and models is an established practice. However, there is little attention to cataloguing AI systems and consequently there is no standardised machine-readable vocabulary that supports cataloguing of not only AI systems but their incorporating components.

2.5 AI and Risk Documentation Approaches for the AI Act

2.5.1 Alignment of AI, Model, and Data Documentation Approaches with the AI Act

Existing documentation approaches, such as Datasheets for Datasets [134] and Model Cards [147], have become de-facto practices for documenting and sharing information regarding datasets and AI models. With the emergence of the AI Act, a key question is the extent to which they could be leveraged to address the Act's requirements in regard to AI and risk documentation. This question has been investigated by the studies mentioned in the following. Pistilli et al. discuss the potential of Model Cards as a compliance tool and anticipate its adoption—among other existing documentation prac-

[42] https://data.europa.eu/en

Table 2.5: Comparison of AI cataloguing approaches

Work	Scope	Format of metadata	Use of standardised vocabularies
Repositories			
Hugging Face Data/-Model Hub	Dataset/Model	JSON	○
Kaggle dataset/-model repository	Dataset/Model	HTML	○
AI-on-Demand (AIoD) platform	AI assets (dataset, model, services)	JSON	○
Approaches			
Croissant [143]	Dataset	JSON-LD	○
MLDCAT-AP [146]	ML models	JSON-LD	•
DPCat [141]	GDPR's ROPA	Turtle	•

tices originated from the AI community—for compliance with the AI Act's documentation obligations [148]. The work by Hupont et al. [93], which investigates the 6 most widely-used AI and data documentation approaches for their alignment with the implementation of documentation provisions, concludes that AI Factsheets [149] offers a higher overall degree of information coverage, followed by Model Cards [147] and the AI Classification Framework proposed by OECD [94]. The research also demonstrates that while data-related information elements are well-covered by most documentation approaches, particularly Datasheets for Datasets [134], the Dataset Nutrition Label [150], and the Accountability for Machine Learning framework [151], they fail to fully cover technical information requirements related to AI systems. This finding is further strengthened in a follow-up comparative analysis of 36 AI system, model, and/or dataset documentation practices, which demonstrates the overall alignment of documentation practices with transparency requirements of the EU AI Act and other recent EU data and AI initiatives, and spots a gap in representing the information related to AI systems in its entirety and its context of use [152].

Given the strong emphasis on the intended use of AI and the central role of the risk management documentation within the AI Act, the remainder of this section focuses only on frameworks for documenting *AI uses* and *AI risks*.

2.5.2 AI Use Documentation Approaches

When it comes to AI systems, it has been increasingly recognised that context matters. The AI Act's high-risk AI classification in Annex III is a reflection of the fact that how and where AI is used significantly affects the types and severity of risks it imposes. Documentation of AI uses also is desired by the AI community [153], however in practice AI use documentation approaches are not common [154]. This subsection reviews a few recent documentation approaches focused on AI use.

Use Case Cards [64] extends UML use case diagram and its respective documentation to present the *intended purpose* of AI systems, aligned with the requirements of the EU AI Act. The Use Case Card's diagram distinguishes between those behaviours of system that are implemented by AI and non-AI use cases. In addition to the common fields in use case documentation such as the flow of the use case, Use Case Cards' template contains information regarding context of use, scope, and application area of the AI system. The template also includes a field to indicate the product(s) in which the AI system is used. The information provided within a Use Case Card aims to facilitate determining the risk level of AI systems according to the AI Act's risk classification. The Use Case Cards has been developed by researchers based in JRC and validated by EU policymakers.

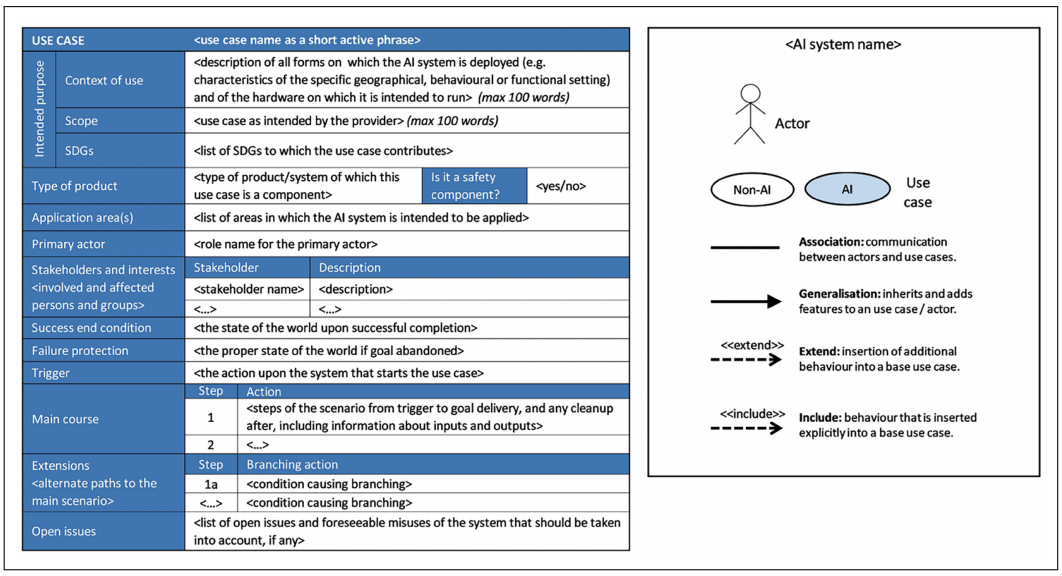


Figure 2.10: An overview of Use Case Cards [64]

AI Usage Cards [137] provides an approach for documenting use of AI, specifically in academic research processes. The documentation is therefore

focused on reporting content generated using AI in different phases of research such as literature review, data collection, writing, and coding. One of the key features of the AI Usage Cards is inclusion of risk management measures applied to eliminate or mitigate risks.

2.5.3 AI Risk Documentation Approaches

The survey conducted by Königstorfer and Thalmann reveals that AI documentation should "support risk management" and has to "ensure that the AI can be used safely" [153]. Therefore, risks and ethical considerations are key in AI documentation. Although many existing documentation approaches include *potential* risk and misuses, disclosing *risk assessments* and *actual measures* in place to address them is not a common practice. In the following, the recently-proposed approaches for AI risk documentation are reviewed.

The **AI Impact Assessment Report Template** [155] is basically a documentation framework centred around use of an AI system and its risks for AI impact assessments required by the EU AI Act, NIST's AI RMF [2], and ISO/IEC 42001 [37] on AI management systems. The template consists of 5 sections wherein the system's use, risks, mitigation measures, benefits, and risk reporting information are described. The template uses the contributions of this work regarding the set of concepts for describing intended purpose (use), which will be discussed in Subsection 3.2.1. This documentation template has been developed through a co-design process with involvement of AI practitioners and compliance experts.

RISKCARDS [156] is focused on documenting risks specifically associated with language models. RISKCARDS breaks down information about a risk into the following: details of context and application of the language model, type of harm as per existing *taxonomies* of risks and harms, impacted entities, pre-conditions required for a harm to be materialised, and sample prompts with the corresponding language model output. In terms of limitations, the RISKCARDS misses the measures to eliminate or mitigate the harm.

AI Risk Profiles [157] provides a two-fold template for documenting and reporting risk of AI use cases in both detailed and summarised forms. AI Risk Profiles has been built upon risk scenarios to provide information about risk management including risks, impacts, and the related mitigation measures. In addition, it incorporates information regarding the use case and lists the regulations and standards the system conforms to. To facilitate creation of the Risk Profiles and further assist with AI risk assessments, the documentation framework is accompanied by a high-level taxonomy of AI

risks, which provides 9 categories of risks.

In the context of the AI Act, a noticeable work directly related to risk management and technical documentation requirement of the Act is the tool proposed for verification and documentation of quality management systems [7]. This tool focuses on risk and data management requirements (Article 9 and Article 11), as components of an AI quality management system. The tool creates a technical documentation, which partially covers the requirements of the Act. This documentation is created utilising the user's input concerning the system's qualities, e.g. performance and explainability, and its risks. Notably, the tool's risk management system component uses the contributions of this work in regard to the 5 concepts for determining risk level, which will be presented in Subsection 3.2.1.

2.5.4 Comparison of AI Use and Risk Documentation Approaches

Table 2.6 provides a comparison of the reviewed AI use and risk documentation frameworks on the grounds of following criteria, which are based on the documentation requirements of the AI Act (refer to Subsection 3.4.2):

1. What is the scope of documentation? This criterion reflects whether the documentation provides a holistic view of an AI system, as required by the AI Act, or only incorporates information regarding specific components or aspects of the system.

2. Does the documentation include technical information regrading the system? This criterion aims to determine whether technical specifications of the AI system are included, as required by technical documentation of the AI Act.

3. Does the documentation include information about the context wherein the system is used? This criterion aims to determine whether information about the use of the system, also known as *intended purpose* by the Act, is included.

4. Does the documentation provide information related to risk management system, e.g. risks, impacts, and control measures? This is to ensure that the documentation approach is transparent regarding risk management and supports risk management documentation, as required by the AI Act, Article 9.

5. Were the requirements of the AI Act considered in the design of the documentation framework? Or has its alignment with the AI Act been explored?

6. Is the documentation represented in a structured machine-readable format? This criterion addresses the use of semantic models and also hints at the potential of the documentation framework to support AI governance and RegTech tools.

7. Does the documentation include a summarised view? This criterion aims to reflect whether the documentation framework considers stakeholders with non-technical knowledge.

Table 2.6: Comparison of AI use case and risk documentation approaches (a black circle (•) indicates that the criterion has been satisfied and a blank circle (○) indicates that the criterion has not been fulfilled)

Work	**Scope**	**Tech.**	**Context of use**	**Risk**	**AI Act**	**Structured**	**Summarised**
Use Case Cards [64]	AI use case	○	•	○	•	○	•
AI Usage Cards [137]	Use of AI in research	○	•	•	○	•	○
AI Impact Assessment Report Template [155]	AI impact assessment	•	•	•	•	○	○
RISKCARDS [156]	Language models risk assessment & documentation	○	○	•	•	○	○
AI Risk Profiles [157]	Risk of AI use cases	•	•	•	•	○	•
Risk Management System tool [7]	AI Act's risk and data management systems	•	•	•	•	○	•

2.6 Findings

The state of the art review presented in this chapter revealed that the AI Act has opened new avenues for both theoretical and empirical research. While the research on the AI Act per se is not mature—which is no surprise given its recent publication—the extensive body of work on trustworthy AI, as well as compliance with digital regulations, in particular the GDPR, can be linked to requirements of the AI Act.

The landscape of AI risk taxonomies seems to be evolving at a rapid pace, which is to some extent due to the increased attention from policy-makers worldwide, as well as the public awareness of AI harms. While plurality of AI risk taxonomies is required, a caveat is that these efforts form a highly-fragmented and logically inconsistent landscape, with no simple way to compare and/or collate various AI risk taxonomies. This raises the necessity for common, standardised, and interoperable representation of AI risk taxonomies, which can be realised through adoption of Semantic Web-based approaches. The state of the art ontologies associated with AI risks have not yet devoted much attention to this direction. However, they illustrate an uptake in the studies that address some of trustworthy AI issues through standardised, machine-actionable, extensible, and interoperable semantic models.

While AI documentation approaches have become increasingly prevalent as a transparency measure, providing a holistic view of AI use cases has been overlooked in the state of the art. Additionally, the existing documentation formats are primarily designed for human consumption, without any practical tools that support (semi-)automated generation, continuous updating, and auditing of documentation.

Concluding this chapter, this thesis acknowledges the growing body of work on RegTech solutions, regardless of their use of Semantic Web technologies, that aim to facilitate navigation of the AI Act.

CHAPTER

THREE

ANALYSIS OF THE AI ACT

This chapter addresses the first research objective (RO1) by providing an analysis of the AI Act to identify the information requirements. First, Section 3.1 describes the methodology for conducting the analysis. Then, Section 3.2 investigates the AI Act's classification of AI systems into high-risk and prohibited categories to address RO1(a). Section 3.3 explores the specification of intended purpose of AI systems to address RO1(b). Section 3.4 provides an analysis of risk management and technical documentation for high-risk AI systems (RO1(c)). Finally, Section 3.5 investigates registration requirements for AI providers and deployers (RO1(d)).

This chapter reflects the interdisciplinary aspect of this research as it delves into the AI regulation and standardisation domains to capture requirements for semantic models and mechanisms that assist with the AI Act compliance tasks. It is important to note that this thesis does not aim to provide legal interpretations; it rather offers a conceptualisation of the AI Act's legal text. In doing so, it highlights the existing ambiguities in the legal text which may affect the analysis. In addition, in the absence of official guidelines and harmonised standards, this thesis uses relevant ISO/IEC standards (mentioned in Subsection 1.2.2), without intending to examine their sufficiency in fulfilling the requirements of the AI Act.

3.1 Methodology for Analysis and Conceptualisation

As mentioned earlier, during the time this research was undertaken, multiple AI Act mandates were published by authorities (refer to Subsection 1.2.1

for the list of mandates). Thus, it was essential to carry out the analysis using a progressive and iterative methodology that enables ongoing updates.

To understand the information requirements of the Act, the first iteration involved scrutinising the text to identify articles, annexes, and recitals that are relevant to each sub-objective outlined in RO1. The identified provisions were then analysed to capture information requirements for developing ontologies and guiding the implementation of Semantic Web-based mechanisms.

The subsequent iterations, which were triggered by publication of the mandates throughout the legislative process, included a preliminary review to determine the extent of changes affecting the previously conducted analysis. With new information requirements identified, the analysis was refined and the already-developed ontologies and mechanisms were updated. The number of refinement iterations varies for each part of the analysis, depending on the extent of changes to the provisions, the time frame in which the analysis was conducted, and the purpose of the iteration. An overview of the iterations for each part of the analysis is shown in the Table 3.1. Among these, changes to Annex III caused significant updates to analysis of high-risk AI systems (RO1(a)) and, in turn, to VAIR (refer to RO2 for the related objective and see Subsection 4.3.1 for the effect of the updates) and SHACL shapes for determining high-risk AI system (refer to RO3a for the relevant objective and see Section 5.1 for the shapes). These revisions provided evidence regarding the suitability of Semantic Web-based approaches for easily updating the ontology and SHACL shapes in accordance with the future amendments to Annex III.

Important to mention that the analysis provided in this chapter is based on the final version of the AI Act [5], except the detailed analysis of technical documentation (outlined in Annex IV), which is based on the Commission's proposal [28]. This is because the analysis of Annex IV was conducted and further validated in collaboration with the Joint Research Centre (JRC) prior to the publication of the final version, thus any subsequent changes could compromise the validation. The investigation of the final text has shown that the extent of changes required to update the Annex IV analysis is minimal, which asserts the continued relevance and validity of the detailed analysis of technical documentation presented in this thesis.

Table 3.1: An overview of analysis iterations

RO	Analysis of	No. of iterations	Justification
RO1(a)	Annex III high-risk AI	5	Considerable changes during the legislative process
	Annex I high-risk AI	1	No changes (the Union harmonisation legislation is already enacted)
	Prohibited AI practices	2	Timing (the analysis was conducted as per the Corrigendum and the final versions)
RO1(b)	Intended purpose of AI systems	1	No changes to the definition of intended purpose during the legislative process
RO1(c)	Risk management requirements	4	Ensuring all changes are reflected, given the central role of this requirement in the AI Act
	Technical documentation requirements	1	The analysis was conducted in collaboration with the JRC researchers and further validated by them
RO1(d)	Registration requirements	2	Timing (the analysis was conducted as per the Corrigendum and the final versions)

3.2 The AI Act's Risk-Based Classification Rules

The classification rules for high-risk AI systems, defined in Article 6, specifies two distinct set of conditions based on the modality of the AI system—which can be a product, a safety component of a product, or an AI application. An AI system is deemed to be high-risk if it meets either of the following conditions:

- An AI system that is referred to in Annex III.
- An AI system that is a *product* or a *safety component of a product*, covered by the Union harmonisation legislation (Annex I) and needs to undergo the third-party conformity assessment under that legislation.

Among these two overall conditions for high-risk AI systems, the AI uses

cases described in Annex III, with the possible exception of the use case listed under the area of critical infrastructure, primarily refer to situations where fundamental rights are more in need of protection, while the main concerns with most of the systems that fall under the already regulated domains, listed in Annex I, are related to health and safety [158]. However, it cannot be presumed that neither critical infrastructure high-risk use cases nor Annex I AI systems impose minimal harm to fundamental rights.

The following addresses RO1a by identifying the information requirements that are needed to determine whether an AI system should be classified as high-risk under Annex III of the AI Act. This analysis aims to discover a minimal set of concepts that enable providing unique specifications for each of the high-risk categories, in alignment with the definition of intended purpose.

3.2.1 Annex III High-Risk AI Systems

Annex III describes 25 categories of high-risk AI systems under 8 areas, where a brief description for each category is provided. For example, under the area of *Education and vocational training* (Annex III, point 3), 4 particular uses of AI are specified as high-risk, with one of them being described as follows: *"AI systems intended to be used for monitoring and detecting prohibited behaviour of students during tests in the context of or within educational and vocational training institutions at all levels"* (Annex III, point 3(d)).

To enable codification of Annex III high-risk AI systems, the "core concepts" whose combination makes an AI system high-risk were identified through unsupervised manual annotation of all of the 25 high-risk categories described within Annex III. In the first iteration, each of the categories was individually annotated based on an intuitive understanding of the AI Act to uncover the key characteristics, in form of atomic concepts that make that specific category high-risk (an example of the annotation for Annex III, Point 5(a) is depicted in Figure 3.1). Consecutively, the identified concepts were clustered together and revised to identify the minimum set of concepts that enable expressing the categories in a way that they can be sufficiently differentiated. The final analysis revealed 5 concepts whose various combinations express different categories of Annex III high-risk AI. These 5 concepts are expressed in the following questions:

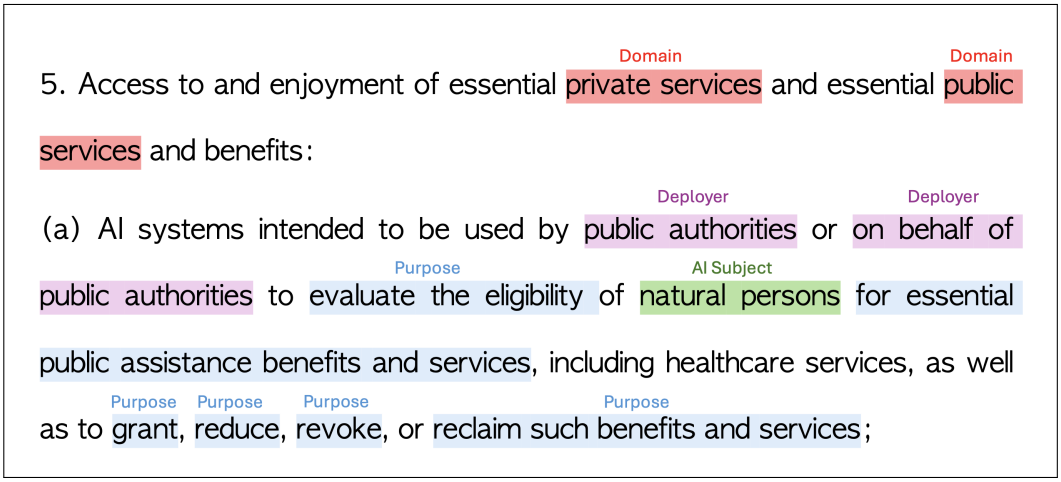


Figure 3.1: Analysis of the high-risk condition described in Annex III, Point 5(a)

1. In which ***domain*** is the AI system used?
2. What is the ***purpose*** of using the AI system?
3. What is the ***capability*** employed by the AI system to achieve the purpose?
4. Who is the ***deployer*** of the AI system?
5. Who is the ***AI subject***?

In the above mentioned questions, ***domain*** represents the area or sector the AI system is intended to be used in. ***Purpose*** is a concept that refers to the end goal of using an AI system within a use case. In this, *purpose* is distinguished from the AI Act's ***intended purpose***, which is a compound concept incorporating the intended use, and the specific context and conditions (refer to Section 3.3 for more detail). The AI system's ***capability*** enables realisation of the purpose and reflects the technological capability; for example *biometric identification* is the capability used towards achieving the purpose of *remote identification of people*. ***AI deployer*** is "a natural or legal person, public authority, agency or other body using an AI system under its authority", as defined by Article 3(4). ***AI subject*** refers to the entity who is subject to the use or influence of AI; *a person entering a territory* is the AI subject in an AI system used for *assessing the risk of irregular immigration*. Using these 5 concepts to determine if an AI system classifies

as high-risk under Annex III is a novel structure that is neither defined in the AI Act nor the state of the art.

Combinations of values assigned to the concepts, which can be treated as rules for high-risk uses for Annex III's high-risk applications, are illustrated in Figure 3.2 and represented in detail in Appendix C. If an AI system meets at least one of these conditions, it is likely to be high-risk. Notably, uses referred to in Annex III can be categorised as *non-high-risk* when they do *"not pose a significant risk of harm to the health, safety or fundamental rights of natural persons"*, except when they perform profiling of individuals (Article 6(3)). The AI Act makes the conditions for this derogation explicit by providing types of *purposes* for which the system intended to be used, which includes:

- Performing a narrow procedural task,
- Improving the result of a previously completed human activity,
- Detecting decision-making patterns or deviations from prior decision-making patterns,
- Performing a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.

AI providers who believe their system is *non-high-risk*, though it meets Annex III conditions, have to document the related (risk) assessment that has led to this conclusion (Article 5(4)). In addition, such providers need to register their systems into the EU database, as will be discussed in Section 3.5. It is therefore essential to carefully assess whether an AI system is high-risk or not, prior to placing the system into market or putting it into service, noting that misidentification of high-risk AI systems as non-high-risk can lead to up to EUR 15 million or 3 percent of the total worldwide annual turnover and can likely cause the system to be removed from the EU market.

The Commission is granted the authority to amend Annex III by modifying existing conditions, adding new high-risk uses, and removing existing ones through adoption of delegated acts (Article 6(6) and Article 7). While foreseeing the future types of AI system that will emerge is challenging, repeating the exercise of analysis over the change history of the AI Act is the closest existing proxy to show how the conceptualisation supports the Annex III revision mechanism. As such, the exercise of analysis was performed 5 times for the following key mandates:

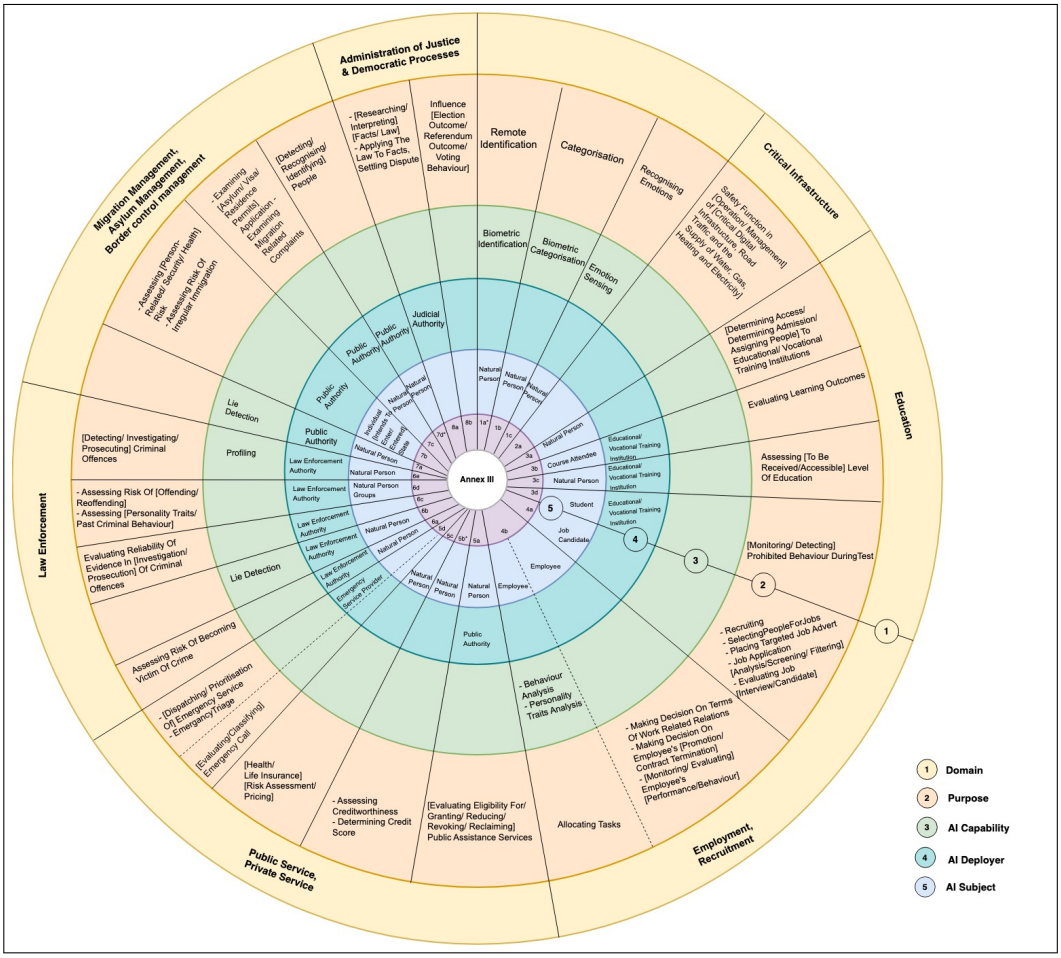


Figure 3.2: Describing high-risk AI uses as per Annex III, AI Act using the 5 concepts

- The European Commission's proposal [28], published in April 2021,
- The Council of EU's Common Position [29], issued in November 2022 by the Czech presidency,
- The European Parliament's mandate [31], published in June 2023,
- The Corrigendum [33], published in April 2024,
- The final version of AI Act [5], published in the Official Journal of the European Union in July 2024.

In these iterations, the first analysis was focused on identification of the core concepts. The consecutive iterations focused on verification of the 5

identified concepts for describing high-risk conditions, as these were iterated through the legislative process. Despite the amendments applied to the high-risk conditions throughout the legislative process, the analysis illustrated that all the various high-risk AI categories can be adequately described using the 5 concept model in a way that the categories are sufficiently differentiated. The modification of values assigned to each concept for describing a high-risk use case is supported through applying changes to the rules defined using SHACL, as will be discussed in Section 5.1.

3.2.2 Annex I High-Risk AI Systems

For background in the context of implementation of EU product rules, the *Old Approach* refers to the traditional manner in which national authorities drew up technical legislation, providing detailed texts containing all the necessary technical and administrative requirements [9]. The New Legislative Framework (NLF) refers to the most recent approach of the EU for product safety legislation, adopted in July 2008, which is restricted to essential requirements and leaves the technical details to European harmonised standards [9]. As mentioned earlier the AI Act follows the NLF structure, with its classification rules for high-risk AI systems interlinked to existing EU product safety regulations. Annex I of the AI Act provides a list of this existing law in two sections (Section A and Section B), representing EU product safety regulations and directives. While the 12 pieces of legislation that are listed in Section A follow the EU's NLF structure, the 5 pieces of legislation listed under Section B follow the Old Approach [158, 8].

When an AI system is a product or a safety component of a product within the scope of at least one of the Annex I regulations **and** is subjected to third-party conformity assessment, it is considered to be high-risk (Article 6(1)). Therefore, for determination of Annex I high-risk AI systems, it is necessary to first examine the scope of each of the regulations outlined in Annex I and second assess the need for third-part conformity assessments therein. With the long-standing presence and enforcement of product safety law, less legal uncertainty is expected in regard to *determination* of Annex I high-risk systems, compared to Annex III. Additionally, analysis of these regulations requires extensive domain-specific knowledge, which is not possessed by the author. Given the depth of regulations outlined in Annex I, a full assessment is not feasible. Therefore, rather than undertaking an analysis of all of these regulations, this section explores the potential of the 5 concepts identified for determining Annex III high-risk AI, previously presented in Subsection 3.2.1, in sufficiently expressing two of the Annex I high-risk AI systems, in way that they can be adequately differentiated.

EU Directive on Safety of Toys

There is a considerable growth in the AI-enabled smart toys market, as reported by the Market Research Future group [159], which has provoked debates around risks and ethical concerns in relation to such toys (see [160] and [161]). Given that the primary users of AI-enabled toys are minors, and considering their potential applications in educational settings [162], such toys have the potential to significantly impact fundamental rights. Thus, within Annex I, Directive 2009/48/EC on the Safety of Toys [163] emerges as a key regulation that, in combination with the AI Act, addresses the risks caused by AI toys.

The scope of the EU Directive on the Safety of Toys includes "*Products designed or intended, whether or not exclusively, for use in play by children under 14 years of age*", with a few exceptions (vide Article 2(2) of the Directive). If a product that meets the aforementioned definitions is an AI system or has an AI-based safety component, therefore it is likely to fall into the AI Act's high-risk AI category; examples are Mattel's Hello Barbie[1] and Wonder Workshop's Dash robot for use in classrooms[2]. Analysis showed that for conceptualisation of the scope, from the list of 5 concepts, purpose and AI subjects can be used. Given that the classification of the AI Act also covers the safety components of Annex I products, ***modality*** is considered as a new concept to reflect the form in which the AI system is placed on the market. Based on this discussion, the scope of the directive is expressed as combination of system's modality, purpose, and AI subjects, as shown in Table 3.2. These concepts can be used for determining potentially high-risk AI toys as per Annex I, Point 2; however further analysis is required for assessing if such toys require third-party conformity assessment under the EU Directive on the Safety of Toys.

Table 3.2: Conceptualisation of high-risk AI systems covered by the EU Directive on the Safety of Toys [163]

Modality	**Purpose**	**AI Subject**
Product	Use in play	Children under 14
Safety component	Safety function of toys	Any

[1] Discontinued due to privacy concerns.
[2] https://www.makewonder.com/en/classroom/

EU Medical Devices Regulation (MDR)

There has been a huge amount of AI investment in the medical and healthcare sectors for drugs, cancer, molecular, and drug discovery [164], resulting in increased concerns around risk of using AI in the medical domain [165]. Within the EU, the Medical Devices Regulation (MDR) [166] aims to ensure safety of medical devices through a risk-based approach. The similarities in the legislative instruments of the MDR and the AI Act, as well as the convergence between the requirements arising from both [167], underscore the critical role the interplay between the MDR and AI Act play in the AI regulatory landscape.

The scope of the MDR includes: (i) medical devices for human use and accessories for such devices, (ii) clinical investigations concerning the aforementioned medical devices and accessories, and (iii) products without an intended medical purpose (e.g. contact lenses) listed in Annex XVI of the MDR. Based on this scope, as well as the definition of *medical devices* in Article 2(1) of the MDR, high-risk AI systems according Annex I, Point 10 of the AI Act can be expressed using combination of domain, purpose, and AI subjects, as shown in Table 3.3. Similar to toys, safety function of medical devices is high-risk under the AI Act. For this reason and to accurately declare specific types of medical devices, the concept of modality is employed.

As emphasised earlier, full assessment of Annex I high-risk AI systems was not feasible in this thesis. The **partial** examination of two of the Union harmonisation legislation demonstrated that the 5 concepts used for describing Annex III rules are **not adequate** for expressing Annex I high-risk categories. In addition, the results cannot be generalised for all the Annex I categories, due to the limited scope of the analysis. Therefore, with the current knowledge, the extent to which the 5 concepts can assist in determining Annex I high-risk AI is inconclusive. However, this section demonstrates that the overall approach for conceptualisation is promising and has the potential be extended to include Annex I conditions.

Table 3.3: Conceptualisation of the AI Act's high-risk AI systems covered by the Medical Devices Regulation [166]

Modality	**Domain**	**Purpose**	**AI Subject**
Instrument, apparatus, appliance, software, implant, reagent, material	Medical	- Diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease - Investigation, replacement or modification of the anatomy or of a physiological or pathological process or state - Providing information by means of in vitro examination of specimens derived from the human body, including organ, blood, and tissue donations	Natural person
	Medical, administration of justice	Clinical investigations concerning medical devices and accessories	Any
Specific type of medical products (Annex XVI of MDR)	Any	Purposes other than medical	Natural person
Safety component	Medical	Safety function of medical devices that fall under the MDR	Any

3.2.3 Prohibited AI Practices

The AI Act prohibits provision and deployment of certain AI systems that impose *unacceptable risk* to health, safety, and fundamental rights. With a penalty up to EUR 35 million or 7 percent of the offender's total worldwide annual turnover, whichever is higher (Article 99(3)), misidentification of prohibited AI systems is highly consequential. The list of prohibited AI practices is provided in Chapter II, Article 5, whereby 8 main categories are specified. Within this list, 4 categories are prohibited under all circumstances, while exception rules are defined for the other 4. In relation to prohibited AI systems, the Commission is required to provide an annual report on the use of real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes (Article 6(7)). Further, the Commission

is required to annually assess the *need* for amending the list of prohibited AI practices and report the findings to the European Parliament and the Council (Article 112). However, it is not clear whether this review will lead to revising the scope of prohibited systems, given that amending the list of prohibited systems through delegated acts is not considered (refer to Article 97 wherein the conditions for adopting delegated acts is specified).

This section aims to identify the essential information elements needed to determine prohibited AI system as per Article 5. Building on top of the information elements for determining high-risk AI systems, the following steps were followed to create a framework for determining prohibited AI practices:

1. Identify the 5 concepts from each prohibited condition described in Article 5(1),

2. Determine whether the 5 concepts are sufficient to describe prohibited AI practices in a unique way that sufficiently distinguish them from each other,

3. Where the 5 concepts are not sufficient, identify the minimal set of additional concepts needed for describing the prohibited AI condition.

Conceptualisation of Prohibited AI Systems

Similar to conceptualisation of high-risk AI systems, this analysis aims to identify the minimum set of concepts that are adequate to uniquely describe prohibited AI practices. Following the steps outlined above, Article 5(1) clauses that describe prohibited AI practices were annotated and deconstructed into atomic concepts that shape prohibited conditions. As an example, annotation of Article 5(1a) is shown in Figure 3.3.

The annotation showed that from the 5 previously identified concepts, deployer does not play a role in determination of prohibited AI practices. In addition, it revealed 4 additional concepts that are: data processed by the system, locality of use, consequence, and impact. ***Locality of use*** refers to the environmental area in which the system is used, for example publicly accessible spaces. ***Consequence*** refers to the immediate negative effect of using the system, whether it leads to harms to individual, groups, and society. ***Impact*** expresses the overall effect of the system on individual, groups, and society or not. These concepts are expressed in the following questions:

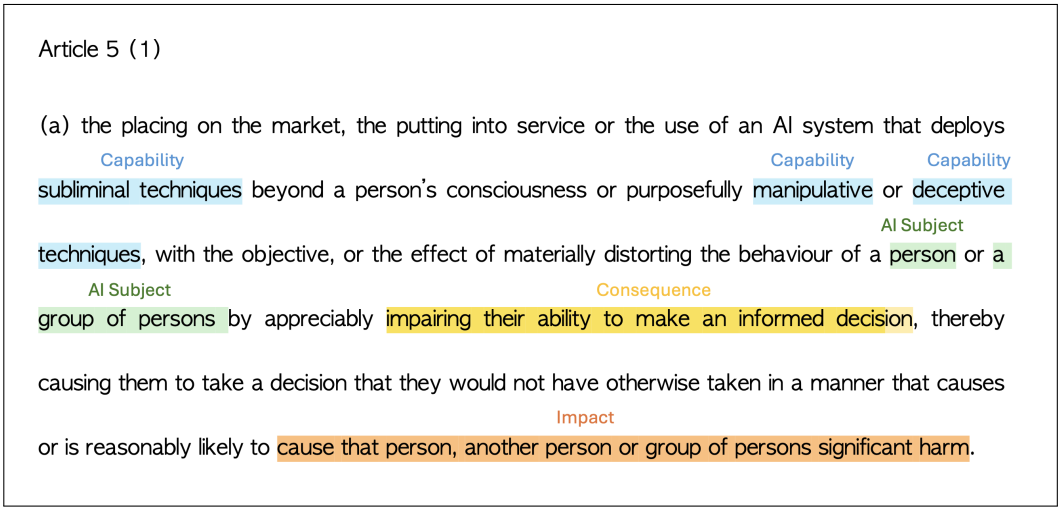


Figure 3.3: Analysis of prohibited AI practice described in Article 5(1a)

1. In which ***domain*** is the AI system used?
2. What is the ***purpose*** of using the AI system?
3. What is the ***capability*** of the AI system?
4. What ***data*** is processed by the AI system?
5. Who is the ***AI subject***?
6. What is the ***locality of use***?
7. What is the ***consequence*** of using the system?
8. What is the ***impact*** of using the AI system?

A summary of the conditions is illustrated in Figure 3.4 and the detailed analysis is presented in Appendix B. In the following, the reasoning behind the key design decisions made in this conceptualisation is discussed.

Discussion on the Key Design Decisions in Conceptualisation of Prohibited AI Conditions

Prohibited AI practices are described using a broadly interpretable terminology. Though definitions of the concepts related to biometrics-based systems are provided in Article 3, there is a crucial lack of established definitions for the core concepts used in describing Article 5(1a), specifically

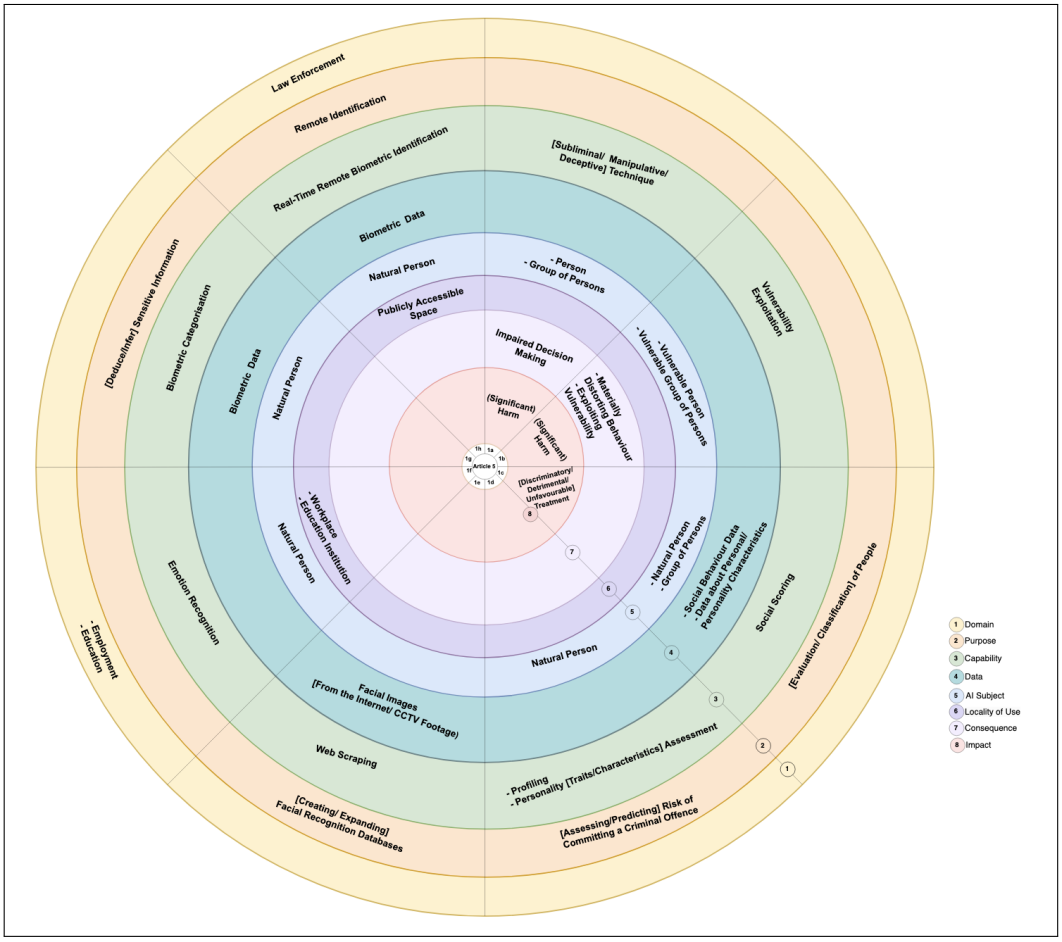


Figure 3.4: Describing Article 5(1) prohibited conditions using the 8 concepts

definitions for *subliminal*, *manipulative*, and *deceptive* techniques [79, 78, 77]. Given that the scope of prohibited practices heavily relies on how these concepts are interpreted [78], the body of literature on defining subliminal techniques and manipulation in the context of the EU AI Act is growing (see [78, 168, 79]), however at the time of writing, there is no authoritative guidelines published in this regard. Therefore, this work highlights the need for clear definitions for subliminal, manipulative, and deceptive techniques and represents them as *AI capabilities*.

The wording used in Article 5 (1a) and (1b) suggests that *materially distorting behaviour* can be either an objective or an effect of employing an AI system. As an objective, behaviour distortion can be conceivably interpreted as the characteristics of targeted advertising and personalised content. As it is highly unlikely that an AI provider explicitly declares purpose of its system

as behaviour distortion, the analysis considers it as a consequence.

Another challenge is dependency of some prohibited systems (Article 5 (1a), (1b), and (1c)) on what some scholars call the *"harm requirement"* [8, 6]. In terms of conceptualisation, the harm requirement can be expressed using two concepts of *consequence* and *impact*. As mentioned earlier, in Subsection 1.3.1, consequence refers to the immediate negative influence followed by the use of an AI system and impact defines the harmful impacts, resulting from materialisation of consequence, on individuals, groups, and society. For example, in Article 5(1a), as shown in Figure 3.3, impaired decision making is a consequence that can potentially lead to significant harms (impact) to the person, third-parties, or groups. In this, impact could be broken down into more granular concepts, namely *impacted stakeholder* and *severity of impact*, however in the absence of thresholds for significance of harm [168] and guidelines on its types, they are not included in the conceptualisation. As will be demonstrated in Chapter 4 and Chapter 5, through adoption of Semantic Web, the framework for determining risk category can be further expanded as guidelines become available.

3.3 Intended Purpose of AI Systems

Within the AI Act, there is a strong emphasis on *intended purpose*, which is described as the **use** of the system as specified by the provider. A description of the intended use of a high-risk AI system is required and must be accompanied by information about context and conditions of use (Article 3(12)). Intended purpose should be stated in instruction for use (Article 13) and technical documentation (Annex IV), and should be submitted upon the registration of high-risk AI systems into the EU database (Annex VIII). These imply a need for transparent communication of intended purpose.

Intended purpose is also a key element in assessing the risk category that an AI system belongs to under the AI Act [64]. In addition, this thesis argues that intended purpose is where the AI Act requires the provider to specify the benefits of the system. This, then addresses the concerns raised in [55] regarding the lack of risk-benefit analysis in the Act.

Approaching RO1(b), this section provides a minimal, clear, and comparable descriptive structure for expressing the intended purpose of an AI system. Given the link between assessing the risk category and the intended purpose of the system, the 5 concepts identified for determination of high-risk AI in Subsection 3.2.1 are proposed. From the additional elements for determination of prohibited AI practices (Subsection 3.2.3), *locality of use* and *data processed by the system* are proposed to be included to better capture

the context of use. These concepts also allow transparent and minimal expression of the conditions that AI system should not be used in, i.e. *precluded uses*, as required by Recital 72 of the AI Act.

To facilitate sharing the intended purpose, this thesis proposes expressing it as a part of a *use policy* that incorporates intended purposes (of use), precluded uses, and the conditions of use. While the provider might consider a myriad of conditions for using a given system, for the purpose of this work, the scope of the policy only includes two conditions that are directly tied to deployers' legal obligations and require transparent declarations from the provider, which are (i) implementing human oversight measures identified by the provider as per Article 14(3b) and (ii) risk and incident reporting referred to in Article 26(5). According to this scope, the questions that shape a minimal set of requirements for AI use policies are:

1. What is the **intended purpose(s)** of the AI system?
2. What is the **precluded use(s)** of the AI system?
3. To use the system as intended, what **human oversight measure(s)** should be implemented by the deployer?
4. What are the **risk reporting obligations** of the deployer?

3.4 Risk Management and Technical Documentation Requirements for High-Risk AI Systems

The AI Act outlines 7 key requirements for high-risk AI systems in Chapter III, Section 2 (see Table 3.4). The onus of ensuring that a high-risk AI system is compliant with these requirements is on providers of the system (Article 16). To accomplish this, the provider is mandated to put a quality management system in place (Article 17).

Considering the risk-centred nature of the AI Act, the risk management system requirement takes the lead in ensuring that the potential harms of AI are reduced to an acceptable level through continuous identification, evaluation, and mitigation of risks across the AI system's entire lifecycle. Demonstration of compliance with the high-risk AI requirements, including risk management system, entails maintaining, querying, and sharing information about the AI system and its risk management in the form of *technical documentation* (Article 11). Given the strong link between documentation and

Table 3.4: Requirements of high-risk AI systems as per Chapter III, Section 2 of the AI Act

Article	Title	Key requirement
9	Risk management system	Establish, implement, document, and maintain a risk management system
10	Data and data governance	Ensure training, validation, and testing datasets are of high quality and representative, and ensure that measures are in place to detect or correct biases
11	Technical documentation	Generate technical documentation with elements, set in Annex IV, in a way that it demonstrates conformity with the Act's requirements
12	Record-keeping	Ensure traceability through automatic recording of events (logs)
13	Transparency and provision of information to deployers	Generate instructions for use and share them with AI deployers
14	Human oversight	Identify and implement human oversight measure to address risks to health, safety, or fundamental rights
15	Accuracy, robustness and cybersecurity	Ensure appropriate level of accuracy, robustness, and cybersecurity

risk management to ensure compliance, this section investigates Articles 9 and 11 to address RO1(c), which ultimately aims at identification of the key AI and risk information elements that should be documented.

3.4.1 Article 9 - Risk Management System

For each high-risk AI system, it is mandatory to establish, implement, document, and maintain a risk management system—"*a continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic review and updating.*" (Article 9(1)). Risk management system documentation also should be included within the technical documentation (Annex IV, Point 5). Given that the scope of the Act is limited to essential requirements, it does not specify the details required to be featured in the documentation of risk management system. Therefore, to elaborate on the information obligations, the analysis uses two ISO/IEC standards that are candidate for harmonisation:

- **ISO/IEC 42001 - Information technology — Artificial intelli-**

gence — Management system [37], published in December 2023, is a key and certifiable standard that offers a framework to assist with implementation of AI management systems, which is defined as "*set of interrelated or interacting elements of an organization to establish policies and objectives, as well as processes to achieve those objectives*" (ISO/IEC 42001, 3.4). Since compliance with the high-risk AI obligations—including risk management—should be ensured through implementation of a quality management system (Article 17), ISO/IEC 42001 can potentially play a key role in addressing risk management requirements.

- **ISO/IEC 23894 - Artificial intelligence — Guidance on risk management** [39], published in February 2023, extends the ISO 31000 risk management guideline [40] specifically for AI system. The aim of ISO/IEC 23894 is to guide organisations in managing AI risks through integration of risk management tasks into AI development tasks or any activity that incorporates AI.

The grounds for selecting these two standards are: first, ISO/IEC 42001 follows the "harmonised approach" for management system standards (MSS), that defines a unified and consensus-based template for all ISO management system standards [169]. Historically, ISO management system standards have been harmonised for conformity with the requirements of the EU product safety regulations [170], notably in regard to the medical devices and toys [171]. In implementation of the requirements of the AI Act under a quality management system, ISO/IEC 42001 is a key standard. Although it is in partial alignment with the requirements of the Act [38], it can be integrated with other management systems that are already in place to address other applicable Union harmonisation legislation. In addition, ISO/IEC 42001 is closely linked to risk management, requiring organisations to establish an AI risk assessment process and have a plan for treating AI risks. Second, the conformity assessment process under a quality management system aligns well with the ISO/IEC 17000 series, also known as the ISO CASCO toolbox[3], which provides a set of conformity assessment tools to support certification and accreditation based on management systems. Worth mentioning that standards from the ISO 17000 series are already harmonised in relation to conformity assessment and CE marking mechanism in the NLF (see the list of harmonised standards in [170]). Based on these, this thesis considers ISO/IEC 42001 as a strong candidate for structuring the process for compliance with the AI Act.

[3] https://casco.iso.org/toolbox.html

In regard to risk management, ISO/IEC 23894, which extends ISO 31000, has the potential to be used in sync with AI management systems (ISO/IEC 42001). This is due to the fact that ISO/IEC 23894 is a guideline that can be customised and even integrated with an AI management system. In relation to harmonisation, ISO/IEC 23894 has been reported as a "valuable guidance", yet insufficient for compliance with Article 9 requirements [38]. while well-known standards, such as NIST's AI Risk Management Framework (AI RMF) [2] and IEEE Std 7000's process for addressing ethical concerns [43], can also be helpful in compliance with the Act, the mechanism for presumption of conformity only applies to harmonised European standards and common specifications, as per Article 40. Therefore, in the existing landscape of risk standardisation, ISO/IEC 23894 is a key candidate for harmonisation in relation to risk management, as requested in the standardisation request [13].

The requirement of technical documentation in regard to risk management is stated in the following statement: "a detailed description of the risk management system in accordance with Article 9" (Annex IV, Point 5). Based on the above mentioned discussion, ISO/IEC 42001 and ISO/IEC 23894 were used as a sources to expand on the minimum information elements that should be featured within the risk management documentation. The identified information elements are listed in Table 3.5 and Table 3.6. In this analysis, the overall structure of AI management systems was used for extracting key organisational activities required for implementation of AI risk management systems. For each activity, information requirements were identified from both ISO/IEC 42001 and 23894 (for a summary of the analysis see Figure 3.5). This information can be categorised into 4 overall categories:

- Information about the *context of the AI system and the organisation*, for example, the AI system's intended purpose and the role of the organisation in relation to the system.
- Details of the *risk management system* in place, e.g. the policies, responsibilities, and resources required for implementation of the risk management system itself. This category of information is relevant to the ISO/IEC 23894's AI risk management framework, whose intention is to help with AI governance and integration of AI risk management activities into an organisation's existing processes.
- Documentation of *risk management processes* across different phases: planning (ex-ante), operation (ongoing), and post-operation (ex-post).
- *Results of AI risk management*, which can be represented in artefacts

produced throughout the risk management process, e.g. risk assessment documentation that lists the identified AI risks, their likelihood, severity, sources, consequences, and impacts.

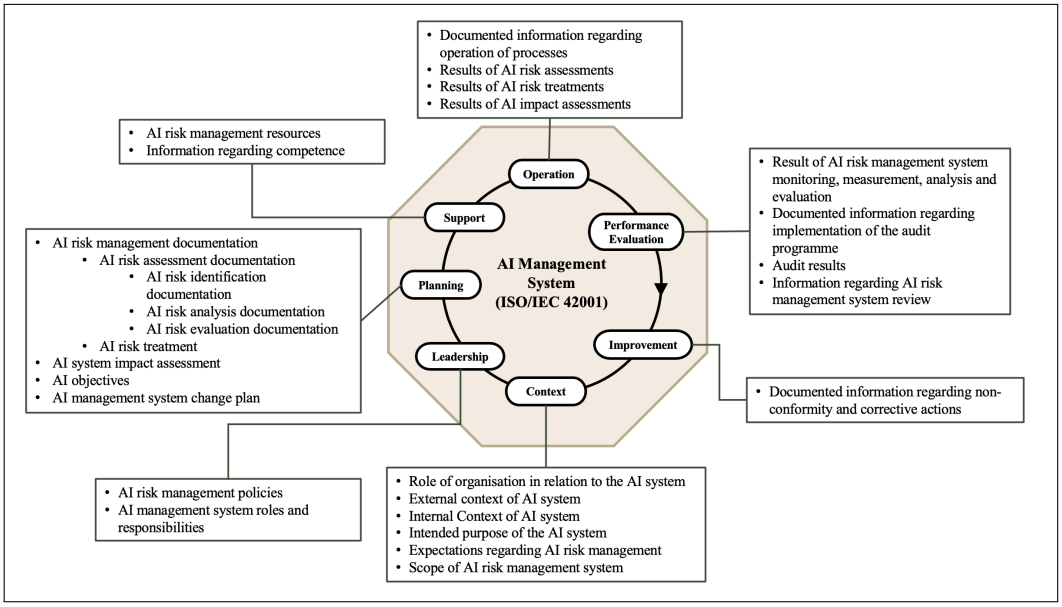


Figure 3.5: Summary of AI risk management system information requirements

Table 3.5: Risk management information elements extracted from ISO/IEC 23894 and ISO/IEC 42001 - part 1

ID	source	Information element
4-1	Article 9 & Annex IV(5)	Description of AI risk management system
4-1-1	ISO/IEC 42001 (4.1), ISO/IEC 23894 (5.4.1)	Role of the organisation in relation to the AI system
4-1-2	ISO/IEC 42001 (4.1), ISO/IEC 23894 (5.4.1)	External context of AI system
4-1-3	ISO/IEC 42001 (4.1), ISO/IEC 23894 (5.4.1)	Internal context of AI system
4-1-3-1	ISO/IEC 42001 (4.1)	Intended purpose of the AI system
4-1-4	ISO/IEC 42001 (4.2)	Needs and expectations of stakeholders (interested parties) in regard to AI risk management
4-1-5	ISO/IEC 42001 (4.3)	Scope of AI risk management system
4-1-6	ISO/IEC 42001 (5.2)	AI risk management policies
4-1-7	ISO/IEC 42001 (5.3)	AI management system roles and responsibilities
4-1-8	ISO/IEC 42001 (6.1)	AI risk management (documented information)
4-1-8-1	ISO/IEC 23894 (6.3.2)	Scope of AI risk management
4-1-8-2	ISO/IEC 23894 (6.3.2)	AI risk management objectives
4-1-8-3	ISO/IEC 23894 (6.3.2)	AI risk management tools
4-1-8-4	ISO/IEC 23894 (6.3.2)	AI risk management techniques
4-1-8-5	ISO/IEC 23894 (6.3.2)	AI risk management resources
4-1-8-6	ISO/IEC 23894 (6.3.2)	AI risk management responsibilities
4-1-8-7	ISO/IEC 23894 (6.3.3)	Internal context of AI system
4-1-8-8	ISO/IEC 23894 (6.3.3)	External context of AI system
4-1-8-9	ISO/IEC 23894 (6.3.4)	AI Risk criteria (for evaluation of risk significance)
4-1-8-10	ISO/IEC 23894 (6.4), ISO/IEC 42001 (6.1.2)	AI risk assessment (documented information)
4-1-8-10-1	ISO/IEC 23894 (6.4.2)	AI risk identification (documented information)
4-1-8-10-1-1	ISO/IEC 23894 (6.4.2.2)	Assets and their value
4-1-8-10-1-2	ISO/IEC 23894 (6.4.2.4)	Risk sources
4-1-8-10-1-3	ISO/IEC 23894 (6.4.2.3)	Entities associated with risk sources
4-1-8-10-1-4	ISO/IEC 42001 (6.1.2)	Risks
4-1-8-10-1-5	ISO/IEC 23894 (6.4.6), ISO/IEC 42001 (6.1.2)	Consequences
4-1-8-10-1-6	ISO/IEC 23894 (6.4.6), ISO/IEC 42001 (6.1.2)	Impacts
4-1-8-10-1-7	ISO/IEC 23894 (6.4.5), ISO/IEC 42001 (6.1.2)	Controls

Table 3.6: Risk management information elements extracted from ISO/IEC 23894 and ISO/IEC 42001 - part 2

ID	source	Information element
4-1-8-10-2	ISO/IEC 23894 (6.4.3)	AI risk analysis (documented information)
4-1-8-10-2-1	ISO/IEC 23894 (6.4.3.3)	Assessment of (likelihood) of risk sources
4-1-8-10-2-2	ISO/IEC 23894 (6.4.3.3)	Assessment of (likelihood) of risks
4-1-8-10-2-3	ISO/IEC 23894 (6.4.3.2)	Assessment of (likelihood and severity) of consequences
4-1-8-10-2-4	ISO/IEC 23894 (6.4.3.2)	Assessment of (likelihood and severity) of impacts
4-1-8-10-3	ISO/IEC 23894 (6.4.4)	AI risk evaluation (documented information)
4-1-8-11	ISO/IEC 23894 (6.5), ISO/IEC 42001 (6.1.3)	AI risk treatment (documented information)
4-1-8-11-1	ISO/IEC 42001 (6.1.3)	Statement of applicability
4-1-8-11-1-1	ISO/IEC 23894 (6.5.2 & 6.4.5), ISO/IEC 42001 (6.1.3)	Controls
4-1-8-11-1-2	ISO/IEC 42001 (6.1.3)	Control objectives
4-1-8-11-1-3	ISO/IEC 23894 (6.5.2)	Residual risk
4-1-8-12	ISO/IEC 42001 (6.1.4)	AI system impact assessment
4-1-9	ISO/IEC 42001 (6.2)	AI quality objectives
4-1-10	ISO/IEC 42001 (6.3)	AI management system change plan
4-1-11	ISO/IEC 42001 (7.1)	AI risk management resources
4-1-12	ISO/IEC 42001 (7.2)	Documented information about competence
4-1-13	ISO/IEC 42001 (8.1)	Documented information about operation of processes
4-1-14	ISO/IEC 42001 (8.2)	Results of AI risk assessments
4-1-15	ISO/IEC 42001 (8.3)	Results of AI risk treatments
4-1-16	ISO/IEC 42001 (8.4)	Results of AI impact assessments
4-1-17	ISO/IEC 42001 (9.1)	Result of AI risk management system monitoring, measurement, analysis and evaluation
4-1-18	ISO/IEC 42001 (9.2)	Documented information about implementation of the audit programme
4-1-19	ISO/IEC 42001 (9.2)	Audit results
4-1-20	ISO/IEC 42001 (9.3)	Documented information about AI risk management system review
4-1-21	ISO/IEC 42001 (9.4)	Documented information about nonconformity and corrective actions

3.4.2 Article 11 - Technical Documentation

Technical documentation is an essential document in assessing legal compliance with high-risk AI requirements, as stated in Article 11. The elements of technical documentation are described in Annex IV at a high-level. The preliminary analysis of Annex IV reveals that technical documentation is a compound document incorporating other documents required by the Act. Figure 3.6 summarises the AI Act's required documentation for high-risk AI systems and illustrates how they are related to each other.

To help with the generation and auditing of technical documentation, a minimum set of information elements is outlined in Annex IV, which is subject to the European Commission's potential amendments (Article 11(3)). Annex IV serves as a primary template wherein information elements are described with varying degrees of detail, with the majority articulated in a high-level manner. For example, the requirement for documenting risk management is succinctly stated as: "*A detailed description of the risk management system in accordance with Article 9*". It is therefore clear that further guidelines, templates, or standards are essential to support the implementation of Article 11. In the current standardisation landscape, of relevance to the Act's technical documentation is ISO/IEC DIS 12792 on AI transparency taxonomy [44], which provides terminology for and a taxonomy of information elements to assist in provision of transparency in AI systems. At the time of writing, ISO/IEC 12792 is a draft international standard and is subject to potential changes, therefore it is not included in the analysis. Its exclusion is also justified on the grounds of the Commission's standardisation request [13], which does not demand European standards specifically for technical documentation.

It is important to mention that the analysis was conducted based on the Commission's proposal [28], as it was conducted and validated in collaboration with the JRC before the publication of the final AI Act. However, the examination of the final version revealed minimal changes to Annex IV, which impacts the analysis provided in Appendix D negligibly.

To serve its purpose, technical documentation needs to be extensive and detailed. However, this thesis focuses only on capturing the information requirements from Annex IV to the extent necessary for addressing risk management; rather than establishing a comprehensive set of information elements required to be documented for compliance with the AI Act. Additionally, information regarding processes need to be excluded to fit the scope of the thesis defined in Subsection 1.2.3. Informed by the analysis of Annex IV and in collaboration with JRC researchers, the key AI and risk information requirements were identified. These requirements are presented in the

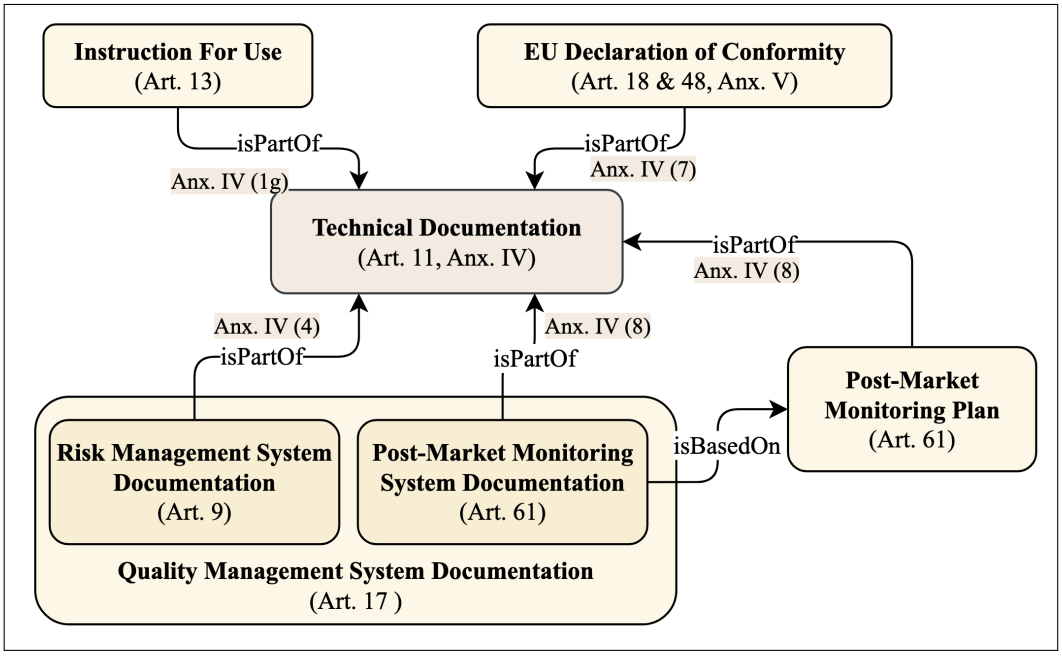


Figure 3.6: Documentation requirements for high-risk AI systems according to the AI Act

form of competency questions (CQ), that will be employed in development of the ontologies in Chapter 4 as part of their requirements and will be used in forming queries for document generation in Section 5.3. Further, these questions shape the AI Cards—the documentation framework that will be introduced in Chapter 6. The competency questions, which are categorised thematically, are:

1. General Information

CQ1-1. What is the version of the AI system?
CQ1-2. What is the modality of the AI system?
CQ1-3. What techniques are used in the AI system?
CQ1-4. Who is the AI provider?
CQ1-5. Who is the AI developer?

2. Intended use (purpose)

CQ2-1. What is the domain in which the AI system intended to used in?
CQ2-2. What is the purpose of the AI system?
CQ2-3. What is the capability of the AI system?
CQ2-4. Who is the deployer of the AI system?
CQ2-5. Who is the AI subject?
CQ2-6. What is the locality where the system is used?

3. Components

CQ3-1. What is the input of the system?
CQ3-2. What is the output of the AI system?
CQ3-3. What are the incorporating components of the system?
CQ3-4. For each component:
CQ3-4-1. What is the version of the component?
CQ3-4-2. What is the type of the component?
CQ3-4-3. What is the link to the component's documentation?
4. Data Processing
CQ4-1. What processing does the AI system perform on data?
CQ4-2. For each processing:
CQ4-2-1. What is the legal basis for processing?
CQ4-2-2. What data is processed?
CQ4-2-3. For each data, what is the source?
5. Human Involvement
CQ5-1. What is the level of automation?
CQ5-2. Who are the AI subjects?
CQ5-3. Who are the AI end-users?
CQ5-4. For each AI subject and end-user:
CQ5-4-1. Is the involvement intended?
CQ5-2-2. Is the involvement active?
CQ5-2-3. Is the involvement informed?
CQ5-2-4. What is the level of control over the AI outputs?
6. Risk Profile
CQ6-1. For each risk:
CQ6-1-1. What is the risk?
CQ6-1-2. What is the risk source?
CQ6-1-3. What is the consequence?
CQ6-1-4. What is the impact?
CQ6-1-5. Who is impacted?
CQ6-1-6. What is impacted?
CQ6-1-7. What is the likelihood of the risk source/risk/consequence/impact?
CQ6-1-8. What is the severity of consequence/impact?
CQ6-1-9. What measures are in place to detect/mitigate/eliminate risk source/risk/consequence/impact?
CQ6-1-10. What is the residual risk after applying the measures?
7. Quality
CQ7-1. What qualities are measured in regard to the AI system?
CQ7-2. For each quality, What is the quality measurement?
8. Pre-determined change
CQ8-1. For each pre-determined change:

CQ8-1-1. Which entity is changed?
CQ8-1-2. What is the frequency of change?
CQ8-1-3. What is the purpose of change?
9. Compliance & Certification
CQ9-1. Which regulations does the system comply with?
CQ9-2. Which standards does the system conform to?
CQ9-3. Which codes of conduct does the system follow?

3.5 Article 49 - Registration Requirements

Under the AI Act, providers and deployers of Annex III high-risk AI systems and providers of non-high-risk Annex III systems, i.e. systems that meet the conditions of Annex III but are considered as non-high-risk by the provider, are required to register their systems into the EU database (Article 49). According to Article 71, the EU database should be set up and maintained by the European Commission, in collaboration with the Member States. It shall be *"accessible and publicly available"* (with some exceptions), provided in a *"user friendly manner"*, and should be *"easily navigable and machine-readable"*. The EU database aims to act as an instrument for the Commission and the Member States to facilitate monitoring the current uptake of Annex III AI systems—regardless of their associated risk category—within the EU market and to serve as a transparency measure for sharing information regarding such systems with the public (Article 71 and Recital 131). The EU database therefore is a key data interoperability point between the Commission, AI providers, AI deployers, and the public.

Table 3.7 provides a summary of the registration provisions specified in Article 49. As shown in the table, the list of information elements that should be registered and their level of openness, i.e. publicly accessible or not, depends on the role of the registrant and the type of the system. In this, notably, submitting information regarding incorporating AI models, whether they are general-purpose or not, is not needed. However, information about general-purpose AI models should be made available to downstream AI providers that intend to use the model within their systems (Article 53).

Annex VIII, wherein the information to be submitted upon the registration of high-risk AI systems is outlined, was analysed to identify the *general* information that should be provided when registering an AI system into the EU database. Detailed information, such as the system's logic, instructions for use, and summary of fundamental rights impact assessment are not included, due to their descriptive nature and the lack of guidelines. In addition, for the general description of the general-purpose AI model, the key

information elements listed in Annex XII, Point 1, were included to enable representation of AI components in the ontologies (Chapter 4). Table 3.8 shows the key information elements extracted from Annex VIII and XII.

Table 3.7: Registration requirements for high-risk AI systems under the EU AI Act

AI Act Article	**AI System**	**Where?**	**What Information?**	**Who?**	**When?**
49(1)	High-risk as per Annex III, P. 3, 4, 5, 8	Public EU database	Annex VIII (A)	AI provider or authorised representative	Before placing on the market or putting into service
49(1) & (4)	High-risk as per Annex III, P. 1, 6, and 7	Non-public EU database	Annex VIII (A), points 1 to 10 (except 6, 8, and 9)	AI provider or authorised representative	Before placing on the market or putting into service
49(2)	Meets Annex III, P. 2, 3, 4, 5, 8 conditions but non-high-risk as per assessment of the provider	Public EU database	Annex VIII (B)	AI provider or authorised representative	Before placing on the market or putting into service
49(2) & (4)	Meets Annex III, P. 1, 6, and 7 conditions but non-high-risk as per assessment of the provider	Non-public EU database	Annex VIII (B), points 1 to 5 and points 8 and 9	AI provider or authorised representative	Before placing on the market or putting into service
49(3)	High-risk as per Annex III, P. 3, 4, 5, 8	Public EU database	Annex VIII (C)	AI deployer (public authorities, Union institutions, bodies, offices, or agencies)	Before putting into service or using
49(3) & (4)	High-risk as per Annex III, P. 1, 6, and 7	Non-public EU database	Annex VIII (C), points 1 to 3	AI deployer (public authorities, Union institutions, bodies, offices, or agencies)	Before putting into service or using
49(5)	High-risk as per Annex III, P. 2	Register at national level	Not mentioned	Not mentioned	Not mentioned

Table 3.8: Key information elements to be registered into the EU database

Annex	Clause	Requirement
Information about **operators**, including providers and deployers		
VIII	A1, B1	AI provider's name
	A1, B1	AI provider's address
	A1, B1	AI provider's contact details
	C1	AI deployer's name
	C1	AI deployer's address
	C1	AI deployer's contact details
Information about **AI system**		
VIII	A4, B4	AI system's trade name
	A4, B4	AI system's additional reference
	A5, B5	AI system's intended purpose
	A7, B8	AI system's market status
	A10, B9	Countries where system is available
Information about **components**, i.e. datasets and models		
VIII	A6	Data used by the system
	A6	Input data used by the system
	A5, B5	Component's intended purpose
	–	AI models used within the system
XII	1-1b	Model's use policy
	1-1c	Model's date of release
	1-1g	Model's input data
	1-1g	Model's output data
	1-1h	Model's license

3.6 Bridging the Analysis to the Thesis Artefacts

The analysis provided in this chapter sets out the functional requirements of the ontologies (AIRO and VAIR), which will be introduced in Chapter 4, and contributes is specification of the Semantic Web-based approaches and mechanisms to assist with the AI Act compliance tasks regarding risk management, documentation, and registration, which will be detailed in Chapter 5 and Chapter 6. For a better navigation, a more accurate representation of how this chapter provides inputs to the subsequent chapters is illustrated in Table 3.9.

Table 3.9: Bridging the analysis of the AI Act (provided in this chapter) to the thesis artefacts

Analysis of	**Functional requirements for**
Section 3.2. Risk-based classification (Article 5 & Article 6)	Chapter 4. AIRO and VAIR
Subsection 3.2.1. Annex III high-risk AI	Section 4.3. Population of VAIR Section 5.1. Rules to be expressed by SHACL shapes
Section 3.3. Intended purpose (Article 3(12))	Chapter 4. AIRO and VAIR Section 5.2. The profile for AI use policies (AIUP)
Subsection 3.4.1 Risk management (Article 9) & Subsection 3.4.2. Technical documentation (Article 11)	Chapter 4. AIRO and VAIR Section 5.3. SPAQRL queries for generating documentation Chapter 6. AI Cards framework
Section 3.5. Registration requirements (Article 49)	Chapter 4. AIRO and VAIR Section 5.4. The profile for cataloguing AI systems (AICat)

CHAPTER

FOUR

AIRO AND VAIR: ONTOLOGIES FOR THE AI ACT

Building on the initial conceptualisation in Chapter 3, this chapter addresses RO2 by developing two ontologies: (i) AIRO (AI Risk Ontology) that provides a minimal set of concepts and relations for modelling AI use cases and (ii) VAIR (Vocabulary of AI Risks) that is a specialisation of AIRO providing instances within a formal taxonomy. This means that AIRO contains the foundational concepts and VAIR is where the changes, potentially introduced through the previously mentioned delegated acts, reside. This modular design enables development of various extensions of AIRO and VAIR, whose overlaps and inconsistencies can be easily identified.

In the remainder of this chapter, Section 4.1 discusses the methodology used for developing the ontologies. Section 4.2 and Section 4.3 introduce AIRO and VAIR, respectively. Section 4.4 provides proof-of-concept regarding how AIRO and VAIR can be used in modelling AI use cases. Finally, Section 4.5 discusses the potential benefits and applications of AIRO and VAIR.

4.1 Methodology for Ontology Engineering

The development of both AIRO and VAIR follows the Linked Open Terms (LOT) methodology [56]—a lightweight methodology for developing ontologies and vocabularies that consists of four overall steps: requirements specification, implementation, publication, and maintenance. Grounded in extensive experience in ontology engineering, each step is supported with a set of useful tools, methods, and guidelines, that assist in the development of ontologies. Further, LOT's iterative workflow makes it an appropriate methodology for managing changes throughout the development process, in

particular to reflect the changes that emerged from various mandates of the AI Act. LOT's alignment with industrial practices, its collaborative approach that considers coordination of various actors (including domain experts) in ontology development, and its flexible and iterative approach, facilitate future efforts to extend the ontologies. In creation of the ontologies, Noy and McGuinness's "Ontology Development 101" guideline [172] are also utilised to follow best practices.

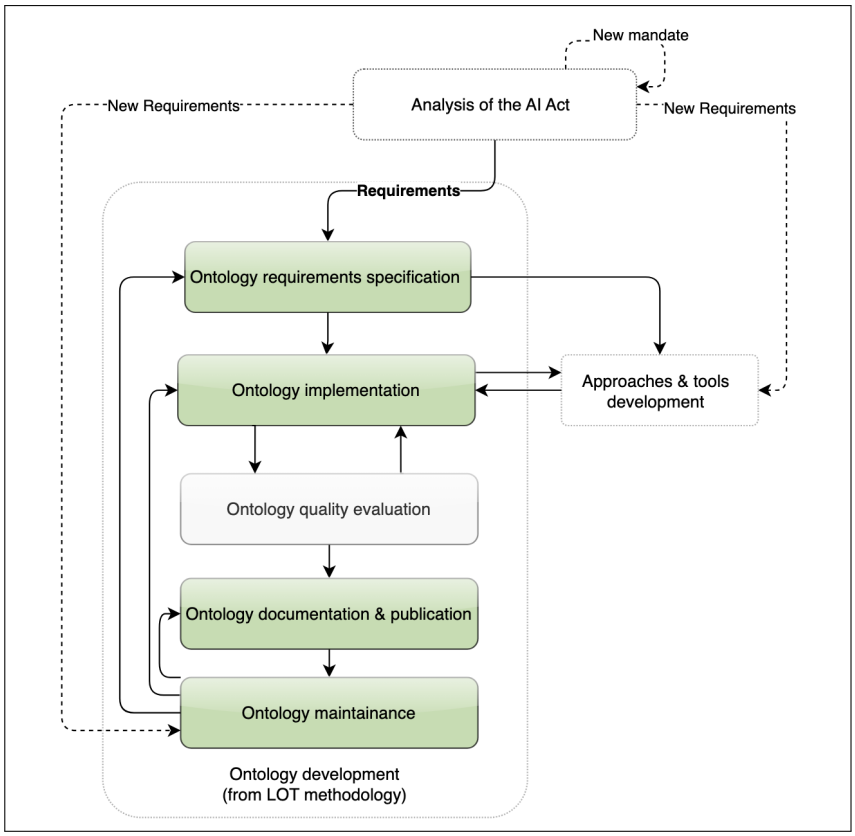


Figure 4.1: Ontology development methodology (the grey-coloured box illustrates that evaluation is not covered in this chapter)

A summary of the steps followed for creating AIRO and VAIR is illustrated in Figure 4.1 and is described in the following:

1. *Ontology requirements specification*: Based on the analysis presented in Chapter 3, the functional requirements are specified in 4 sets of competency questions addressing (a) identification of high-risk AI systems, (b) describing intended purpose of AI systems, (c) documentation of AI and risk information, and (d) registration of high-risk AI systems into the EU database.

2. *Ontology implementation*: In this phase, the concepts, their definitions, and relations are primarily derived from the AI Act. As discussed in Subsection 1.2.2, where the details required for satisfying the competency questions are not provided by the AI Act, the information modelling utilises ISO/IEC standards. The ontologies are encoded in OWL 2 Web Ontology Language [49] using Protégé ontology editor [173]. Additionally, for representing concepts in VAIR, the W3C SKOS data model [50] is used. In the encoding process, existing vocabularies for describing metadata and existing relevant ontological resources are reused (see Subsection 4.2.3). It is worth mentioning that the ontologies reviewed in Subsection 2.3.3 were identified as improper for reuse, mainly due to the fact that they are not aligned with the AI Act.

3. *Ontology evaluation*: AIRO and VAIR are evaluated against the competency questions to ensure their sufficiency in addressing those questions. Their quality is ensured by following the W3C Data on the Web Best Practices [59], and FAIR best practices for vocabularies and ontologies [62]. They are tested iteratively using OOPS! (OntOlogy Pitfall Scanner!) [60] and FOOPS! (ontology pitfall scanner for FAIR principles) [174], as will be discussed in Subsection 7.3.2. AIRO and VAIR are also evaluated through their application in modelling AI use cases (see Section 4.4). In addition, their applicability in (a) determining Annex III high-risk AI systems, (b) expressing AI intended purposes within a use policy, (c) documentation of AI and risk information, and (d) cataloguing AI systems for registration into the EU database is demonstrated in Chapter 5.

4. *Ontology documentation and publication*: The HTML documentation is created in a semi-automated process using the ReSpec template[1]—a common template for W3C specifications. WIDOCO [175], a well-known tool for automated generation of ontology documentation, is not used due to its limitations in supporting grouping of taxonomies of concepts [57]. Both AIRO and VAIR use permanent URLs (PURLs) using the W3ID service[2] and are made available online at https://w3id.org/airo and https://w3id.org/vair under the CC-BY-4.0 license[3], allowing free reuse, distribution, and modification conditioned upon appropriate attribution. Both ontologies are indexed in the Linked

[1] https://www.w3.org/respec/
[2] https://w3id.org/
[3] https://creativecommons.org/licenses/by/4.0/deed.en

Open Vocabularies (LOV) registry[4] and published in Zenodo for better findability and accessibility.

5. *Ontology maintenance*: During the time this research has been carried out, the AI Act was undergoing the legislative process and therefore requirements and concepts derived from the Act needed to be revised as newer versions were published. Additionally, multiple drafts of the relevant ISO/IEC standards were published over the course of the last 4 years, which to some extent influenced the ontology design. AIRO and VAIR have been iteratively updated following the LOT methodology in order to keep requirements, implementation, and documentation up to date.

4.2 AI Risk Ontology (AIRO)

AIRO provides a formal representation of AI systems and their risks as per the EU AI Act. As emphasised before, the AI Act only defines essential requirements and relies on future harmonised standards and common specifications for supporting its interpretation and implementation. With no such standards available at the time of writing, AIRO uses existing international standards as sources to extract concepts and their relations.

4.2.1 Iterative Development of AIRO

The first iteration of the development, in 2021, employed the following standards from ISO's well-known 31000 series on risk management: ISO 31000:2018 on risk management guidelines [40] and ISO Guide 73:2009 on risk management vocabulary [41], which was later replaced by ISO 31073:2022 Risk management — Vocabulary [42]. This replacement required the sources (`dct:source`) to be updated, however it did not impact the design of AIRO. With official publication of ISO/IEC 22989 [36] in 2022, the AI concepts and definitions were revised. While the terms used within the AI Act and ISO/IEC 22989 are overlapping, in some cases definitions and taxonomical relations are not exactly the same. In the process of development, the AI Act served as the primary resource and therefore definitions from standards were used when there was a lack of definition for a specific concept in the AI Act. Another round of revision was conducted when ISO/IEC 23894 [39] on AI risk management was published in 2023. Given that this standard relies on the definitions provided in ISO 31000:2018, ISO 31073:2022, and ISO/IEC

[4] https://lov.linkeddata.es/dataset/lov/

22989:2022, AIRO's concepts and definitions did not require any changes to be aligned with ISO/IEC 23894. In addition to the aforementioned standards, the definitions provided in ISO/IEC TR 24028:2020 — Overview of trustworthiness in artificial intelligence [176] were used for some of the AI-related concepts. Further, the definition for `Standard` was extracted from Regulation (EU) No 1025/2012 on European standardisation [12], as the AI Act relies on the definition of harmonised standards in this regulation. Table 4.1 shows the non-ontological sources from which definitions for AIRO concepts were extracted. To ensure tractability, from the Dublin Core Metadata Terms [177], `dct:source` is used to link the concepts to the source wherein they are defined.

Table 4.1: Non-ontological sources of AIRO

Source	**No. of concepts**	**AIRO Concepts**
The AI Act, Article 3	5	`AI System`, `AI Operator`, `AI Deployer`, `AI Provider`, `General Purpose AI Model`
ISO 31073:2022 [42]	8	`Risk Source`, `Hazard`, `Threat`, `Vulnerability`, `Consequence`, `Risk Control`, `Frequency`, `Likelihood`
ISO/IEC 22989:2022 [36]	3	`AI Component`, `Stakeholder`, `AI Developer`
ISO/IEC TR 24028:2020 [176]	2	`AI User`, `AI Model`
Regulation (EU) No 1025/2012 on European standardisation [12]	1	`Standard`

In addition to the iterations triggered by publication of standards, AIRO was revised throughout the AI Act's ordinary legislative process to incorporate the changes introduced in different phases[5]. The extent of changes to AIRO's concepts and relations were minimal, primarily affecting the definitions extracted from Article 3 of the Act.

It should be noted that in the process of ontology development, initially all the concepts were included in a single ontology. As the ontology expanded with population of the concepts, it became less reusable and more complex to manage. Therefore, the ontology was broken down into a minimal ontology

[5]For the historic timeline of the AI Act refer to https://artificialintelligenceact.eu/developments/.

(AIRO), that provides foundational concepts and relations, and a separate vocabulary (VAIR) extending AIRO with hierarchies of instances.

4.2.2 AIRO Overview

AIRO's core concepts and relations are illustrated in Figure 4.2. The upper half shows the main concepts required for describing a specific use of an `AI System` (green boxes), and the lower half represents key concepts for expressing `Risk` (yellow boxes). The relation `hasRisk` links these two halves by connecting risk to an AI system or a component of the system. AIRO incorporates 56 classes, 61 object properties, and 3 data properties. 10 of the concepts, 9 of the object properties, and all of the data properties are reused from existing ontologies (see Subsection 4.2.3 for details).

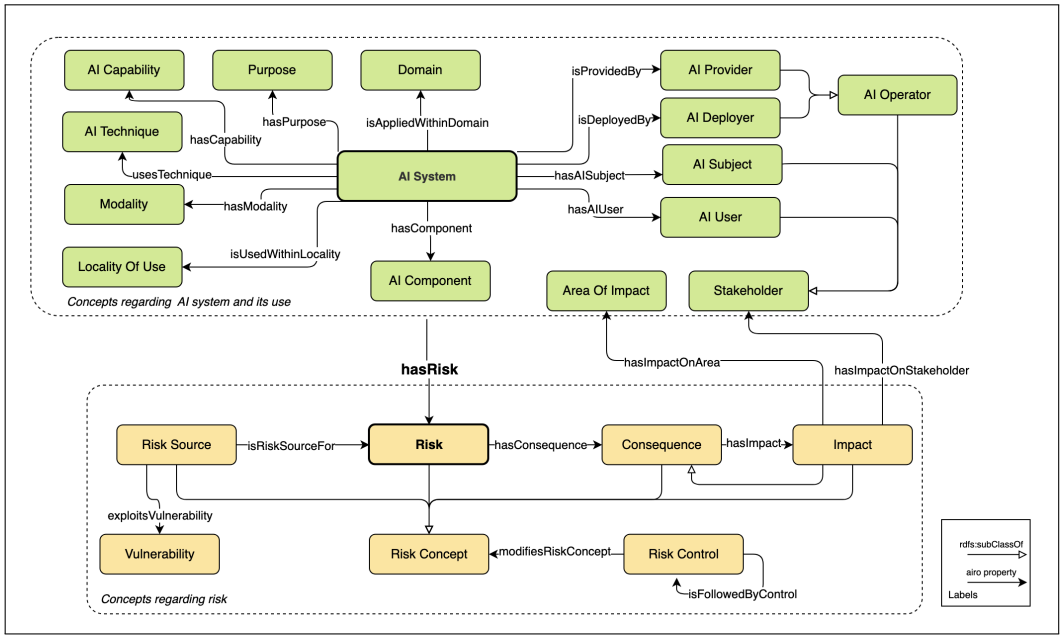


Figure 4.2: Overview of AIRO's main concepts and relations

The core concepts related to the use of an AI system are as follows:

- `Domain`: indicates the area, sector, or industry in which the AI system is used.
- `Purpose`: specifies end goal for which the system is used.
- `Locality Of Use`: specifies the locality the system is designed to be used within, e.g. publicly accessible spaces.

- `AI User`: refers to any individual or group that interacts with a system (ISO/IEC TR 24028, 3.43).

- `AI Subject`: represents an entity that is subject to or impacted by the use of AI.

Regarding technical details, the key concepts are:

- `AI Technique`: refers to computer science approaches and techniques used in development of a system, e.g. supervised machine learning.

- `AI Capability`: refers to the capability of an AI system that enables realisation of the system's purposes, e.g. facial recognition.

- `AI Component`: specifies incorporating components of the system.

- `Modality`: indicates the way or type in which the system exists, e.g. service.

The key actors involved in the AI development and deployment that are modelled in AIRO are as follows:

- `AI Provider`: is a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge (Article 3(3)).

- `AI Deployer`: is any natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity (Article 3(4)).

- `AI Developer`: is an organisation or entity that is concerned with the development of AI services and products (ISO/IEC 22989, 5.19.3.2).

- `AI Operator`: refers to a provider, product manufacturer, deployer, authorised representative, importer or distributor (Article 3(8)).

Concerning risk, the key concepts are:

- `Risk`: The AI Act defines risk as *"the combination of the probability of an occurrence of harm and the severity of that harm"*. Risk in ISO 31073 is defined as *"effect of uncertainty on objectives"*, with a note indicating that the effect can be negative, positive, or both. While both definitions provide quantitative perception of risk, ISO 31073 considers a broader scope of effects by including positive impacts, compared to the AI Act that only focuses on harms, i.e. negative effects. On another hand, the AI Act's definition of risk, together with the language used throughout its text, indicates a focus on harm to fundamental rights, health, and safety, whereas ISO focuses on broad organisational objectives. Based on this and the discussion provided in Subsection 1.3.1, AIRO considers a novel definition of `Risk` as **the state of uncertainty associated with an AI system, that has the potential to cause harms and is expressed in terms of risk sources, consequences, impacts, likelihood, and severity**. `hasRisk` relation is used to associate a specific risk to an AI system or a component of the system.

- `Vulnerability`: refers to properties of an entity, e.g. AI system or AI component, resulting in susceptibility to a risk source (ISO 31073:2022, 3.3.21 with modifications).

- `Risk Source`: indicates the element which alone or in combination has the potential to give rise to risk (ISO 31073:2022, 3.3.10). `Threat` and `Hazard` are both types of risk sources that have potential to cause harm.

- `Consequence`: is the direct outcome of a risk that affects objectives (ISO 31073:2022, 3.3.18, event is replaced with risk for better accuracy).

- `Impact`: represents the outcome of a consequence on individuals, groups, society, and environment. While consequence indicates the direct outcome of a risk on objectives, which may or may not involve external entities, the notion of impact puts emphasis on the effect and influence on individuals, groups, and society. This distinction addresses the concern raised by Soler et al. [38] in regard to non-alignment of ISO/IEC and AI Act, rooted in organisation-focused view of risks in ISO/IEC documents that misses the AI Act's centre of attention: the effect of AI on individuals' health, safety, and fundamental rights.

- `Risk Control`: ISO 31073 defines *risk control* as a measure that maintains and/or modifies risks[6]. In AIRO, risk control is not restricted

[6]ISO 31073:2022, 3.3.33

to risks and indicates a measure that is applied to detect, mitigate, or eliminate risk sources, risks, consequences, or impacts.

To enhance flexibility in modelling various AI use cases, domain or range of some properties are not restricted to any concepts, for example for `airo:hasRisk`, no domains is specified. Another design decision made concerning object properties in AIRO is the exclusion of inverse relations to avoid redundancy and maintain simplicity. AIRO is published online at https://w3id.org/airo[7] under the CC-BY-4.0 license and is indexed in the linked open vocabularies (LOV) registry at https://lov.linkeddata.es/dataset/lov/vocabs/airo. Evaluation of AIRO will be discussed in Section 7.3.

4.2.3 Ontology Reuse

In addressing competency questions, as per established Semantic Web practices, where more specialised and well-known open ontologies exists, appropriate concepts and relations are reused from them, rather than modelling them from scratch. In AIRO, classes and properties from Data Protection Vocabulary (DPV)[8] [57], Data Quality Vocabulary (DQV)[9] [178], Data Catalog Vocabulary (DCAT) [54], and DCMI Metadata Terms [177] are reused. The list of reused concepts and properties is presented in Table 4.2.

4.3 VAIR: A Vocabulary of AI Risks

The high-level model provided by AIRO is not sufficient for annotating AI use cases, expressing rules for identification of the AI Act's high-risk AI systems, and documenting and registering AI and risk information. These require enriching AIRO with specialisations of concepts that are represented formally and organised in hierarchies. The Vocabulary of AI Risks (VAIR) is a specialisation of AIRO, providing a taxonomy for expressing and exchanging information regarding AI systems and their risks aligned with the requirements of the AI Act.

4.3.1 Iterative Development of VAIR

As previously mentioned, the first version of AIRO included key concepts plus their specialisations, which were modelled as sub-classes. However,

[7] Also available on Zenodo at https://doi.org/10.5281/zenodo.10894952
[8] https://w3id.org/dpv/
[9] http://www.w3.org/ns/dqv

Table 4.2: Ontologies reused in AIRO

Source	Classes	Object properties	Data properties
DCAT	dcat:Dataset	None	None
DPV	dpv:Processing, dpv:Data, dpv:SensitiveData, dpv:DataSource, dpv:DataSubject, dpv:LegalBasis	dpv:hasProcessing, dpv:hasData, dpv:hasDataSource, dpv:hasDataSubject, dpv:hasLegalBasis	None
DQV	dqv:QualityMeasurement, dqv:Metric, dqv:Dimension,	dqv:hasQualityMeasurement, dqv:isMeasurementOf, dqv:inDimension, dqv:expectedDataType	dqv:value
DCT	None	None	dct:date, dct:title

emergence of new standards and publication of multiple AI Act mandates have shown that while the key top-level concepts remain largely intact, new sub-classes are likely to be iteratively identified from multiple sources as the understanding of AI risks evolve over time.

Therefore, the instances were de-integrated from AIRO and provided under the VAIR namespace. In this modular structure, AIRO acts as a foundational ontology, including only key high-level concepts and relations, while VAIR serves as a specialisation of AIRO that contains a taxonomy of these key concepts. This design decision contributes to better reusability of AIRO and VAIR, in particular for creating domain-specific risk taxonomies.

VAIR was updated iteratively, upon publication of different mandates of the AI Act. In addition to the AI Act, the EU AI Watch's[10] AI taxonomy [179, 180] and ISO/IEC 22989:2022 on AI terminology [36] were used to further populate VAIR.

Providing an exhaustive and commonly-agreed taxonomy in this thesis was not practical, firstly, due to the extensive state of the art and non-Semantic Web efforts to build AI risk taxonomies. Secondly, building such a taxonomy requires involvement of stakeholders to build consensus around the

[10]AI Watch is the European Commission's JRC knowledge service on AI, that supports development and uptake of trustworthy AI in the EU.

taxonomy. Therefore, VAIR focuses on inclusion of the concepts that are essential to satisfy the competency questions and are needed for implementing the Semantic Web approaches that are in the scope of this thesis (listed in RO3). The latter particularly includes specialisations for the 5 concepts identified for determining Annex III high-risk AI systems. With the publication of the final version of the AI Act, some of the concepts extracted from the Act mandates were no longer essential in implementation of the approaches, however, for better comprehensiveness no concepts have been removed, but the wording of some concepts has been modified for better clarity.

Regarding stakeholders' involvement in the development process, as will be discussed in Section 8.3.2, AIRO and VAIR have been proposed to the W3C Data Privacy Vocabularies and Controls Community Group[11] (DPVCG) to be included in DPV through a consensus-based process.

4.3.2 Overview of VAIR

VAIR is a hierarchical taxonomy, providing a specialisation of AIRO's concepts and therefore serves as a vocabulary for expressing AI use cases in an interoperable manner. VAIR is provided as an OWL 2 ontology where the concepts are defined using both `skos:Concept` and `rdfs:Class`, with hierarchical relations between them modelled using `rdfs:subClassOf`. This promotes extensibility and reusability by enabling the concepts to be used as both instances and classes in specification of an AI system.

In total, VAIR includes 402 classes, with the following thematically-categorised taxonomies:

- **AI**: contains taxonomies of *techniques* (number of instances in the taxonomy: 17), *capabilities* (32), *types* of AI systems (13), *components* (23), *lifecycle phases* (10), *modality* (4) and *outputs* (5).
- **Use of AI**: includes taxonomies for defining AI use cases namely *purposes* (118) and *domains* (11). In addition, 7 *levels of automation*, 6 types of *human involvement*, and 6 *modes of output controllability* are included (refer to Section 6.2 for more details).
- **Risk**: contains taxonomies of *risk sources* (44), *consequences* (4), *impacts* (5), *risk controls* (20), and *areas of impact* (7).
- **Stakeholder**: contains types of *AI operators* (17) and *AI subjects* (17), mainly extracted from Annex III of the AI Act.

[11] https://www.w3.org/groups/cg/dpvcg/

- **Document and standard**: contains a list of *documentation* (13), including those required for conformity assessments, and *standards* (23) that can potentially be used in implementation of the AI Act.

VAIR is published online as an open resource under the CC-BY-4.0 license at https://w3id.org/vair and is indexed in LOV at https://lov.linkeddata.es/dataset/lov/vocabs/vair[12]. Evaluation of VAIR will be presented in Section 7.3.

4.4 Modelling Use Cases Using AIRO and VAIR

This section aims to demonstrate the applicability of AIRO and VAIR in modelling AI use cases and their associated risks. To this end, one synthetic use case was created and two incidents from the AIAAIC were selected. The criteria for creation and selection of use cases are described in the following:

- The initial analysis of the use case demonstrates that it is highly likely to be classified as high-risk as per Annex III,
- The domains in which the systems are used should not overlap. This is to demonstrate the capability of the ontologies to be used in and/or extended for various domains,
- Diversity in types of harmful impacts is preferred in order to include harms to health, safety, and fundamental rights.
- For use cases selected from the AIAAIC repository, documentation that provides detailed information about the AI system and its risks should be available. In this, topicality of the use case is preferred.

Based on these criteria, the following use cases were selected to be modelled:

- *Proctify*: is a synthetic, but realistic, AI use case in the education domain (referred to in Annex III, Point 3) for student proctoring. This use case has been designed to show the potential of AIRO and VAIR, without being restricted to model the information provided by third-parties.

[12] Also available on Zenodo at https://doi.org/10.5281/zenodo.10894914

- *Uber's Real Time ID (RTID)* (incident ID = AIAAIC0756): is a system to verify the identity of Uber drivers using facial recognition. This incident was selected as an example of AI systems used in the employment domain (Annex III, Point 4). The system has negatively affected certain demographics and has led to violation of fundamental rights.
- The *Spanish Ministry of the Interior's AI system (VioGén)* (incident ID = AIAAIC0848): is an AI system that is being used for assessing risk of domestic violence. This use case is related to both public services (Annex III, Point 4) and law enforcement (Annex III, Point 6) domains. The incidents caused by VioGén have already impacted health and safety of victims of gender violence.

In the modelling process, each use case was initially annotated manually using AIRO. If the identified individual, which was annotated as an `rdf:type` of an AIRO concept, belonged to a more specific class in VAIR, then it was defined as `rdf:type` of that specific VAIR class. If the individual matched an `owl:NamedIndividual` in VAIR, it was replaced with the corresponding VAIR concept. It should be emphasised that this exercise demonstrated applicability by presenting how AIRO and VAIR can be utilised for modelling use cases, rather than illustrating their comprehensiveness, given that development of comprehensive AI risk ontologies was not feasible.

4.4.1 Proctify

To demonstrate the potential scalability and applicability of AIRO and VAIR, and inspired by the use cases described in [63] and [64], an AI-based student proctoring tool called *Proctify* is used as an illustrative proof-of-concept. Under the AI Act, AI systems used in the education domain for monitoring and detecting suspicious behaviour of students during tests are considered high-risk (Annex III, Point 3(d)). Therefore, it is highly likely that Proctify is classified as high-risk. It should be noted that the specification of Proctify was created in collaboration with researchers from the European Commission's Joint Research Centre (JRC) who had practical experience in development and legal analysis of facial analysis systems.

Proctify is intended to detect suspicious behaviour during online exams by analysing facial behaviour from a student's facial video captured throughout the exam using a webcam. Prior to this, students have explicitly consented to be recorded during the exam and informed that they must be alone in the room. The system incorporates a graphic interface displaying an analysis of the student's face including the head pose, gaze direction, and face landmarks' positions. This extracted information is then provided as an input to

SusBehavedModel, which has been trained in-house by the system's provider using *SusBehavedDataset*, to determine whether the student is displaying suspicious behaviour, e.g. looking away from the screen, leaving the room, or a third person detected in the room. Detection of suspicious behaviour raises an alarm in the interface to inform and let the human oversight actors, e.g. human instructors, take appropriate actions, e.g. communicating with the student.

The machine-readable representation of Proctify using AIRO/VAIR in Turtle serialisation is available online on GitHub[13]. Figure 4.3 depicts a visualisation of this representation, created using RDF Grapher[14].

[13] https://github.com/DelaramGlp/airo/blob/main/usecase/proctify.ttl
[14] https://www.ldf.fi/service/rdf-grapher

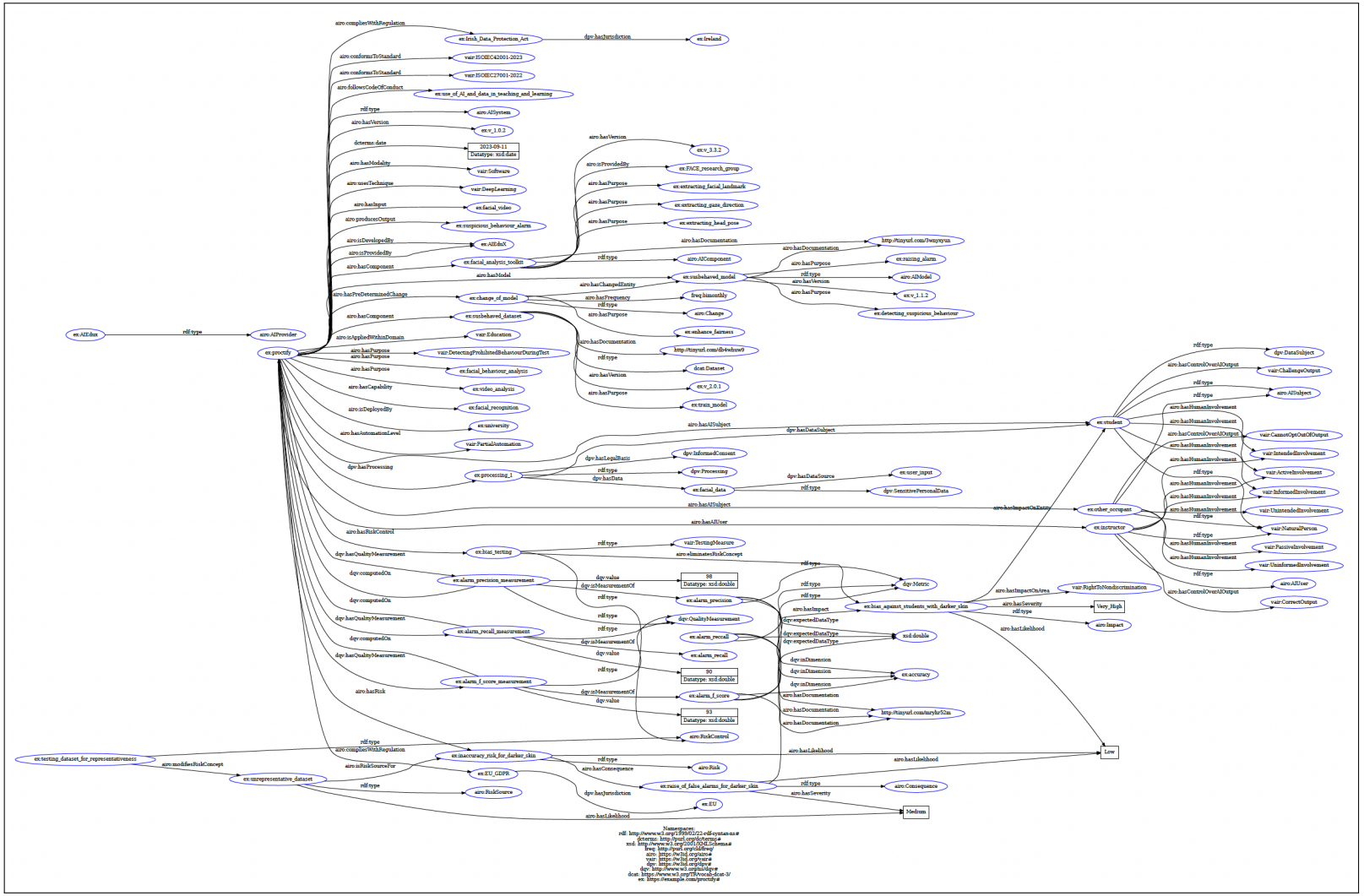


Figure 4.3: Visual representation of the graph specifying Proctify

4.4.2 Uber's Real-time ID Check System

Uber's facial recognition identification system, known as the Real Time ID (RTID), aimed to ensure that the driver's account is not used by anyone other than the registered Uber driver. When the system failed to recognise a person for two consecutive times, the driver's contract would be terminated and their driver and vehicle licenses would be revoked. Multiple incidents were reported concerning the system's failure in verifying drivers, in particular those belong to ethnic minorities or vulnerable groups, leading to unfair dismissal of drivers (vide [181] and [182]). This system can be considered high-risk as per Annex III, Point 4(b) as it is used to "make decisions affecting terms of work-related relationships". Figure 4.4 depicts the graph expressing Uber's Real-time ID Check use case using AIRO and VAIR, and is available online[15].

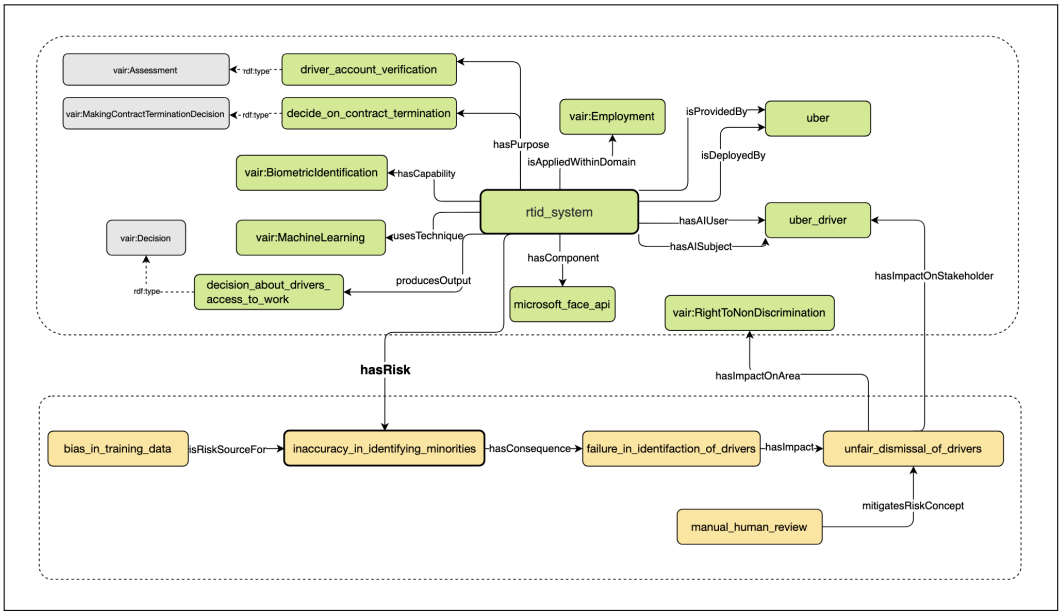


Figure 4.4: Visual representation of the graph specifying the Uber's Real-time ID Check use case

4.4.3 VioGén Domestic Violence Risk Assessment System

The VioGén System is used by the Spanish law enforcement agencies to determine the victim's eligibility for police protection by assessing the likeli-

[15] https://github.com/DelaramGlp/airo/blob/main/usecase/uber.ttl

hood of a victim of gender violence to be assaulted by the same perpetrator again [65]. Its use of statistical models to predict the risk faced by a victim raised questions regarding the accuracy of its predictions since these would be highly dependent on the quality of data fed into the models. The input data was generated based on a questionnaire answered by victims who filed a report. The ambiguity of questions and timing of questionnaire could have lead to inaccurate or biased predictions, and if the score was not modified by police officers, the victims would not receive the police protection they need. To address this risk, police officers were granted the power to increase the risk score calculated by the system. However, according to [183], in most cases the officers trusted the system's scoring despite warning signs, which led to "automation bias", i.e. over-reliance on the system's outcomes. Under the AI Act, VioGén is likely to be high-risk as it is used by the police to assess the"risk of a natural person becoming the victim of criminal offences" (Annex III, Point 6(a)). It also can fall under the high-risk AI systems used by public authorities to "evaluate the eligibility of natural persons for essential public assistance" (Annex III, Point 5(a)), in this case police protection. The graph expressing the VioGén use case is provided online on GitHub[16] and is illustrated in Figure 4.5.

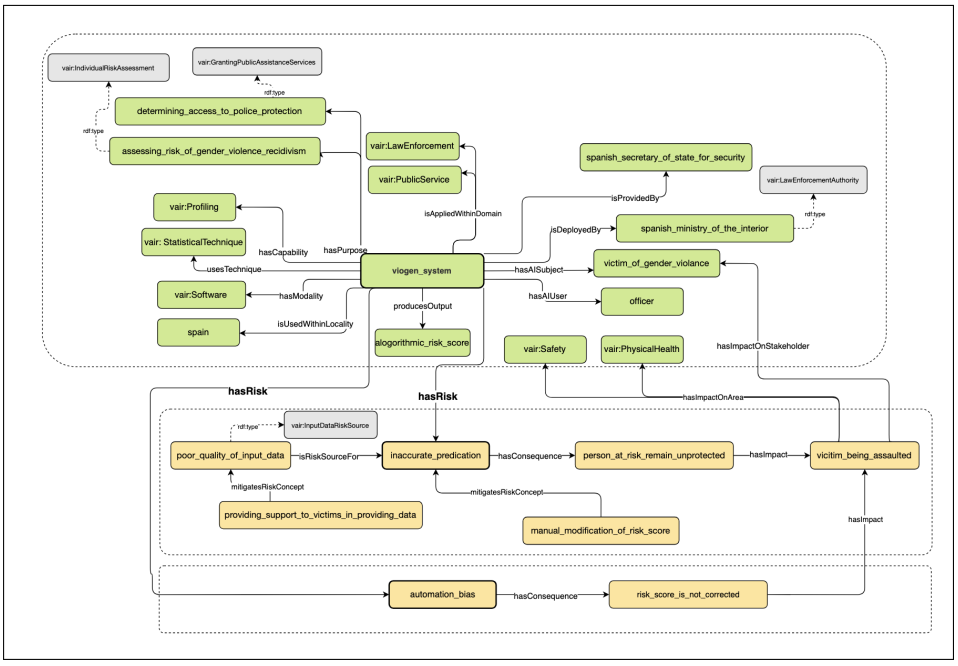


Figure 4.5: Visual representation of the graph specifying the VioGén use case

[16] https://github.com/DelaramGlp/airo/blob/main/usecase/viogen.ttl

4.5 Discussion of the Benefits and Potential Applications of AIRO and VAIR

As shown in the previous section, AIRO and VAIR enable modelling and maintaining information regarding AI use cases and their associated risks in an interoperable and queryable format. The three use cases demonstrated the capability of the ontologies in modelling real-world AI use cases, in a way that satisfies the competency questions. However, the modelling relies on the information available about the AI use case and its associated risks.

It should be noted that AIRO and VAIR are insufficient for creating technical documentation, as per Annex IV, given that they only support modelling a *subset* of the information elements that should be featured in technical documentation. Notably, AIRO and VAIR do not support modelling activities and processes across the AI lifecycle, which need to be demonstrated for legal compliance. This is particularly important in cases where an obligation entails provenance of plans as well as executed activities, and therefore requires documentation at both ex-ante and ex-post phases [184]. An example of such obligations in the AI Act is the post-monitoring system requirements (Article 72).

As shown in Subsection 4.3.2, there are not considerable number of specialisations for risk concepts within VAIR, due to (i) the lack of information about AI risks in the knowledge sources this thesis investigated and (ii) the context-dependant nature of AI risks. This is not particularly a limitation as addition of AI risk specialisations is supported through use of open and standardised formats.

In terms of benefits, AIRO and VAIR lay the ground for development of RegTech solutions that facilitate tasks related to compliance with the AI Act including determining Annex III high-risk AI systems, generating and exchanging AI and risk documentation, and expressing policies for using AI systems, as will be demonstrated in the next chapter.

Utilising AIRO and VAIR for modelling AI incidents helps with classification, collation, and comparison of AI risks and impacts over time. This can be helpful in addressing the gaps that exist between the ongoing AI regulation and standardisation activities and real-world AI incidents. Both AIRO and VAIR are provided as open and extensible resources, enabling accommodation of sector-specific requirements, as well as the requirements arising from the highly-anticipated European Commission-issued guidelines, future amendments to the AI Act (via delegated acts), and case laws.

AIRO has the potential to function as a basis for a minimal pan-European

AI vocabulary[17] to help establish a common language across different actors involved in the AI ecosystem. This consistency in the language, accompanied by a machine-readable representation, promotes interoperability and streamlines the information exchange required for incident reporting, compliance checking, and sharing best practices. As an added advantage, these semantic models can further be improved to evolve into multilingual ontologies supporting official EU languages. These ontologies can be implemented by adopting multilingual labels in RDF graphs for annotating concepts and their definitions with language-tagged strings [185].

In a broader context, AIRO and VAIR could be helpful for AI providers and deployers that operate in different jurisdictions in addressing challenges of cross-border compliance and interoperability by providing an extensible and adaptable structure to maintain information. AIRO and VAIR also can be useful resources in development of risk management catalogues, such as the CEN/CLC's undergoing project on risk management catalogues.

[17] See existing examples of EU vocabularies here: https://op.europa.eu/en/web/eu-vocabularies

CHAPTER

FIVE

SEMANTIC WEB-BASED APPROACHES TO SUPPORT THE AI ACT

The previous chapter introduced AIRO and VAIR as ontologies for representing and maintaining information about AI use cases as knowledge graphs. This chapter demonstrates how through utilisation of Semantic Web technologies these ontologies can assist with compliance with the requirements of the AI Act. To address RO3(a), Section 5.1 shows how the rules for determining Annex III high-risk AI systems can be modelled using the SHACL shapes constraint language [52]. Section 5.2 proposes an extension of the ODRL policy language [53] for expressing AI systems' intended purposes as use policies to address RO3(b). Section 5.3 demonstrates the use of SPARQL queries [51] for retrieving information to generate documentation (RO3(c)). Finally, Section 5.4 provides an extension of DCAT [54] for cataloguing AI systems for registering into the EU database (RO3(d)).

5.1 SHACL for Determining Annex III High-Risk AI Systems

Based on the analysis of Annex III high-risk AI systems (discussed in Subsection 3.2.1), this section explores the extent these classification rules can be expressed using Semantic Web technologies, addressing RO3(a). To this end, multiple Semantic Web languages that offer rule-checking capabilities can be employed, including the Shapes Constraint Language (SHACL) [52], the Semantic Web Rule Language (SWRL) [186], N3 (Notation3) rules [187], and the Shape Expressions (ShEx) language [188]. Among these, the SHACL constraint language is preferred to automate reasoning required for deter-

mining Annex III high-risk AI systems since it is a W3C recommendation. Further, it always produces a Boolean output and the results of validation process (rule-checking in the context of this section) can be annotated to link outputs to the related Annex III clause, to increase the explanatory power of the checks. Since SHACL uses closed-world assumption [189], it enables expression of rules to ensure the required information is present.

The rules for determining whether or not a particular use of an AI system is qualified as high-risk as per Annex III are expressed in a shape graph that contains SHACL shapes for each of the high-risk conditions. These shapes express different combinations of the 5 concepts identified in Subsection 3.2.1, depicted in Figure 5.1.

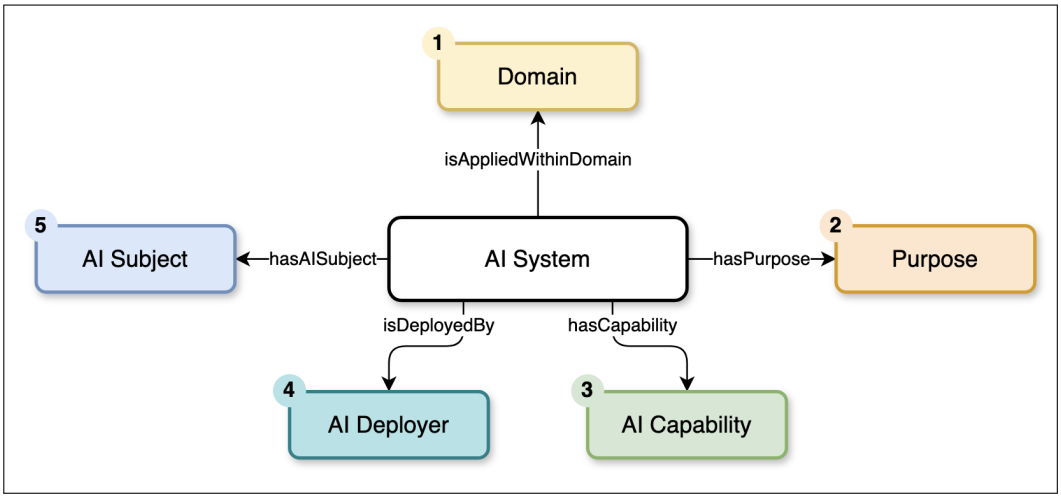


Figure 5.1: Semantic model of the 5 concepts required for determining high-risk AI systems as per Annex III

As an example, consider the high-risk condition described under *law enforcement* in Annex III, Point (6e) as follows: "*AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of the detection, investigation or prosecution of criminal offences*". Figure 5.2 depicts an illustration of the conceptual model of the high-risk rule drawn up in aforementioned clause using the Chowlk visual notation [190]. The pseudocode of the high-risk rule for Annex III, Point (6e) is also expressed in Algorithm 1.

SHACL's validation report is utilised to support generation of annotation that describes the Annex III clause(s) based on which a use case is likely to be deemed as high-risk. As detailed validation reports are generated in case

Algorithm 1: Pseudocode to determine whether an AI system is high-risk as per Annex III, Point (6e)

```
highrisk ← False
if system.Domain ∈ vair:LawEnforcement then
    if system.Purpose ∈ (vair:DetectingCriminalOffences ∪
      vair:InvestigatingCriminalOffences ∪
      vair:ProsecutingCriminalOffences) then
        if system.AICapability ∈ vair:Profiling then
            if system.AIDeployer ∈ (vair:LawEnforcementAuthority
              ∪ vair:LawEnforcementAuthorityAgent ∪
              vair:EUInstitution ∪ vair:EUAgency ∪ vair:EUOffice ∪
              vair:EUBody) then
                if system.AISubject ∈ vair:NaturalPerson then
                    highrisk ← True
```

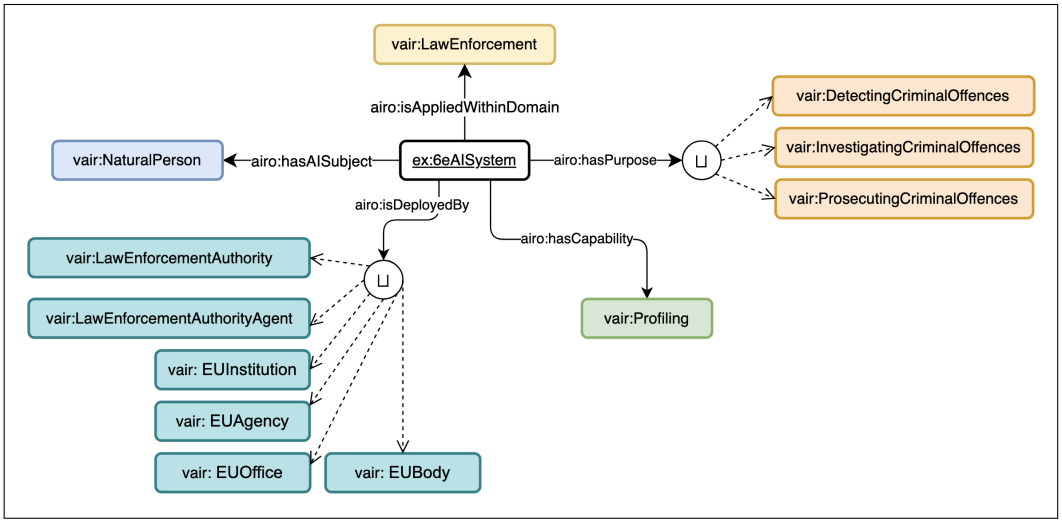


Figure 5.2: High-risk AI condition as per Annex III, Point 6e

of non-compliance with the shapes graph, the shapes are encoded in a way that they fail when the high-risk criteria are met. Therefore, all the target shapes are expressed as negation of the high-risk rules using `sh:not`, given that non-compliance with negation of a statement is equivalent to compliance with that particular statement. When applying the shapes graph over use cases, satisfying the constraints expressed within the `sh:not` leads the shape to fail. As the result of this failure, a validation report is generated

providing additional information in a `sh:ValidationResult`, which includes a `sh:resultMessage` providing annotation that assists in identification of the source in the legal text.

Each of the rules for high-risk AI, detailed in Appendix C, is encoded using `sh:NodeShape`, against which data graphs describing AI systems can be validated to determine whether they are qualified as high-risk under Annex III of the AI Act. To ensure that the model correctly detects all the clauses of Annex III, each shape was tested during implementation using two synthetic use case specifications: one that satisfies the not condition (i.e. is high-risk as per the clause under investigation) and another that does not. Listing 1 shows an example of a SHACL shape expressing the high-risk use described in Annex III, Point (6e). Figure 5.3 depicts the validation report generated by the SHACL Playground validator[1] for an example input use case that satisfies the conditions of the aforementioned clause[2].

Validation Report (1 results)

```
[
        a sh:ValidationResult ;
        sh:resultSeverity sh:Violation ;
        sh:sourceConstraintComponent sh:NotConstraintComponent ;
        sh:sourceShape ex:AnnexIII-6e ;
        sh:focusNode <https://example.com/myexample#example1> ;
        sh:value <https://example.com/myexample#example1> ;
        sh:resultMessage "High-Risk as per AI Act Annex III-6e: Law enforcement, AI
systems intended to be used by or on behalf of law enforcement authorities or by
Union institutions, bodies, offices or agencies in support of law enforcement
authorities for the profiling of natural persons as referred to in Article 3(4) of
Directive (EU) 2016/680 in the course of the detection, investigation or
prosecution of criminal offences."@en ;
] .
```

Figure 5.3: `sh:ValidationResult` for a use case that meets the high-risk use of AI specified in Annex III, 6(e)

The shapes were implemented for both the Council's Common Position [29] and the final version of the AI Act [5][3]. This indicated the ability to check a given use case against multiple set of SHACL checks, as an important feature of using SHACL for determining high-risk AI systems. This is particularly useful when new categories of high-risk AI introduced by the Commission to evaluate the influence of the proposed changes on high-risk and

[1] https://shacl.org/playground/

[2] See the example at https://github.com/DelaramGlp/airo/blob/main/high-risk-shacl/example-6e.ttl

[3] The SHACL shape graphs are available online at https://github.com/DelaramGlp/airo/tree/main/high-risk-shacl

non-high-risk AI systems that are already available on the market. However, a caveat of the proposed approach is its reliance on the correct specification of the AI use case under investigation using AIRO and VAIR.

Based on this, a simple web-based tool to assist in determining high-risk uses of AI was developed (see Figure 5.4 for its user interface). To determine high-risk AI systems, the user needs to provide the value for 5 concepts from a list of instances, populated from VAIR. Based on the user's input, an RDF graph expressing the AI system is generated and then it is validated against the SHACL shapes to determine if conditions for high-risk AI are met. The output of the tool, shown in Figure 5.5, includes the result of the assessment (high-risk or not high-risk) and an assessment report. This tool can be accessed online at https://regtech.adaptcentre.ie/highrisk and its source code, along with the instructions to run the tool, is available on GitHub[4] under the MIT license[5].

Compared to the state of the art approaches for classification of AI systems under the AI Act, reviewed in Section 2.4, this thesis provides a novel codified approach for determining Annex III high-risk AI systems using 5 key concepts. A key feature of this approach is the ability to produce classification reports that can be investigated and compared. Moreover, utilising SHACL shapes provides a structured, transparent, interoperable, and adaptable framework for specifying high-risk AI rules. Compared to the available tools on the market, the developed web application is free to use, without requiring users to share any personal information. Additionally, its source code is available under a permissive software license, allowing further enhancements and modifications, including addition of rules for prohibited AI practices and Annex I high-risk AI systems.

[4] https://github.com/DelaramGlp/highrisk_app
[5] https://mit-license.org

```
1
2  ex:AnnexIII-6e
3      a sh:NodeShape ;
4      sh:targetClass airo:AISystem ;
5      sh:message "High-Risk as per AI Act Annex III-6e: Law enforcement, AI
       ↪  systems intended to be used by or on behalf of law enforcement
       ↪  authorities ... for the profiling of natural persons ...in the
       ↪  course of the detection, investigation or prosecution of criminal
       ↪  offences."@en ;
6      sh:description "AI systems for detection, investigation or
       ↪  prosecution of criminal offences"@en ;
7      sh:not [sh:and (
8                  sh:property [
9                      a sh:PropertyShape ;
10                     sh:path airo:isAppliedWithinDomain ;
11                     sh:class vair:LawEnforcement ; ]
12                 sh:property [
13                     a sh:PropertyShape ;
14                     sh:path airo:hasPurpose ;
15                     sh:or (
16                         [ sh:class vair:DetectingCriminalOffences ; ]
17                         [ sh:class vair:InvestigatingCriminalOffences; ]
18                         [ sh:class vair:ProsecutingCriminalOffences; ] )
                           ↪  ]
19
20                 sh:property [
21                     a sh:PropertyShape ;
22                     sh:path airo:hasCapability ;
23                     sh:class vair:Profiling ; ]
24
25                 sh:property [
26                     a sh:PropertyShape ;
27                     sh:path airo:isDeployedBy ;
28                     sh:or (
29                         [ sh:class vair:LawEnforcementAuthority ; ]
30                         [ sh:class vair:LawEnforcementAuthorityAgent ; ]
31                         # omitted other deployers here for brevity
32                     ) ]
33
34                 sh:property [
35                     a sh:PropertyShape ;
36                     sh:path airo:hasAISubject ;
37                     sh:class vair:NaturalPerson ; ] ) ] .
38
```

Listing 1: SHACL shape for identifying high-risk AI Systems from Annex III, Point 6e of the AI Act

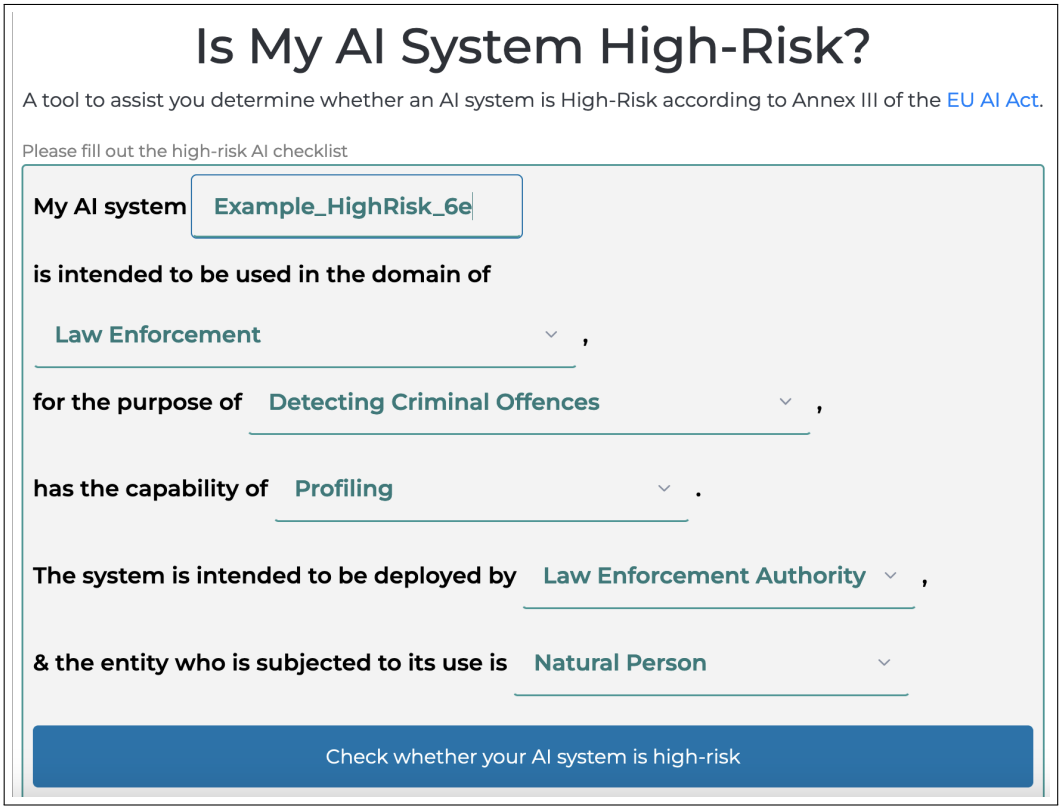


Figure 5.4: User interface of the tool developed for determining high-risk AI

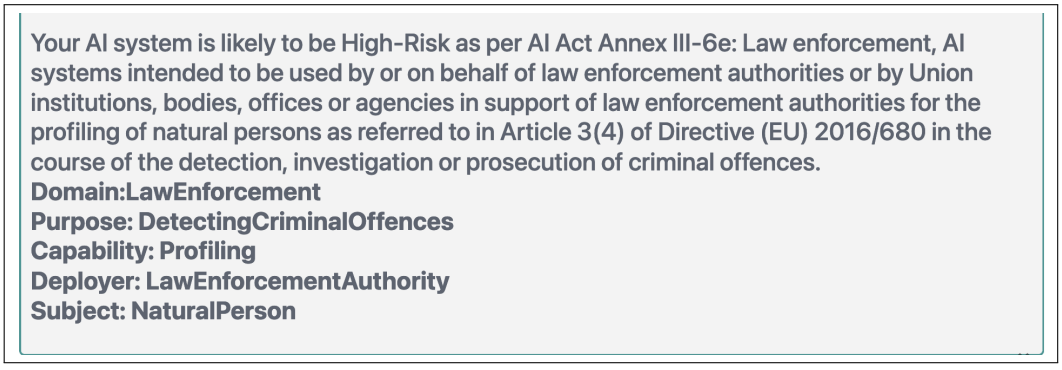


Figure 5.5: The results shown for an AI system meets Annex III, 6e conditions

5.2 AIUP - an ODRL Profile for Expressing AI Use Policies

The previous discussion in Section 3.3 highlighted the importance of *intended purpose* in risk assessment and further proposed specifying this concept within *AI use policies*. To reiterate, the following were identified as key constituting elements of an AI use policy:

1. **Intended purpose(s)** of the AI system,
2. **Precluded use(s)** of the AI system,
3. **Human oversight measure(s)** that should be implemented by the AI deployer,
4. AI deployer's **risk reporting obligation(s)**.

To include the aforementioned information within a policy, this section proposes the AI Use Policy (AIUP) profile as an extension of the W3C's recommendation on Open Digital Rights Language (ODRL) [53]. Based on the previous analysis, the requirements of AIUP are specified in Table 5.1 using the requirements specification convention established in [191].

5.2.1 AIUP Overview

As discussed in Section 3.3, policies regarding intended and precluded uses of AI can be expressed using 7 concepts, namely, domain, purpose, AI capability, AI deployer, AI subject, data, and locality of use. These concepts therefore are included in the AIUP profile as left operands, enabling describing use policies by assigning of constraints to each of them.

AIUP introduces 3 types of `aiup:UsePolicy`, that are `aiup:UseOffer`, `aiup:UseRequest`, and `aiup:UseAgreement`. These, in addition to the aforementioned left operands, enable expressing offers, requests, and agreements from/between AI providers and deployers. By inclusion of different types of `aiup:AIComponent`, the profile also allows expressing use policies for general-purpose AI models, as required by Annex XI of the AI Act. However, for the actors involved in defining, negotiating, and using such policies, it only includes general concepts of `aiup:Provider` and `aiup:Deployer`.

While `odrl:eq` enables checking equivalency, i.e. a given value equals the right operand of the Constraint, `odrl:isA` allows indicating that a given value is an instance of the right operand (`rdf:type`). In the ODRL specification, it is not clear whether `odrl:isA` is also an indicator of a value

Table 5.1: AIUP profile requirements specification

AIUP Profile Requirements Specification Document
1. Purpose
The purpose of this ODRL profile is to support specification of policies that express the conditions of using an AI system aligned with the requirements of the EU AI Act.
2. Scope
The scope of AIUP is limited to policies specifying intended purpose(s) and precluded use(s) of AI systems. In terms of the conditions for using an AI system as intended, the scope is limited to specification of human oversight measures and risk reporting requirements.
3. Implementation Language
OWL, SKOS, ODRL
4. Key Uses
USE 1. Expressing and communicating use policies, including those that describe intended purposes, conditions of use, and precluded uses by AI providers. USE 2. Comparing multiple AI use policies. USE 3. Expressing AI agreements between AI providers and deployers USE 4. Investigating and auditing compliance with AI use policies. USE 5. Identifying misuses of the system based on the agreements.
5. Ontology Functional Requirements
a. Non-Functional Requirements
NFR 1. The profile shall be published online with standard documentation.
b. Functional Requirements
CQ1. What is the intended purpose(s) of the AI system? CQ2. What is the precluded use(s) of the AI system? CQ3. To use the system as intended, what human oversight measures should be implemented by the AI deployer? CQ4. What are the reporting obligations of the AI deployer?

being a sub-class of (`rdfs:subClassOf`) or semantic equivalence of (e.g. `skos:broader`) the right operand[6]. An example of such a use in AIUP is where the purpose is one of the sub-classes or instances of `vair:ServingSafetyFunction`. To address the ambiguity around the function of `odrl:isA`, semantic equality (`aiup:seq`) was added as an Operator to indicate presence of either "instance of" or "sub-class of" relationships.

The development of AIUP followed the ODRL V2.2 Profile Best Practices [192], which requires the terms to be defined in the policy namespace with `skos:exactMatch` to link the proposed terms to existing vocabular-

[6]See the related GitHub issue that has raised this matter here: https://github.com/w3c/odrl/issues/28

ies. Therefore, the concepts are specified within the AIUP namespace (aiup) and further aligned with concepts form AIRO and DPV's TECH extension through use of `skos:exactMatch`, as shown in Table 5.2.

An overview of the AIUP profile is illustrated in Figure 5.6. Expressing intended purposes and precluded uses of an AI system or component within a policy are enabled by employing `odrl:permission` and `odrl:prohibition` rules, respectively. For expressing the conditions of use, i.e. obligations that should be fulfilled by a party in order to use the system or component, the `odrl:duty` property should be employed.

The documentation of the AIUP profile was generated using WIDOCO [175] and is accessible online at https://w3id.org/aiup under the CC-BY-4.0 license.

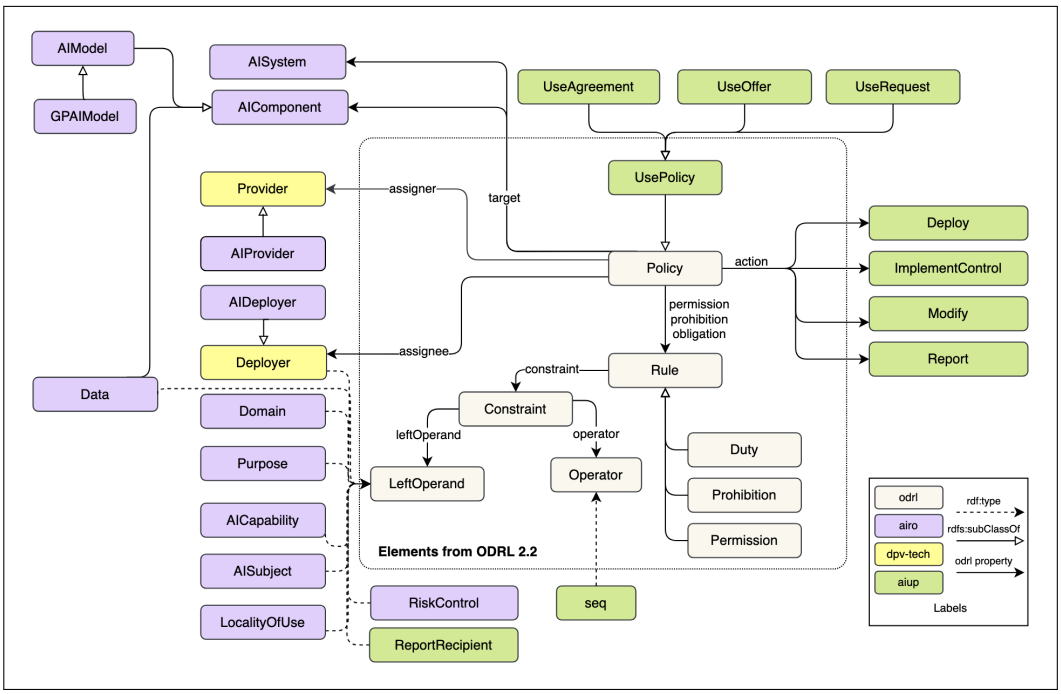


Figure 5.6: AIUP core classes and properties

Table 5.2: Key elements of AIUP profile

Type	AIUP term	skos:exactMacth
odrl:Policy	aiup:UsePolicy	N/A
	aiup:UseOffer	N/A
	aiup:UseRequest	N/A
	aiup:UseAgreement	N/A
odrl:Asset	aiup:AISystem	airo:AISystem
	aiup:AIComponent	airo:AIComponent
	aiup:AIModel	airo:AIModel
	aiup:GPAIModel	airo:GPAIModel
	aiup:Data	airo:Data
odrl:Action	aiup:Deploy	N/A
	aiup:Modify	N/A
	aiup:ImplementControl	N/A
	aiup:Report	N/A
odrl:Party	aiup:Provider	tech:TechnologyProvider
	aiup:Deployer	tech:TechnologyDeployer
	aiup:AIProvider	airo:AIProvider
	aiup:AIDeveloper	airo:AIDeveloper
odrl:LeftOperand	aiup:Domain	airo:Domain
	aiup:Purpose	airo:Purpose
	aiup:AICapability	airo:AICapability
	aiup:Deployer	tech:TechnologyDeployer
	aiup:AISubject	airo:AISubject
	aiup:Data	airo:Data
	aiup:LocalityOfUse	airo:LocalityOfUse
	aiup:RiskControl	airo:RiskControl
	aiup:ReportRecipient	N/A
odrl:Operator	aiup:seq	N/A

5.2.2 Proof-of-Concept and Potential Benefits

To demonstrate how AIUP can be applied, a policy for Proctify, the example use case described in Subsection 4.4.1, is expressed in the following. Listing 2 presents a use policy, expressing the conditions of deploying Proctify in a form of an `aiup:UseOffer`. For brevity, the policy only includes 3 constraints for describing the intended domain, purpose, and subjects. Further, the use offer indicates that the deployer should provide training to users of the system as a control to address the risk of over-reliance of users on the system's output. As will be shown in Subsection 5.4.2, the policies regarding use of Proctify and its components can be included in AI catalogues.

In terms of potential benefits, the AIUP profile supports modelling and comparison of policies in different stages of procurement. Given that the AIUP profile can be extended according to the EU's draft for standard contractual clauses for AI procurement [193], it can potentially be used by the

```
:proctify-offer-01 a aiup:UseOffer ;
  odrl:uid :proctify-offer-01 ;
  odrl:profile aiup: ;
  rdfs:comment "Offer for using Proctify"@en ;
  odrl:permission [
      a odrl:permission ;
      odrl:assigner :aiedux ;
      odrl:target :proctify ;
      odrl:action aiup:Deploy ;
      odrl:constraint [
      odrl:and  [
              odrl:leftOperand aiup:Domain ;
              odrl:operator aiup:seq ;
              odrl:rightOperand vair:Education  ] ,
       [
              odrl:leftOperand aiup:Purpose ;
              odrl:operator aiup:seq ;
              odrl:rightOperand
              ↪  vair:DetectingProhibitedBehaviourDuringTest  ] ,
       [
              odrl:leftOperand aiup:AISubject ;
              odrl:operator aiup:seq ;
              odrl:rightOperand vair:Student  ]  ] ;
      odrl:duty [
          dct:title "User training to address over-reliance" ;
          odrl:action aiup:ImplementControl ;
          odrl:constraint [
              odrl:leftOperand aiup:RiskControl  ;
              odrl:operator aiup:seq ;
              odrl:rightOperand vair:UserTraining  ]  ]  ] .

```

Listing 2: An example of `aiup:UseOffer` describing Proctify's use policy

public sector for expressing contractual arrangements when procuring AI solutions. AIUP also assists AI auditors and authorities in investigation of non-compliance and ascertaining liable parties when investigating claims concerning AI systems. When presented as an element within an AI catalogue (see Section 5.4), discovery and filtering of AI systems or components based on use policies would become possible.

Compared to the state of the art, reviewed in Subsection 2.4.2, AIUP positions itself as a novel approach for specifying AI use policies through fostering the ODRL policy language.

```
SELECT  ?system ?domain ?purpose ?capability ?deployer ?subject
    WHERE { ?system a airo:AISystem;
                    airo:isAppliedWithinDomain ?domain ;
                    airo:hasPurpose ?purpose ;
                    airo:hasCapability ?capability ;
                    airo:isDeployedBy ?deployer ;
                    airo:hasAISubject ?subject . }
```

Listing 3: SPARQL `SELECT` query to retrieve the 5 key concepts for determining Annex III high-risk AI from an AIRO-based representation of a use case

5.3 SPARQL Queries for Retrieving the Information Featured in Documentation

To show usefulness of AIRO and VAIR in automated generation of documentation, SPARQL queries [51] are utilised in order to retrieve information from an AIRO and VAIR-based specification of an AI use case. In doing so, the competency questions established in Subsection 3.4.2, which also represent the elements of the AI Cards (will be introduced in Chapter 6), are considered in implementation of the queries.

The information retrieval to generate each theme within the set of competency question is mainly implemented using `SELECT` queries. An example of such queries is shown in Listing 3, where the query retrieves the information related to identification of Annex III high-risk AI systems.

In high-level compliance checking and auditing of AI systems, verification of presence of information rather than provision of that information might be required; for example, verifying if there are any measures in place to address the impacts on the fundamental right to non-discrimination. For this, SPARQL `ASK` queries can be utilised, returning a Boolean value indicating whether the query pattern has any matches. An example of determining if there are any logging measures in place that addresses harms to the right to non-discrimination is encoded using an `ASK` query in Listing 4.

One of the most significant advantages of using consistent and standardised formats is the ability to easily integrate and collate information provided by multiple sources. In the context of documentation, when components' documentation is provided in standardised linked data formats, such as JSON-LD, XML, and Turtle, information about incorporating components can be retrieved from their documentation and integrated with an AIRO-based specification of an AI system. As an example of how this can be implemented,

```
ASK {
   ?system a airo:AISystem ;
           airo:hasRisk ?risk .
   ?risk a airo:Risk ;
       airo:hasConsequence ?consequence .
   ?consequence a airo:Consequence ;
       airo:hasImpact ?impact .
   ?impact a airo:Impact ;
       airo:hasImpactOnArea vair:RightToNondiscrimination .
    ?control  a airo:RiskControl, vair:LoggingMeasure .
    ?riskRelation rdfs:subPropertyOf* airo:modifiesRiskConcept .
       { ?control  ?riskRelation  ?risk .}
       UNION { ?control ?riskRelation  ?consequence . }
       UNION { ?control ?riskRelation  ?impact . }   }
```

Listing 4: SPARQL `ASK` query to determine if there are any logging measures are in place to address harmful impacts to the right to non-discrimination using AIRO and VAIR

the Python script presented in Listing 5 demonstrates retrieval of information from the JSON-based HuggingFace's Model Cards. The script uses the models' ID from the HuggingFace Model Hub, which is indicated in the system's specification using `airo:hasDocumentation` relation. In cases where components' documentation is presented in a form of a standardised knowledge graph with SPARQL endpoints available for querying, federated queries using SPARQL `SERVICE` can be executed to retrieve and combine information from multiple sources.

The set of SPARQL queries for generating documentation is available online on GitHub[7]. As mentioned before, this thesis only supports modelling and retrieval of a subset of the information that should be featured in the AI Act's technical documentation (this subset will be later represented within the AI Cards documentation framework in Chapter 6). However, the overall approach of querying discussed in this section can be further used to generate the technical documentation and retrieve information for demonstration and investigation of legal compliance.

[7] https://github.com/DelaramGlp/aicards/tree/main/sparql_queries

```python
import rdflib
import json
from huggingface_hub import ModelCard

g = rdflib.Graph()
g.parse("/huggingFaceExample.ttl" , format="turtle")
airo = "https://w3id.org/airo"
g.parse(airo, format="ttl")
model_query = """
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX airo: <https://w3id.org/airo#>
SELECT ?model ?doc
    WHERE {
    ?system a airo:AISystem ;
            airo:hasModel ?model .
    ?model airo:hasDocumentation ?doc . } """

for row in g.query(model_query):
    card = ModelCard.load(str(row.doc))
    card_json_string = json.dumps(card.data.to_dict(), indent=-2)
    card_json = json.loads(card_json_string)
    license = card_json["license"]
    print("HuggingFaceModel:", row.doc, "license:", license)
```

Listing 5: Python script showing retrieval of information from HuggingFace's Model Cards linked to AI documentation

5.4 AICat - a DCAT Extension for Cataloguing AI Systems

As indicated in Section 3.5, Article 49 requires providers and deployers of Annex III high-risk AI system, with the exception of those systems that fall under the critical infrastructure, to register their systems into the EU database. This requirement is also applied to those providers that declare their system, which meets Annex III conditions, as not high-risk on the basis of derogation.

The EU database shall be *"accessible and publicly available"* (with some exceptions), provided in a *"user friendly manner"*, and should be *"easily navigable and machine-readable"* (Article 71(4)). Implementation of these desired features requires the EU database to be supported by a layer of metadata. To address this need, which was reflected in RO3(d), this section aims to demonstrate how AIRO and VAIR can be applied in combination with the Data Catalog Vocabulary (DCAT) [54] to support governance of AI registries, including the EU database.

The key information elements identified from the AI Act's registration obligations, discussed in Section 3.5, shape the functional requirements of AICat. These requirements, which are expressed in the form of competency questions following the methodology described in [191], are shown in Table 5.3.

5.4.1 AICat Overview

AICat extends DCAT version 3 [54], since this version of DCAT supports cataloguing resources beyond datasets. By extending DCAT, AICat aims to scale the cataloguing to include AI systems and models to address the needs of the EU database. Table 5.4 illustrates how the identified requirements are mapped into concepts from DCAT, AIRO, DPV, and DPV's TECH extension. As shown in the table, intended purpose is represented as a policy modelled using the AIUP profile (proposed in Section 5.2).

Figure 5.7 depicts an overview of AICat's information model. As illustrated in the figure, `aicat:Catalog` is a sub-class of `dcat:Catalog` that provides a curated collection of metadata about AI systems, models, and datasets. AICat extends DCAT by introducing `airo:AISystem` and `airo:AIModel` as sub-classes of `dcat:Resource`, enabling inclusion of their metadata in an `aicat:Catalog`. Given that `airo:Data` is a sub-class of `dcat:Dataset`, cataloguing data is also supported by AICat. While the inclusion of AI

Table 5.3: AICat profile requirements specification

AICat Requirements Specification Document		
1. Purpose		
The purpose of the AICat profile is to use DCAT and AIRO to describe catalogues of AI systems and their associated components, such as datasets and AI models.		
2. Scope		
The scope of AICat is limited to the *atomic* information that should be submitted upon the registration of high-risk AI systems into the EU database, outlined in Annex VIII. This means that descriptive information, for instance the system's logic and findings of the fundamental rights impact assessment, is not included in the scope.		
3. Implementation Language		
OWL, DCAT		
4. Key Uses		
USE 1. Maintaining and managing metadata about AI systems, datasets, and models in interoperable and standardised catalogues. USE 2. Discovering and comparing AI solutions. USE 3. Cataloguing and sharing information about AI systems with the public in a transparent manner. This includes the use by the European Commission for sharing metadata of the high-risk AI systems indexed in the EU database.		
5. Ontology Requirements		
a. Non-Functional Requirements		
NFR 1. AICat shall be published online with standard documentation. NFR 2. AICat shall reuse concepts and relations from AIRO to the fullest extent possible.		
b. Functional Requirements: Groups of Competency Questions		
CQG1. AI systems	**CQG2. Datasets**	**CQG3. AI models**
CQ1-1. What is the name of the system? CQ1-2. Who is the system's provider? CQ1-3. Who is the system's deployer? CQ1-4. What is the system's intended purpose? CQ1-5. What is the system's market availability status? CQ1-6. In which countries is the system made available? CQ1-7. What are the additional references to the system?	CQ2-1. Which datasets are used by the system? CQ2-2. What is the system's input data? CQ2-3. What is the dataset's use policy?	CQ3-1. Which models are used by the system? CQ3-2. What is the model's release data? CQ3-3 What is the model's input data? CQ3-4 What is the model's output data? CQ3-5. What is the model's license? CQ3-6. What is the model's use policy?

systems was directly linked to the scope of the EU database, whose aim is to index AI systems, the inclusion of models and datasets was driven by the existing focus in the state of the art on cataloguing these AI components, as reviewed in Subsection 2.4.4. To enable modelling the relationships between the resources, for example to show which datasets were used for training a model, `airo:hasTrainingData`, `airo:hasTestingData`, `airo:hasValidationData`, `airo:hasInput`, `airo:hasOutput`, and `airo:hasModel` are reused from AIRO.

AICat's documentation was generated using WIDOCO [175] and is available online at https://w3id.org/aicat under the CC-BY-4.0 license.

Table 5.4: Specifications for representing AI systems and models in AICat

CQ	AI Act Annex	Requirement	Metadata Field	Range
Information about **AI system**				
1-1	VIII, A4 & B4	AI system's trade name	`dct:title`	`rdfs:Literal`
1-2	VIII, A1 & B1	Provider's information	`airo:isProvidedBy`	`airo:AIProvider`
1-3	VIII, C1	Deployer's information	`airo:isDeployedBy`	`airo:AIDeployer`
1-4	VIII, A5 & B5	AI system's intended purpose	`odrl:hasPolicy`	`aiup:UsePolicy`
1-5	VIII, A7 & B8	AI system's market status	`tech:hasMarketAvailabilityStatus`	`tech:MarketAvailabilityStatus`
1-6	VIII, A10 & B9	Countries where system is available	`dpv:hasCountry`	`dpv:Country`
1-7	VIII, A4 & B4	AI system's additional reference	`dct:isReferencedBy`	`dcat:Resource`
Information about **components**				
2-3	VIII, A5 & B5	Component's intended purpose	`odrl:hasPolicy`	`aiup:UsePolicy`
Information about **datasets**				
2-1	VIII, A6	Data used by the system or model	`airo:hasTrainingData`, `airo:hasValidationData`, `airo:hasTestingData`	`airo:Data`
2-2	VIII, A6	Input data used by the system	`airo:hasInput`	`airo:Data`
Information about **models**				
3-1	–	AI models used within the system	`airo:hasModel`	`airo:AIModel`
3-2	XII, 1-1c	Model's date of release	`dct:issued`	`xsd:date`
3-3	XII, 1-1g	Model's input data	`airo:hasInput`	`airo:Data`
3-4	XII, 1-1g	Model's output data	`airo:hasOutput`	`airo:Data`
3-5	XII, 1-1h	Model's license	`airo:hasLicense`	`airo:License`
3-6	XII, 1-1b	Model's use policy	`odrl:hasPolicy`	`aiup:UsePolicy`

By following DCAT-AP [144], AICat can further distinguish between *mandatory*, *recommended*, *optional*, and *deprecated* elements based on the requirements of the AI Act. Even though implementing such normative profiles can easily be realised by defining the aforementioned property types for

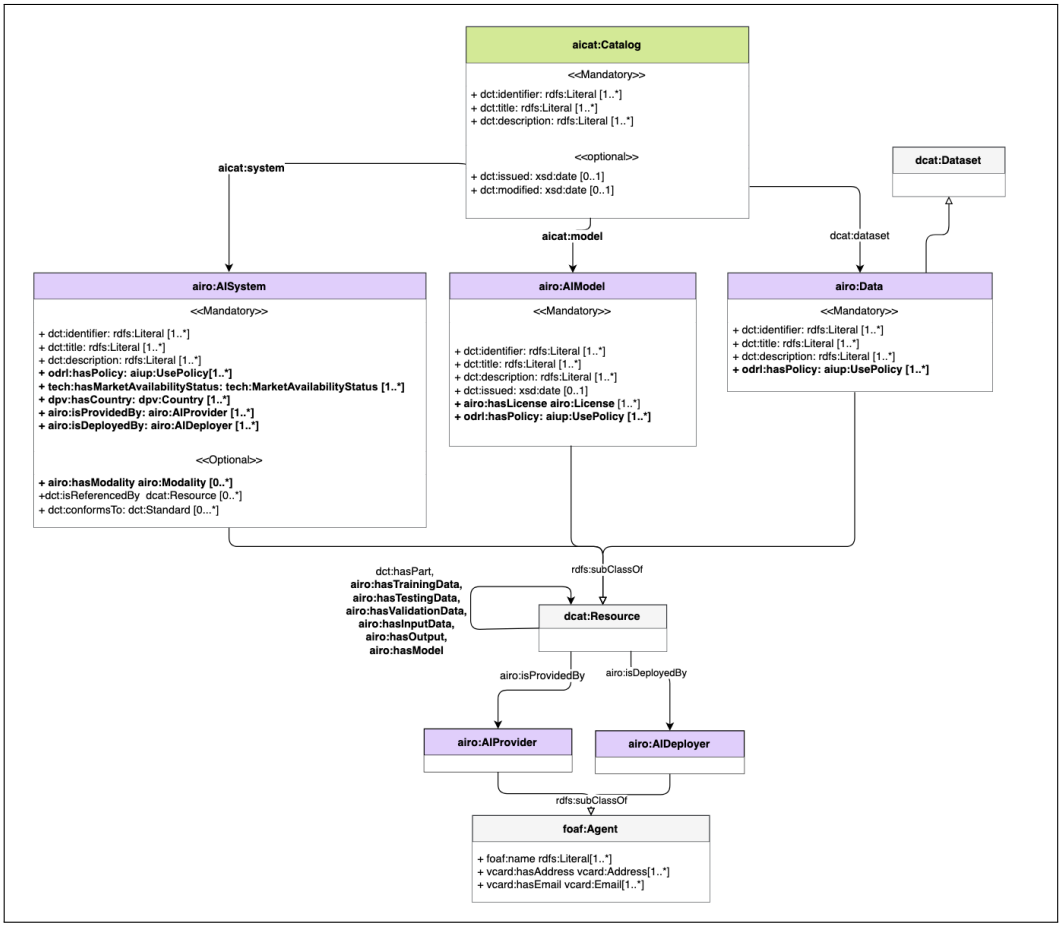


Figure 5.7: An overview of the AICat Profile

each of the information elements, in the context of the AI Act, identification of whether provision of an information element is mandatory, recommended, optional, or deprecated requires additional guidelines and codes of conduct.

AICat is introduced as a minimal extension of DCAT. This extension introduces the `aicat:Catalog` class and its relations with `airo:AISystem` and `airo:AIModel`, both added as new types of `dcat:Resource`. It is implemented under the *aicat* namespace to follow the principle of modularity and further enhance reuse. One of the key directions for improving AICat is using SHACL shapes to specify the level of necessity for information elements—that can be mandatory, recommended, or optional. Listing 6 shows an example of a SHACL shape indicating that each AI system should have at least one provider. Currently, AICat does not define such a normative profile due to the absence of recommendations and guidelines in regard to the AI Act.

```
<https://w3id.org/aicat#AIProviderShape> a sh:NodeShape;
                                         sh:targetClass airo:AISystem ;
                                         sh:property [
                                           a sh:PropertyShape ;
                                           sh:path airo:isProvidedBy;
                                           sh:minCount 1 ] .
```

Listing 6: Example of a SHACL shape that specifies the requirement for presence of at least one provider for an AI system

5.4.2 Proof-of-Concept and Potential Benefits

Listing 7 presents a summarised version of an `aicat:Catalog` that contains metadata about *Proctify* and its components. As shown in the listing, the policies for using the AI system and its incorporating components are expressed using the AIUP profile.

In terms of potential benefits, through reusing widely-used W3C standardised vocabularies, the AICat enables expressing metadata regarding AI systems and AI components within catalogues, wherein common vocabularies and open linked data-based formats are used. Therefore, the AICat addresses the AI market needs for a consistent and interoperable mechanism for cataloguing AI solutions [194], in a way that enables federated search and comparison across AI, model, and data catalogues offered by different vendors—a crucial feature often required in AI procurement processes. In relation to this, the European Commission's dataset of selected uses of AI in the public sector [195] is a prominent resource, whose interoperability and searchability can be enhanced through adoption of a cataloguing mechanism such as AICat.

At the organisational level, AICat could assist AI providers and deployers in providing structured catalogues of AI systems and components.At the European level, a similar approach to AICat is expected to be adopted for the implementation of the database of high-risk AI systems as required by Article 71 of the AI Act. Given that AICat ensures traceability while protecting privacy by providing metadata without revealing sensitive information within a dataset, it supports the implementation of the non-public section of the EU database and provides a structure for registration forms. AICat potentially addresses the gap in the European open data portal in providing FAIR information regarding existing AI systems and models provided or deployed by public organisations. AICat also has the potential to promote cross-border interoperability required by the recently-enforced Interopera-

```
:aieduX-catalogue-01 a aicat:Catalog, dcat:Catalog  ;
    dct:identifier "aiedux-cat01"^^xsd:string ;
    dct:title "AIEduX catalogue"@en ;
    dct:description "AI systems and models provided by AIEduX"@en ;
    dct:created "2024-05-05"^^xsd:date ;
    dcat:dataset :susbehaved_dataset ;
    aicat:model :susbehaved_model;
    aicat:system  :proctify .

:susbehaved_dataset a dcat:Dataset, airo:Data ;
   dct:identifier "aiedux-d012"^^xsd:string ;
   dct:title "SusBehavedDataSet"@en ;
   dct:description ".. includes suspicious behaviour data.."@en ;
   odrl:hasPolicy :susbehaved_dataset_policy  .

:susbehaved_model a dcat:Resource, airo:AIModel ;
    dct:identifier "aiedux-m022"^^xsd:string ;
    dct:title "SusBehavedModel"@en ;
    dct:description ".. determines suspicious behaviour .."@en ;
    dct:issued "2024-02-15"^^xsd:date ;
    airo:hasTrainingData :susbehaved_dataset ;
    odrl:hasPolicy :susbehavedmodel_policy .

:proctify a dcat:Resource, airo:AISystem ;
    airo:isProvidedBy :aiedux ;
    dct:identifier "aiedux-ai031"^^xsd:string ;
    dct:title "Proctify"@en ;
    dct:description "An AI-based proctoring system..."@en ;
    tech:hasMarketAvailabilityStatus tech:MarketAvailable ;
    dpv:hasCountry <http://dbpedia.org/resource/Italy> ;
    dcat:contactPoint <http://example.org/aieduX-AI031/contact> ;
    airo:hasModel :susbehaved_model ;
    odrl:hasPolicy :proctify_use_policy .

 :susbehaved_dataset_policy a aiup:UseOffer  .
 :susbehavedmodel_policy a aiup:UseOffer .
 :proctify_use_policy a aiup:UseOffer .
```

Listing 7: An example of `aicat:Catalog` for describing a catalogue describing *Proctify* and its components

ble Europe Act [196], particularly in the implementation of the *Interoperable Europe portal*—the EU's single point of entry for information related to cross-border interoperability of trans-European digital public services (Interoperable Europe Act, Article 8). In this, AICat can be employed to facilitate sharing information and best practices to support interoperability in public procurement of AI-based solutions.

Compared to existing cataloguing approaches, reviewed in Subsection 2.4.4, AICat expands the scope of cataloguing to AI systems. From this literature review, MLDCAT-AP [146] bears a close resemblance to AICat, especially in the use of DCAT. MLDCAT-AP has been supported by the European Commission's Semantic Interoperability Community (SEMIC), and therefore it might be a candidate to be adopted in the implementation of the EU database. However, prior to this, it needs to be extended to include specifications of AI systems in the catalogue in alignment with the requirements of the AI Act. This can be realised through the integration of MLDCAT-AP and AICat. Another key feature of MLDCAT-AP, in comparison with AICat, is the inclusion of risk information in the catalogue. While AICat can support DCAT-based documentation of risks by reusing `airo:hasRisk`, in its current form it does not go beyond the general, non-descriptive information elements of Annex VIII, mainly due to the absence of related official guidelines. Despite this, AICat offers sufficient evidence of the potential of AIRO, VAIR, and AIUP in cataloguing AI systems, as required for registration of AI systems into the EU database.

CHAPTER

SIX

AI CARDS

As discussed in Section 2.5, the existing landscape of AI documentation approaches exhibits three critical gaps: first, there is a lack of holistic view of AI systems that focuses on *intended purposes*. Second, information reflecting the efforts undertaken to manage AI risks, which is essential for AI transparency and accountability [2], is underrepresented. Third, the inconsistencies in the terminology used in the documentation approaches, in addition to their unstructured formats, hinder information sharing and interoperability.

This chapter addresses RO4 by proposing AI Cards as a documentation framework, providing a summary of key AI and risk information based on the AI Act's risk management system and technical documentation requirements. The AI Cards framework addresses the above-mentioned gaps by offering a structured approach to represent a given use of an AI system in both human- and machine-readable formats. The human-readable representation of AI Cards takes the form of a visual summary card, providing the wide range of AI stakeholders with a transparent, yet summarised overview of an AI use case. This design decision is made to address the challenges of sharing the AI Act's technical documentation, which arise due to its extensiveness and confidentiality of its content. The machine-readable representation of AI Cards builds upon AIRO and VAIR, ensuring the AI Cards flexibility and interoperability and further enabling its automated generation using SPARQL queries, as previously discussed in Section 5.3.

The rest of this chapter is organised as follows: Section 6.1 describes the AI Cards development process. Section 6.2 introduces the human-readable representation of the AI Cards framework and Section 6.3 presents its accompanying machine-readable specification. The chapter concludes with a discussion of the AI Cards' benefits and potentials in Section 6.5

6.1 AI Cards Development Process

The AI Cards framework was co-designed in consultation with researchers from the Joint Research Centre (JRC)—the European Commission's internal research services to support EU policies at various stages of the policy cycle. This collaboration primarily took place during a 6-week onsite secondment, as part of the PROTECT project (refer to Subsection 7.4.1 for more information). In this collaborative work, first, the requirements were identified (listed in Table 6.1). Then, the information elements were selected from the detailed analysis of technical documentation requirements (outlined in Appendix D) and the competency questions represented earlier in Subsection 3.4.2. The extensiveness of technical documentation and confidentiality of its content may hinder collaboration and communication with AI stakeholders—many of whom are not necessarily technical AI experts. Thus, in consultation with JRC experts a deliberate decision was made to create a summarised overview of a given use of an AI system, with the criteria for selection defined in Table 6.1.

At the first attempt, some of the selected information elements were not suitable for inclusion in a summary card due to their sensitivity or high-level of details. Therefore, condensed views for these information elements through visual aids were considered. After reaching a solid structure, the AI Cards was further revised based on the feedback received from an online anonymous survey, which will be discussed in Subsection 7.4.3. To ensure that RDF-based representation of AI Cards can be generated using AIRO and VAIR, both ontologies were reviewed and updated when solid representation of AI Cards was reached. Therefore, AI Cards served as an additional means for further evaluation and revision of AIRO and VAIR.

6.2 AI Cards Information Elements

Figure 6.1 shows a visualisation of the AI Cards, which condenses the information elements into 9 sections. To enable representing AI Cards in a time series and to allow version control required to reflect the evolving nature of AI systems, the AI Cards' metadata includes essential information describing its **version**, **issuance date**, **language**, **publisher**, and **contact** information. Further, in addition to the **URL of the machine-readable specification**, if the system is registered into the EU database, the **URL of the entry of the AI system in the EU database by its provider** is included.

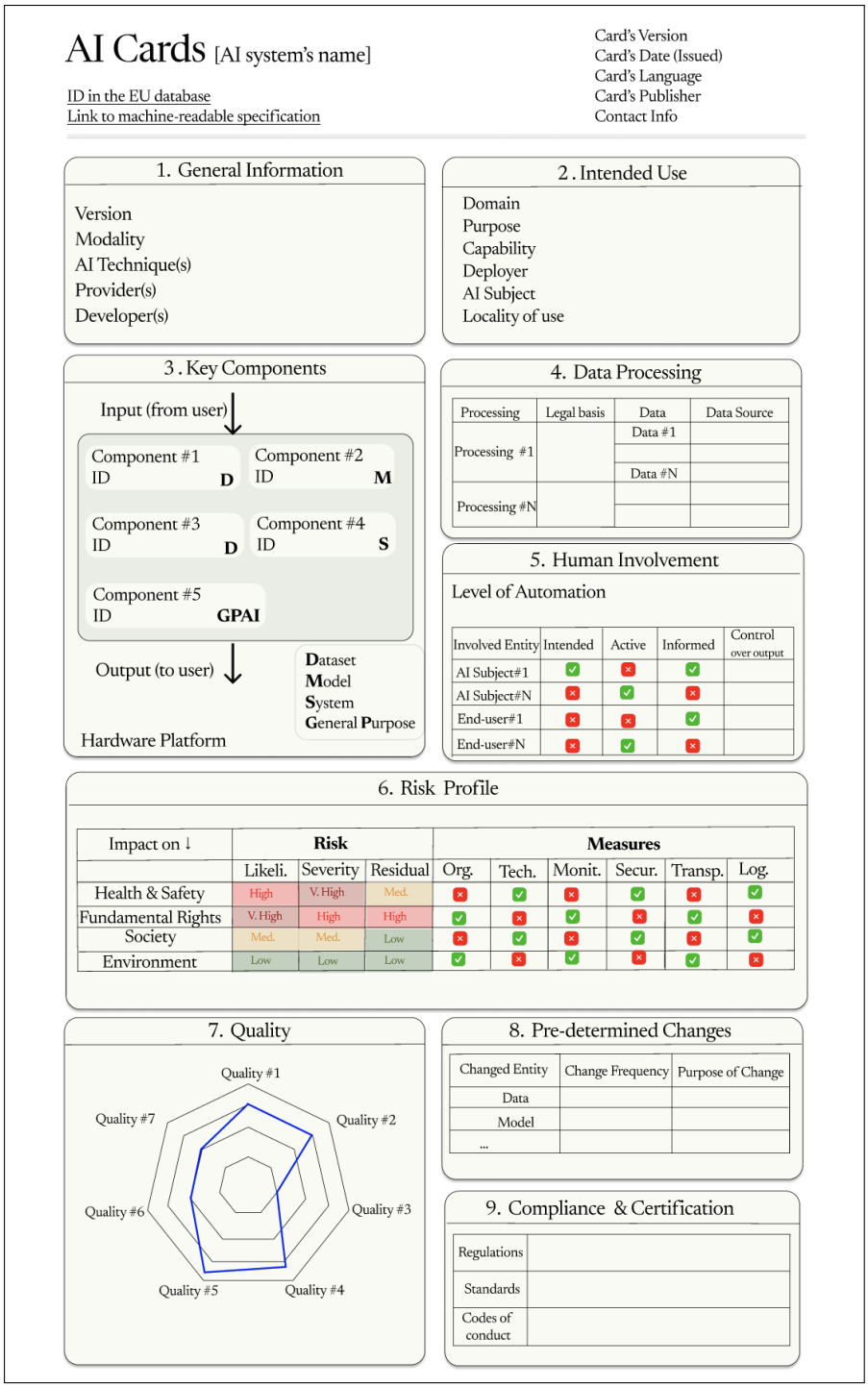


Figure 6.1: Human-readable representation of the AI Cards

Table 6.1: AI Cards framework requirements

Purpose	Providing an overview of technical documentation that effectively conveys key information about a given use of an AI system.
Knowledge sources	The EU AI Act (Article 9 on risk management system, Article 11 and Annex IV on technical documentation), ISO/IEC 42001 on AI management system, ISO/IEC 23894 on AI risk management.
Scope	*In the scope*: AI system (as a whole), its context of use, and information relevant to trustworthy AI concerns including risk management. *Out of the scope*: organisational processes, details of management systems, and documentation of AI components, e.g. datasets and models. The framework is horizontal and does not take sector-specific nuances into account.
Audience	AI providers, developers, deployers, end-users, subjects, auditors, and policymakers.
Representation formats	An easy-to-understand visual human-readable representation accompanied by a machine-readable specification.

(1) General Information

This section provides the key information about the AI system including its **version**, **modality**—which defines the form in which the system is placed on the market, e.g. standalone software, API, or safety component of a product—main **AI techniques** used, **provider(s)**, and **developer(s)**.

(2) Intended Use

As discussed earlier, the AI Act puts a considerable emphasis on intended purpose of the system, which refers to the *use* for which an AI system is intended by the provider. Based on the analysis provided in Section 3.3, 6 of the concepts are included in this section of the Card: **domain**, **purpose**, **AI capability**, **AI deployer**, **AI subject**, and **locality of use**. Note that information about data is represented in section (4) of the AI Cards.

(3) Key Components

Many AI systems are built through integration of multiple AI models, datasets, general-purpose AI systems, and other software elements, each of which has an effect on the system's behaviour and in turn its risks [152]. This section provides the system's *high-level architecture* in terms of incorporating **components**. For each component, its **name**, **version**, and **link to documentation or ID** are presented. For detailed information about components, the AI Cards relies on their documentation, which is presumably offered by the components' providers in the form of **information sheets**, e.g. Datasheets, Model Cards, and AI FactSheets. **Hardware platform** is also included in this part to specify the hardware on which the system is intended to run.

(4) Data Processing

AI systems that do not process data, if they exist at all, are rare. Within the EU digital acquis[1], the GDPR [21], which protects the fundamental right to privacy by regulating *personal data*, is applicable to those AI systems that process natural persons' data. Therefore, having knowledge of whether a given use of an AI system involves processing of personal data is crucial to correctly interpret the resulting legal compliance obligations. This section of AI Cards specifies the **data processing** associated with the system, the **legal basis** for processing **data**, e.g. consent, and the **source of this data**.

(5) Human Involvement

Involvement can take different forms depending on the phase of AI development, the role of human actors, and the system's **level of automation**—which has a range from fully autonomous to fully human-controlled according to ISO/IEC 22989 [36]. The level of automation also has a substantial effect on the safeguards, including human oversight measures, required for controlling AI risks. This section of the AI Cards provides an overview of involvement of two specific human actors: **AI end-users**, who use the AI system's output and **AI subjects**, who are subject to the outputs or effects of the system. For these actors, the following aspects of involvement are considered:

- **Intended involvement**: represents whether the involvement of a specific actor is as intended. An example of an *intended* AI subject in an

[1] Acquis refers to the body of rights and obligations that is binding across all EU Member States.

AI-based proctoring system is a student sitting for an online test. In this case, other persons present in the room are *unintended* AI subjects.

- **Active involvement**: shows whether a specific actor actively interacts with the AI system. For example, in the aforementioned online proctoring system, students actively interact with the system. In contrast, in a CCTV-based AI system that monitors students' behaviour during paper-based exams, students are not actively involved.

- **Informed involvement**: represents whether a specific actor was informed that an AI system is in place; for example, in cases where a decision affecting a person's education is made using AI-based solutions. This can serve as a partial indication of compliance with the AI Act's transparency obligations for certain AI systems, outlined in Article 50, which requires providers of the systems that directly interact with individuals to inform those individuals that they are engaging with an AI system.

- **Control over AI outputs**: shows the extent to which AI subjects and end-users have control over AI outputs, in particular decisions made by AI systems and their subsequent impacts. Inspired by the OECD's modes of optionality [94], the following six levels of control are considered:

  - An AI subject/end-user can opt in the system's output.
  - An AI subject/end-user can opt out of the system's output.
  - An AI subject/end-user can challenge the system's output.
  - An AI subject/end-user can correct the system's output.
  - An AI subject/end-user can reverse the system's output ex-post.
  - An AI subject/end-user cannot opt out of the system's output.

(6) Risk Profile

This part provides a high-level summary of risk management results, which includes an overview of **likelihood**, **severity** and **residual risk** associated with risks that have impact on **health and safety**, **fundamental rights**, **society**—including the often-absent systemic risks in many documentation approaches [152]—and **environment**. Further, this section shows whether any technical, monitoring (human oversight), cybersecurity, transparency, and logging measures applied to control the risks.

(7) Quality

With many AI incidents rooted in poor quality, ensuring that the system is of high quality by participating in AI regulatory sandboxes, benchmarking, and performing tests is necessary. This section of AI Cards aims to illustrate key **AI system qualities** using a radar chart. This can include **accuracy**, **robustness**, and **cybersecurity**, which are the key qualities explicitly mentioned in the AI Act (Article 15). Further, the relevant *product quality* and *quality in use*, described by ISO/IEC 25059 on AI SQuaRE (Systems and software Quality Requirements and Evaluation) [197] can be considered to be included in the quality section.

(8) Pre-determined Changes

This section of the AI Cards provides a list of **pre-determined changes** to the system and its performance in terms of **changed entity**, **frequency of change**, and **purpose of change**.

(9) Compliance & Certification

This section of the AI Cards lists the main digital **regulations** the AI system is compliant with, examples within the EU jurisdiction are the EU AI Act, GDPR, and Union harmonisation legislation. Also, it represents key **standards** to which the system or the provider(s) conform, e.g. ISO/IEC 42001 on AI management system [37], and **codes of conduct** followed in development or use of the AI system, for example UNESCO's recommendations on AI ethics [198].

6.3 Machine-Readable Representation of the AI Cards

Effective AI documentation practices should enable search capabilities, facilitate meta-analysis, allow for easy comparison of multiple AI systems, and support automation in creating, updating, and auditing of documentation [199, 14, 153]. To support these futures, the human-readable representation of the AI Cards requires an accompanying machine-readable specification. Therefore, the AI Cards framework supports machine-readable representation of information by leveraging the standards, methods, and tools, provided by the World Wide Web Consortium (W3C). This is also motivated by the body of work discussed in Subsection 2.4.3 and the rise in the uptake of open data formats for documentation, reporting, and sharing information by the

AI community—for example, HuggingFace's use of JSON Model Cards[2]—as well as the uptake of such data formats by the EU, e.g. DCAT-AP open data portals[3] [144], EU vocabularies and ontologies[4], and machine-readable regulatory reporting [200],.

The provision of machine-readable specifications for AI Cards is realised using AIRO and VAIR, as illustrated in Figure 6.2. With a knowledge graph-based specification of AI use cases, wherein the information elements are modelled, the human-readable specification of AI Cards can be automatically generated, as previously discussed in Section 5.3.

The machine-readable representation of AI Cards aims to foster openness and interoperability—both essential for exchanging information across the AI value chain. This representation is extensible and therefore enables accommodating sector-specific information requirements and allows adaptation as per the potential future amendments to the AI Act via delegated acts, the anticipated guidelines from authorities including the AI office, and case laws. Grounded on formal logic, it also assists in ensuring that the information is complete, correct, and verifiable.

[2] https://huggingface.co/docs/hub/model-cards
[3] https://op.europa.eu/en/web/eu-vocabularies/dcat-ap
[4] https://op.europa.eu/en/web/eu-vocabularies/controlled-vocabularies

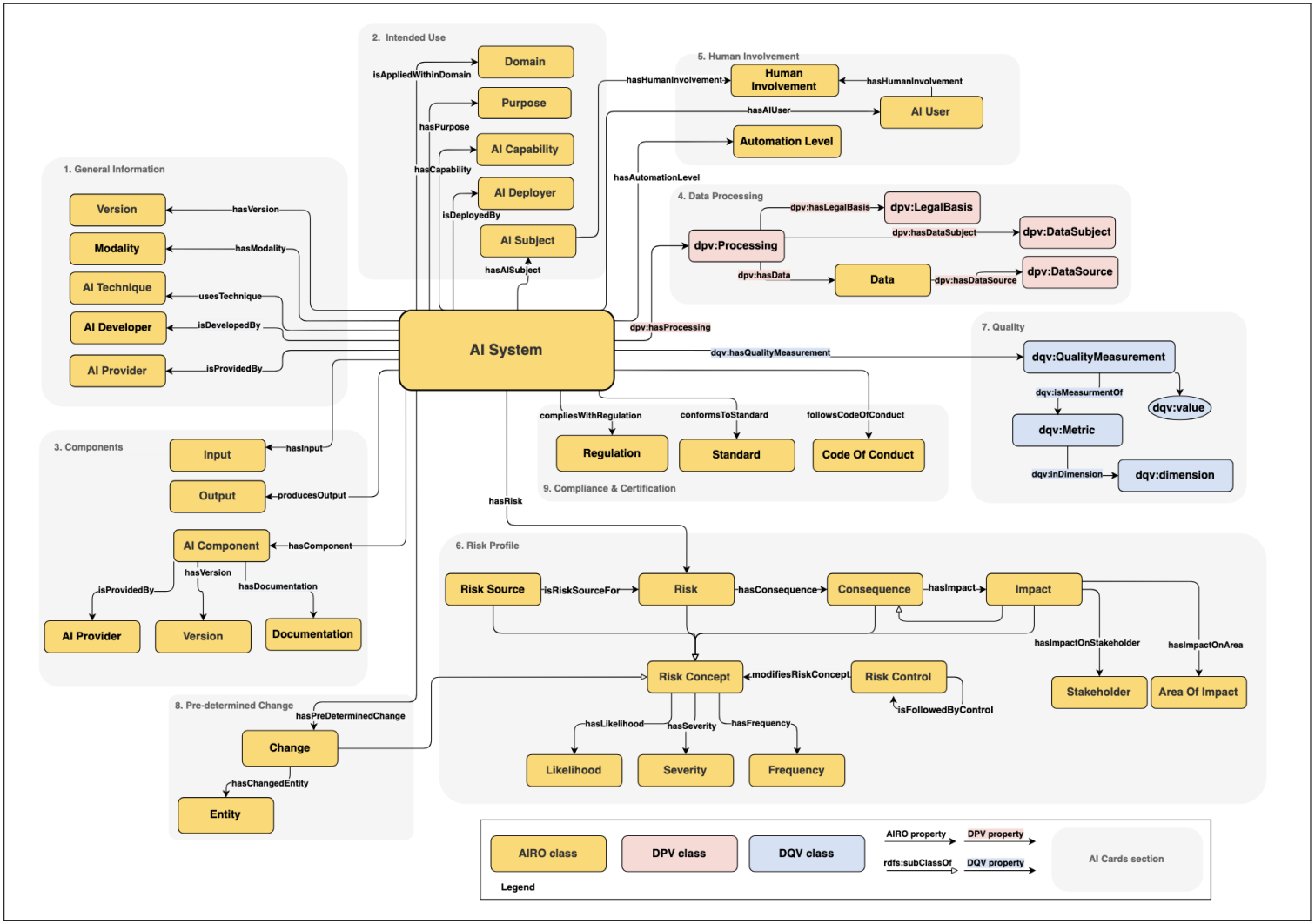


Figure 6.2: An overview of the AIRO-based semantic model for generating the AI Cards

6.4 Proof-of-Concept: AI Cards for an AI-Based Proctoring System

This section provides an illustrative proof-of-concept implementation of AI Cards for *Proctify*—an AI-based student proctoring system that aims to detect prohibited behaviour of students during online exams, introduced in Subsection 4.4.1.

In addition to the use case description in Subsection 4.4.1, it is important to note that throughout the risk management process performed in relation to Proctify, the risks and impacts of the system were identified and assessed by the provider, including the following:

- The system may have lower accuracy for students with darker skin tones.
- The system may have a higher rate of false-positive alarms for students wearing glasses.
- False-negatives and false-positives are more frequent for students with health issues or disabilities that affect their facial behaviour.
- There is a chance of over-reliance of human instructors on the system's output (automation bias).

These risks have the potential to negatively impact students' *mental health*, *future career*, and their rights to *dignity* and *non-discrimination*. Some of the measures applied to address the system's risks and impacts are:

- Ensuring the dataset is representative and diverse in demographic terms,
- Conducting rigorous and frequent testing of accuracy,
- Assigning expert human proctors,
- Creating clear protocols to act upon when an alarm is raised.

The Proctify's AI Cards is visualised in Figure 6.3. A snippet of the machine-readable specification is provided in Listing 8[5].

[5]For a comprehensive specification refer to https://github.com/DelaramGlp/airo/blob/main/usecase/proctify.ttl

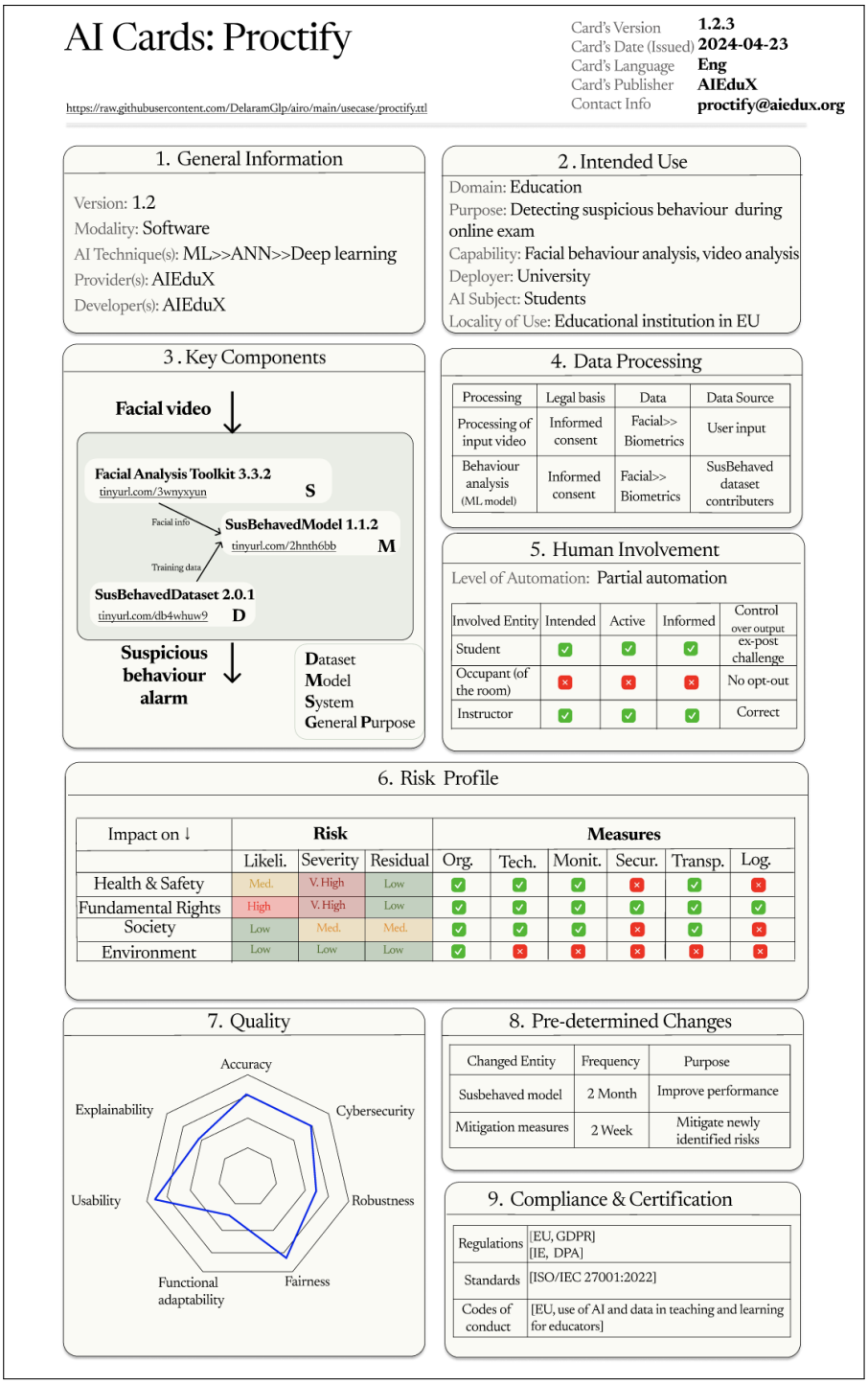


Figure 6.3: An example of AI Cards for an AI-based proctoring system.

```
1
2     ex:proctify
3     a airo:AISystem ;
4     airo:hasVersion ex:v_1.0.2 ;
5     dct:date "2023-09-11"^^xsd:date ;
6     airo:hasModality vair:Software ;
7     airo:usesTechnique vair:DeepLearning ;
8     airo:hasInput ex:facial_video ;
9     airo:producesOutput ex:suspicious_behaviour_alarm ;
10    airo:isProvidedBy ex:AIEduX ;
11    airo:isDevelopedBy ex:AIEduX ;
12    airo:hasComponent ex:facial_analysis_toolkit,
13                      ex:susbehaved_dataset ;
14    airo:hasModel     ex:susbehaved_model ;
15    airo:isAppliedWithinDomain vair:Education ;
16    airo:hasPurpose vair:DetectingProhibitedBehaviourDuringTest,
17                    ex:facial_behaviour_analysis,
18                    ex:video_analysis ;
19    airo:hasCapability ex:facial_recognition ;
20    airo:isDeployedBy ex:university ;
21    airo:hasAutomationLevel vair:PartialAutomation ;
22    airo:hasAISubject vair:Student,
23                      ex:other_occupant ;
24    airo:hasAIUser ex:instructor ;
25    airo:hasRisk ex:inaccuracy_risk_for_darker_skin ;
26    airo:hasRiskControl ex:bias_testing ;
27    dqv:hasQualityMeasurement ex:alarm_precision_measurement,
      ↪  ex:alarm_recall_measurement, ex:alarm_f_score_measurement   ;
28    airo:hasPreDeterminedChange ex:change_of_model ;
29    airo:compliesWithRegulation  ex:EU_GDPR ,
30                                 ex:Irish_Data_Protection_Act ;
31    airo:conformsToStandard vair:ISOIEC42001_2023 ,
32                            vair:ISOIEC27001_2022 ;
33    airo:followsCodeOfConduct
      ↪  ex:use_of_AI_and_data_in_teaching_and_learning   ;
34    dpv:hasProcessing ex:processing_1 .
```

Listing 8: A snippet of machine-readable provision of Proctify in Turtle

6.5 Discussion of the AI Cards' Benefits and Potential Applications

The AI Cards is primarily created by *AI providers*. However, with involvement of multiple actors in the AI development, creation of the AI Cards can transform into a collective activity, requiring exchange of information among these actors. In the AI Cards framework, the information sharing between AI providers and providers of its incorporating components, such as datasets and models, is particularly important. As mentioned earlier, the AI Cards framework relies on the components documentation provided by their providers in form of information sheets, for example, Datasheets and Model Cards. When this documentation is provided in an interoperable manner using common ontologies and vocabularies, it enables federated querying capabilities. Therefore, in terms of documentation management, the machine-readable specification of AI Cards not only assists AI providers in frequent modification and version control, but enables integration of data regarding incorporating components from diverse sources.

Concerning the AI Act, the AI Cards assists AI providers in addressing the obligations regarding risk management (Article 9), technical documentation (Article 11), and provision of information to deployers (Article 13) by offering a means for transparent sharing of AI and risk information with different parties, particularly *AI deployers*. The summarised view offered by the AI Cards enables *AI deployers* as well as *end-users* of the system in capturing a holistic view of the intended use of the system, its technical specifications, qualities, and risk profile. Further, AI deployers can use the AI Cards for the comparative evaluation conducted for selecting appropriate AI systems, in particular in procurement processes.

Regarding enforcement of the EU AI Act, the AI Cards can be helpful for *conformity assessment bodies* in tasks related to auditing and certification. The *AI Office* can benefit from AI Cards, in particular its Semantic Web-based representation, in development of the urgently-needed automated tools for fundamental rights impact assessment (FRIA), as mentioned in Article 27 (4), which has overlaps with technical documentation in terms of information requirements. The AI Cards can further be extended to support publishing and comparison of FRIAs, which in turn enable agile refinement and understanding of best practices in that regard.

The *European Commission* can benefit from the potential of the AI Cards in serving as a template for creating public searchable repositories of AI use cases. While the EU database (Article 71) aims at indexing Annex III high-risk AI systems, such an open repository that contains AI risk information

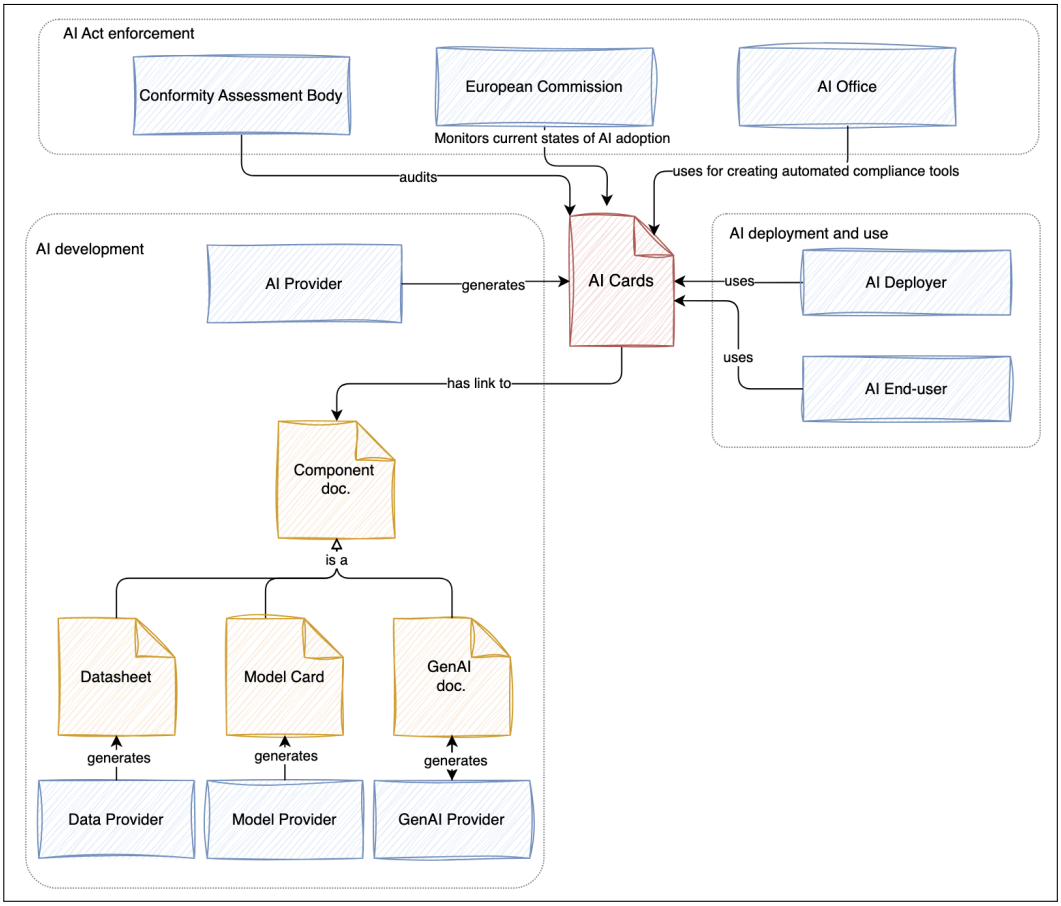


Figure 6.4: Applications of the AI Cards in the AI value chain

accelerates regulatory learning both among authorities and between prospective AI providers and deployers. Where confidentiality concerns may impede the open access sharing of such information, using summarised views offered by AI Cards can maximise sharing and learning. Sharing AI Cards also offers assistance in identifying and sharing best practices in AI risk management and impact assessment and assists in identification of reasonably foreseeable misuses of AI. This is helpful for AI providers and deployers, especially less well-resourced SMEs and public bodies in drawing legal certainty and further in implementation of the AI Act. A public knowledge base of AI use cases might be useful for the Commission in monitoring the current state of AI adoption within the EU and more importantly assessing the need for amending the list of Annex III high-risk AI systems as per Article 7. Adopting such an approach promotes the transparency in the process of adopting the relevant delegated or implementing acts. Figure 6.4 depicts a summary of

the potential applications of the AI Cards in the AI value chain.

The AI Cards will be evaluated in the next chapter in Section 7.4, as part of which the potential uses of the AI Cards as perceived by survey participants will be discussed (Subsection 7.4.3).

CHAPTER

SEVEN

EVALUATION

This chapter first discusses the methodology used for evaluating the research presented so far in Section 7.1. The validation of the AI Act analysis (Chapter 3) is presented in Section 7.2. The evaluation of AIRO and VAIR and evaluation of the AI Cards framework (Chapter 4) are detailed in Section 7.3 and Section 7.4, respectively. Finally, a summary of the proof-of-concept implementations of the artefacts proposed in the thesis is provided in Section 7.5.

7.1 Evaluation Methodology

The methodology adopted for evaluation of the research involves 4 main components:

1. **Validity of the AI Act analysis** (RO1) is determined through consultation with AI and law experts, and is presented in Section 7.2.

2. **The ontologies, i.e. AIRO and VAIR,** (RO2) are evaluated in regard to their (i) sufficiency in satisfying information requirements for determining high-risk AI systems, expressing intended purpose, documenting AI and risk information, and registering AI systems and (ii) technical quality including logical consistency, adherence to best practices, freedom from errors, and FAIRness. Evaluation of AIRO and VAIR is provided in Section 7.3.

3. The usefulness and usability of the **AI Cards framework** (RO4) are evaluated through an anonymous survey. Evaluation of the AI Cards framework is presented in Section 7.4.

4. **Applicability of the ontologies and the Semantic Web-based compliance mechanisms** (RO2 and RO3) was previously demonstrated by the proof-of-concept implementations in Chapter 4 and Chapter 5. In addition, applicability of the AI Cards (RO4) was shown in Chapter 6. A summary of these proof-of-concept implementations is presented in Section 7.5.

7.2 Validity of the AI Act Analysis

Since this thesis stands as an interdisciplinary research at the intersection of AI, law, and standardisation, engagement with subject matter experts was necessary to ensure the validity of the interpretation of the AI Act and related AI standards.

As an initial step of validation, the analysis was continuously refined based on the discussions with the supervisors of this thesis, who have conducted extensive work on capturing requirements and semantic modelling for compliance with the EU GDPR, and have been involved in international and European standardisation activities regarding AI, cybersecurity, and privacy.

As part of the PROTECT project, the inputs from early-stage researchers from the Ontology Engineering Group (OEG) at Universidad Politécnica de Madrid (UPM), the School of Law at Trinity College Dublin (TCD), the Department of Philosophy at University of Twente, and Rathenau Instituut were used in validation of the risk management requirements. These requirements were extracted from the Commission's proposed AI Act and other relevant resources including the EU High-Level Expert Group's (HLEG) ethics guidelines for trustworthy AI [201], its assessment list (ALTAI) [69], and the Medical Devices Regulation (MDR) [166]. This interdisciplinary collaboration within the PROTECT network further contributed to shaping the notion of AI risk in this thesis.

Consultation with AI Act experts from the European Commission's Joint Research Centre (JRC), which provides science and research services to support EU policymaking, played a critical role in validating the analysis. This consultation specifically addressed the AI Act's documentation requirements and therefore was directly related to the following research objectives: RO1(c), which involved analysis of the provisions of the AI Act regarding documentation, RO3(c) that was approached using SPARQL for retrieving information to generate documentation, and RO4, which was addressed by development of the AI Cards framework (refer to Subsection 7.4.1 for more details). Given the involvement of the JRC in assessing European standards at the time of collaboration, the alignment of existing standards with the requirements of

the AI Act's risk management and documentation obligations was discussed, mainly during a 6-week on-site visit to JRC, resulting in refinements in using standards in this work.

With regard to standards, the development of AI standards at both international and European levels were monitored through membership in the National Standards Authority of Ireland (NSAI), ISO/IEC JTC 1/SC 42 and CEN-CLC/JTC 21. This involvement was particularly useful in tracking development of the relevant standards to this work, including ISO/IEC 22989 [36] on AI terms and concepts, published in July 2022, ISO/IEC 23894 [39] on AI risk management, published in February 2023, and ISO/IEC 42001 [37] on AI management systems, published in December 2023. In addition, participation in the recent discussions at CEN-CLC/JTC 21 concerning their response to the European Commission's AI standardisation request [13], further helped in refining the links made between standards and the AI Act's requirements in this thesis.

Throughout this research, the analysis was cross-referenced with documents and statements issued by the EU officials and compared with experts' opinions published in academic papers, previously reviewed in Section 2.2, notably [8, 34, 85, 84, 64].

The use of the 5 concepts for describing the system's intended use has been also validated by Bogucka et al. [155] in both co-design and user study phases of the development of the AI Impact Assessment Report template. Within this template, the 5 concepts are utilised for describing the AI system's use, as depicted in Figure 7.1.

7.3 Evaluation of AIRO and VAIR

This section discusses evaluation of AIRO and VAIR from two aspects:

1. **Ontology verification** determines, on the basis of the ontology requirements, whether the ontologies are built correctly [202]. Therefore, AIRO and VAIR were compared against the information requirements that were identified from the AI Act and the relevant standards in Chapter 3.

2. **Ontology quality assessment** ensures the ontologies are of high-quality, promoting their interoperability and reusability. The criteria considered for AIRO and VAIR quality assessment include consistency, freedom from errors, and FAIRness.

Usefulness of machine-readable specification of the AI Cards, which is created using AIRO and VAIR, will be discussed as part of the user study in

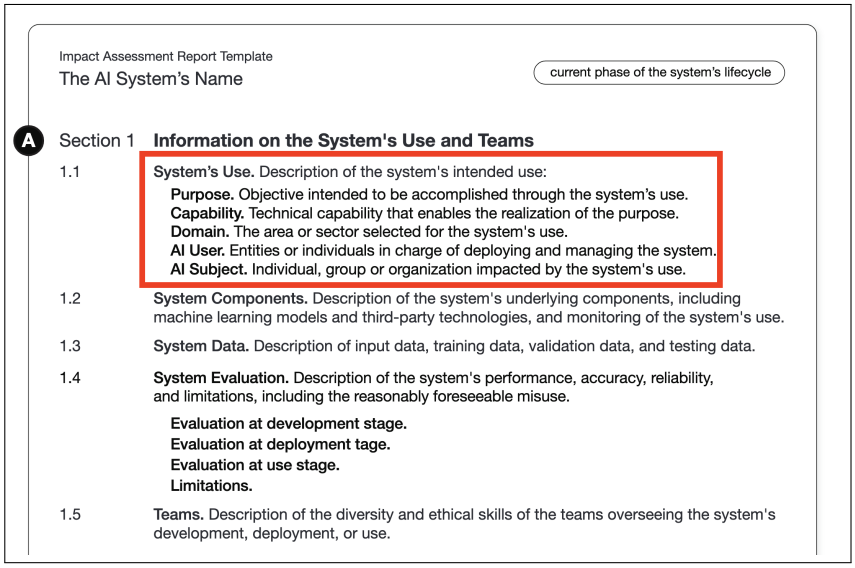


Figure 7.1: AI system's use section in the AI Impact Assessment Report template [155] that uses the 5 concepts proposed by this work

Subsection 7.4.3. The results of the user study were considered as an indirect assessment of the usefulness of the ontologies.

7.3.1 Ontology Verification

The following provides evidence of fulfilment of the information requirements by the ontologies. This demonstrates the extent AIRO and VAIR are in-line with the use they are designed for.

Fulfillment of Information Requirements for Determining High-Risk and Prohibited AI systems

The questions for determining Annex III high-risk AI, presented in Subsection 3.2.1, shaped the foundation of AIRO. The top-level concepts and relations modelled in AIRO to address the information required to answer these questions are presented in Table 7.1. The additional concepts and relations modelled to satisfy the questions for determining prohibited AI systems (Subsection 3.2.3) are shown in Table 7.2.

Moreover, VAIR was assessed to ensure inclusion of the instances of these concepts, that are essential for modelling the rules for determination of Annex III high-risk AI systems (SHACL shapes presented in Section 5.1). The assessment showed completeness of VAIR in covering the instances of the 5

Table 7.1: Fulfilment of competency questions for determining Annex III high-risk AI

Competency question	**Concept**	**Relation with** `airo:AISystem`
1. In which domain is the AI system used?	`airo:Domain`	`airo:isAppliedWithin Domain`
2. What is the purpose of using the AI system?	`airo:Purpose`	`airo:hasPurpose`
3. What is the capability employed by the AI system?	`airo:AICapability`	`airo:hasCapability`
4. Who is the deployer of the AI system?	`airo:AIDeployer`	`airo:isDeployedBy`
5. Who is the AI subject?	`airo:AISubject`	`airo:hasAISubject`

Table 7.2: Fulfilment of additional competency questions for determining prohibited AI practices

Competency question	**Concepts & relations**
What data is processed by the AI system?	`airo:AISystem dpv:hasProcessing dpv:Processing .` `dpv:Processing dpv:hasData airo:Data.`
What is the locality of use?	`airo:AISystem airo:isUsedWithinLocality airo:LocalityOfUse`
What is the consequence of using the system?	`airo:AISystem airo:hasRisk airo:Risk .` `airo:Risk airo:hasConsequence airo:Consequence .`
What is the impact of using the AI system?	`airo:Consequence airo:hasImpact airo:Impact .`

concepts featured in the rules for determining Annex III high-risk AI, presented in Appendix C. It should be emphasised that completeness of VAIR for purposes other than describing Annex III use cases was not assessed.

Describing AI Intended Purpose

Since the information requirements for expressing intended purpose overlap with the ones concerning determination of high-risk and prohibited AI, the evaluation presented earlier in this section is applicable. As shown in Section 5.2, the AIUP profile introduced new concepts (listed in Table 5.2) in order to satisfy competency questions documented in Table 5.1.

Inclusion of the AI Cards' Information Elements as a Subset of Risk Management and Technical Documentation

AIRO and VAIR were also evaluated to determine whether they cover the AI Cards' incorporating information elements, which are considered as a subset of the information requirements arising from the AI Act's risk management and technical documentation obligations. This evaluation ensures presence of the concepts and relations needed for generation of the AI Cards using the queries provided in Section 5.3.

Registration Requirements

Fulfilment of registration requirements was previously illustrated in Table 5.4. The AICat competency questions were addressed using AIRO, VAIR, and AIUP, as well as reusing some existing vocabularies and ontologies, including Dublin Core Terms, DCAT, ODRL, and DPV.

7.3.2 Quality Assessment

Quality assessment of AIRO and VAIR was carried out in accordance with the following criteria:

1. **Syntax validity and logical consistency**: Throughout the development process, the syntax was checked using Protégé. Logical consistency was tested using build-in Protégé reasoners including FaCT++ 1.6.5 and HerMiT 1.4.3.

2. **Adherence to ontology development best practices**: In addition to following LOT as an ontology development methodology, quality of AIRO and VAIR was ensured by following best practices and guidelines from the Semantic Web community, including the W3C's Data on the Web Best Practices [59], the W3C Recipes for Publishing RDF Vocabularies [58], FAIR best practices for vocabularies and ontologies [62], and WIDOCO best practices for ontology documentation [175].

3. **Freedom from common ontology pitfalls**: OOPS! (OntOlogy Pitfall Scanner!) [60], was used to detect and further resolve common errors and pitfalls of AIRO and VAIR[1]. The key unresolved pitfalls raised by OOPS! concerned the following:

[1]Re-assessment of the final version of both ontologies using OOPS! was not possible due to the unavailability of the OOPS! service in September 2024.

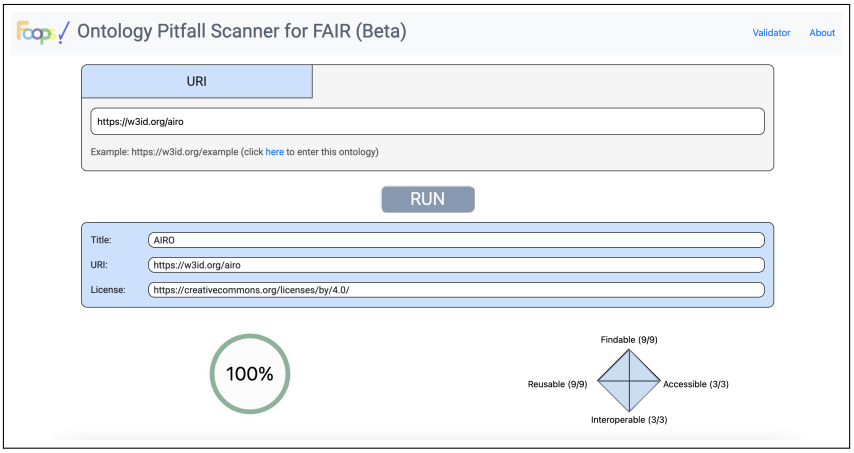


Figure 7.2: A screenshot illustrating the result of evaluating AIRO using FOOPS!

- Missing annotations: It was raised due to the absence of definitions for those classes and properties reused from other ontological resources.
- Missing disjointness: This issue was raised as disjointness between some classes are not defined, given that they can share individuals.
- Missing domain or range in properties: This pitfall addressed the deliberate design decision, discussed in Subsection 4.2.2, to enhance AIRO's flexibility in modelling AI use cases by not restricting domain or range for some properties.
- Inverse relationships not explicitly declared: This issue addressed the lack of inverse relations (`owl:inverseOf`) between object properties. This was intentional to keep AIRO minimal and simple, as also discussed in Subsection 4.2.2.
- Equivalent classes not explicitly declared: This was a false positive, incorrectly identifying equivalent classes.

4. **FAIRness**: Adherence of the ontologies to FAIRness principles was evaluated using FOOPS! [174]. The fairness scores for both AIRO and VAIR, as calculated by the FOOPS! web service[2], showed their full compliance with FAIRness principles for ontologies. Figure 7.2 depicts the result of evaluating AIRO using the FOOPS! web service.

[2] https://foops.linkeddata.es/FAIR_validator.html

7.3.3 Comparison with the State of the Art

With a focus on the coverage of concepts, Table 7.3 compares the 6 key generic AI taxonomies, reviewed in Subsection 2.3.1, with AIRO and VAIR. In this table, and the following ones in this chapter, a black circle (•) indicates presence and a blank circle (○) signifies absence. Further, an asterisk (*) indicates involvement of the author in the development of the ontology.

Table 7.4 provides a comparison of AIRO with the existing ontologies related to AI risks, presented in Subsection 2.3.3. In this table, the ontologies that have been developed based on AIRO with involvement of the author, which are the Health AI Risk Taxonomy (HART) [67] and DPV 2.0 [57], are also included to ensure the comprehensiveness of the evaluation (see Subsection 8.3.2 for the details of these two ontologies).

These comparisons show that AIRO and VAIR include concepts for expressing information related to technical aspects of AI systems, their context of use, and risks in standardised formats that not only enable machine-readable representation of AI use cases, but allow further reasoning. As shown in Table 7.3, AIRO and VAIR do not include explicit categorisation of AI impacts (harms). This is due to the way *impact* is specified in AIRO, using *impacted entities* and *area of impact*. AIRO is a generic ontology and therefore can be applied, and extended if needed, for modelling AI use cases, regardless of the sector they are used within, their technology, and the types of risks they impose. At the time of writing, AIRO and VAIR are the only ontologies aligned with the requirements of the EU AI Act. Both AIRO and VAIR have been published under open and permissive licenses, allowing their free reuse.

Leveraging of Semantic Web technologies and alignment with the AI Act and ISO/IEC standards are two distinctive features of AIRO and VAIR, filling the gap associated with the lack of data models for AI risks. Another key feature of AIRO and VAIR is their practical application, which has been demonstrated in this thesis through implementation of a set of RegTech solutions.

Table 7.3: Comparison of AIRO and VAIR with SOTA generic AI taxonomies (a black circle (•) indicates that the criterion has been satisfied and a blank circle (○) indicates that the criterion has not been fulfilled)

	AIRO & VAIR	OECD [94]	AIAAIC* [68]	CSET [97]	GMF [98]	AITopics [99]	AVID [101]
Machine-readable format	OWL2	○	○	○	○	○	JSON
AI technical concepts	•	•	•	•	•	•	○
AI lifecycle concepts	•	•	○	○	○	○	•
AI use concepts	•	•	•	•	•	•	○
Types of harms	○		•	•			•
Risk management concepts	•	•	○	•	•	○	○
Existing tools empowered by the taxonomy	•	○	○	○	○	•	•
Alignment with the AI Act	•	○	○	○	○	○	○
Alignment with standards	•	○	○	○	○	○	○

Table 7.4: Comparison of AIRO with ontologies related to AI Risks (* illustrates the involvement of the author)

Ontology	Scope	AI technology	AI use	AI risk	Open-source code	Documentation	Peer-review	proof-of-concept
AIRO	AI systems, their use, and associated risks	●	●	Risks as per the AI Act & ISO/IEC standards	●	●	●	●
HART* [67]	AI risks in the health domain	●	●	Types of AI risks in the health domain	●	●	●	○
DPV 2.0* [57]	Processing of personal data and use of technologies	●	●	Risk management based on ISO 31000 series	●	●	●	●
EA-Ontology [110]	Technology ethics assessment	○	●	Ethical risks	●	●	●	●
ASCENT [111]	AI explainability	○	●	Measures related to explainable AI	●	○	○	●
ExplainableML Ontology [112]	AI explainability	●	○	Measures related to explainable AI	●	○	●	●
FIDES [113]	Accountability of statistical ML models	●	○	○	●	●	●	●
Doc-BiasO Ontology [114]	AI Bias documentation	●	●	AI bias & its measures	○	○	●	○
AIO [115]	AI terminology	●	●	Types of bias	●	●	○	○
TAIR* [116]	Alignment of trustworthy AI requirements	○	○	○	●	●	●	●
AIPO [117]	Alignment of trustworthy AI principles	○	○	○	●	○	○	●

7.4 Evaluation of AI Cards

7.4.1 Expert Consultation

The AI Cards was created in close collaboration with 3 experts based in the JRC's digital transformation and data directorate (JRC T). In addition, throughout the development process, the analysis of the AI Act's provisions and the AI Cards framework was discussed with 9 JRC experts, whose backgrounds were in digital policymaking in the EU, particularly the AI Act, AI transparency and documentation, explainability, cybersecurity, standardisation, use of AI in the public sector, and Semantic Web. This consultation aimed to ensure alignment of the AI Cards framework with EU digital policies, adoption of correct terminology, and suitability of the framework for addressing common concerns of AI stakeholders. Based on the feedback received, the visual representation of the AI Cards was refined 4 times, with minor subsequent adjustments to the semantic models, in particular AIRO.

Additionally, once a solid structure for the AI Cards was reached, a subsequent consultation was conducted with JRC experts, who were not directly involved in the development process. In this phase of consultation, a draft of a paper on AI Cards[3] was shared for internal feedback. Constructive feedback was received on the use of ISO/IEC standards, which resulted in minor modification of the analysis and the AI Cards framework.

It is important to note that the views expressed by the JRC researchers were their individual expert opinion and therefore should not be regarded as an official position of the European Commission.

7.4.2 Comparison of AI Cards with the State of the Art

Given its dual representations, the AI Cards framework is compared with the two sets of related work reviewed in Chapter 2. First, it is compared with AI use and risk documentation practices reviewed in Subsection 2.5.2 and Subsection 2.5.3 using the criteria discussed in Subsection 2.5.4. The comparison, which is provided in Table 7.5, shows that in terms of information coverage, the *template for AI impact assessment* [155] and *AI Risk Profiles* [157] are closely aligned with the AI Cards. Compared to the visual representation of AI Cards, the *template for AI impact assessment* [155] features detailed risk management information (risk profile section), which comes at the cost of a summarised view. *AI Risk Profiles* [157] also heavily

[3]The published version of this paper is [72].

Table 7.5: Comparison of AI Cards with existing AI use case and risk documentation approaches

Approach	Tech.	Context of use	Risk	AI Act	Summarised view	Machine-readable
AI Cards	●	●	●	●	●	●
Use Case Cards [64]	○	●	○	●	●	○
AI Usage Cards [137]	○	●	●	○	○	●
AI Impact Assessment Report Template [155]	●	●	●	●	○	○
RISKCARDS [156]	○	○	●	●	○	○
AI Risk Profiles [157]	●	●	●	●	●	○

relies on lengthy unstructured textual information with the summarised view limited to risks, mitigation measures, and certifications.

The second aspect of comparison focuses on setting the AI Cards framework side by side with machine-readable documentation approaches, that were previously investigated in Subsection 2.4.3. As shown in Table 7.6, the scope of existing machine-readable documentation, reviewed previously in Subsection 2.4.3, is limited to semantic specifications for dataset and model documentation[4]. In addition, none of these studies consider alignment with the requirements of the AI Act. The semantic specification for the AI Cards, which is created using AIRO and VAIR, models AI use cases and their risks. This, to some extent, fills the gap in the state of the art regarding open knowledge models that enable expressing AI use cases, aligned with the requirements of the EU AI Act.

The AI Cards framework, consisting of both human- and machine-readable representations, distinguishes itself from existing documentation approaches through:

- Its modular and future-proof design, rooted in its semantic nature, that ensures the scalability required to address documentation requirements arising from forthcoming AI regulations, policies, and standards. This

[4]To avoid confusion, the use of DPV for GDPR-required documentation previously reviewed in Subsection 2.4.3 is not included.

Table 7.6: Comparison of AI Cards (machine-readable representation) with existing approaches for machine-readable documentation of AI and its components

Approach	Scope	License	AI Act	Peer-review
AI Cards	AI use case and risk	CC-BY-4.0	•	•
Open DataSheets [133]	Dataset	MIT License	○	○
MCRO [135]	Model	Attribution 3.0 Unported	○	•
LMDC [136]	Dataset & model	Apache License 2.0	○	•

scalability is achieved by extending the AIRO and VAIR at the conceptual level through inheritance or concept specialisations or by reusing them in semantically richer ontologies within separate namespaces.

- Its alignment with the EU AI Act's provisions in regard to technical and risk management documentation, as well as use of ISO/IEC standards related to AI risk management and management systems.

- Its dual approach towards information representation, which makes the framework accessible to both humans and machines. This facilitates communication, ensures consistency, and promotes interoperability needed in documentation generation and maintenance activities performed by several entities across the AI value chain. In addition, the Semantic Web-based representation enables automation, which is much-needed in AI documentation approaches to ensure the documentation is in sync with the AI system development and usage practices [154]. As discussed before, automation is also crucial for investigation of technical documentation by AI auditors and conformity assessment bodies for compliance checking and certification.

- Its holistic *AI use*-focused approach, which allows inclusion of information regarding the context of use, risk management, and compliance in addition to technical specifications. While most existing documentation practices address potential risks or ethical issues at a very high level in an unstructured format [93], the AI Cards framework draws a picture of the risk management system that is already established by

illustrating an overview of the identified risks and applied mitigation measures.

7.4.3 Evaluation of the AI Cards Framework through a User Study

In continuation of the collaboration with JRC and to evaluate the AI Cards, a user study was conducted to assess (i) the alignment of the AI Cards framework with the overall objectives of the AI Act and (ii) the potential of the AI Cards in implementation and enforcement of the AI Act.

The questionnaire, which is presented in Section E.2, incorporates four groups of questions that inquire about (1) the participant's background, (2) (only) the visual representation of the AI Cards, (3) (only) machine-readable representation of the AI Cards, and (4) the AI Cards framework (both human- and machine-readable representations). In the last of these, to assess how stakeholders envisage usability of the AI Cards, the System Usability Scale (SUS) [203] was used. As the SUS is originally designed for assessing perceived usability of interactive human-computer interfaces, minor wording changes were applied to fit it with the purpose of AI Cards' assessment and to provide more clarity. Table 7.7 shows the list of SUS-based questions for assessing the AI Cards' usability, each answered on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Preliminaries

Ethics approval: Prior to carrying out the survey, it was reviewed and approved by the ethics committee within the School of Computer Science and Statistics at Trinity College Dublin.

Informed consent: At the beginning of the survey, the participants were asked to agree to the informed consent form (it was a mandatory field). Then, they could start completing the survey. The survey was anonymous and participation in the survey was optional.

Introducing AI Cards: Essential introduction about AI Cards was provided within the survey, with links to an example of AI Cards for Proctify (presented in Section 6.4)[5]. Further, a one-pager infographic was shared with participants, attached to the invitation email. Where possible or requested by participants, information about the AI Cards was presented in a preliminary interactive session. In total, four sessions were held. However,

[5]The information that was shared regarding the example is also available at: https://delaramglp.github.io/aicards/example/

Table 7.7: SUS questions used for evaluating usability of AI Cards

No.	Question
SUS01	I think that AI stakeholders would like to use the AI Cards framework frequently.
SUS02	I find the AI Cards framework unnecessarily complex.
SUS03	I think the AI Cards framework is easy to use.
SUS04	I think that AI stakeholders would need the support of a technical person to be able to use the AI Cards framework.
SUS05	I find the various aspects (i.e. information elements and human- and machine-readable formats) in the AI Cards framework are well integrated.
SUS06	I think there is too much inconsistency in the AI Cards framework.
SUS07	I would imagine that most people would learn to use the AI Cards framework very quickly.
SUS08	I find the AI Cards framework very cumbersome to use.
SUS09	I feel very confident using the AI Cards framework.
SUS10	I need to learn a lot of things before I could get going with the AI Cards framework.

due to anonymity of the survey, investigating the effect of running interactive sessions on the results was not possible.

Recruitment

The following three groups of participants were asked to take part in the survey in a sequential manner:

- **Student cohort**: 23 students were recruited from the Data Governance module in Dublin City University, attended by students enrolled in Masters of Art (MA) in Data Protection and Privacy (MDPP) and the European Master in Law, Data and Artificial Intelligence (EMILDAI). It is important to highlight that, due to the nature of this module, some of the participants (30 percent) have notable positions in public organisations, industry, and NGOs.
- **Academic cohort**: Following oral presentations at the ADAPT Centre's annual scientific conference (AASC'2024) and ADAPT Law & Tech working group, an invitation email was sent to the centre's members. 18 participants contributed to validation of AI Cards by completing the survey.

- **Industry, policy, and standard actors cohort**: In the final round, a wider call was issued aiming to capture industry, policy, and standard players' opinions. The call, which targeted those who are familiar with the AI Act, was disseminated through email as well as other communication channels, including LinkedIn[6], X[7], Slack groups, and the NSAI portal. 9 participants contributed to validation of AI Cards by completing the survey.

The responses received from each cohort were analysed separately, considering the different time intervals during which the survey was conducted for each cohort. Minor adjustments were made to the AI Cards design after the first round of the survey with students as well as consultations with JRC experts. However, the overall design of the AI Cards and survey questions shared with the cohorts remained unchanged. The integrated findings and the impact of the survey results on the AI Cards are discussed in the remainder of this section and presented in details in Appendix E.

Findings

In total, 50 responses were collected. The following refers to the comments regarding the missing and additional information elements, which directly impacted the design of the AI Cards. Moreover, a discussion of the potential of the AI Cards, as perceived by the participants, will be provided. The detailed analysis of the responses is presented in Appendix E.

Human-readable representation: Most participants (86%) indicated that the human-readable representation of the AI Cards is useful (useful or very useful) in providing key AI and risk information. Some participants suggested including additional information. Table 7.8 illustrates a summary of the suggestions that impacted the design of AI Cards. Some comments mentioned inclusion of detailed information within the human-readable representation. As inclusion of such detailed information was not aligned with the objective of providing a summarised view, these comments did not impact the visual representation. However, they were considered as new information requirements for the ontologies (for these comments please refer to Table E.1, Table E.2, and Table E.3).

In the context of implementation and enforcement of the AI Act, the results showed that the AI Cards is perceived by the majority of participants

[6]See the call here: https://www.linkedin.com/posts/delaramglp_call-for-participants-i-would-like-activity-7217456286454456321-JcrU?utm_source=share&utm_medium=member_desktop.

[7]See the call here: https://x.com/DelaramGlp/status/1811691465715695887.

Table 7.8: Changes applied to the AI Cards based on the participants' comments

Participant's comment	**Impact on the AI Cards**
General comments	
Descriptive text of key words and abbreviations used (mentioned by 3 participants)	Clarified the fields as much as possible.
Risk Profile	
"be more specific about what kind of rights are meant. Individual rights? Fundamental rights? "	Rights was reworded to fundamental rights.
Data Processing	
"a link to the GDPR ... adding a short version of the requirements of data processing..."	The *structure of the data processing section* was changed by focusing on *processing*. *Legal basis* was added.
".. a place to list input datasets that are not personal or sensitive..."	The *structure of the data processing section* was changed by removing types of data (i.e. sensitive and non-personal) to avoid confusion.
"How the data was sourced"	*Data Source* was added to the data processing section.
Components	
"Where the AI is run"	*Hardware platform* was added to the Components section.
Other comments	
"Information related to the registry process"	*ID in the EU database* was added.

as adaptable for the following tasks related to the AI Act: compliance checking, creating technical documentation, risk management documentation, and crafting guidelines (for details see Figure E.7). The potential uses of the human-readable representation of AI Cards mentioned by the participants highlights three themes:

- Sharing and communicating AI and risk information with different parties with different levels of AI literacy and technical knowledge (this was the most common use mentioned by the participants),
- Auditing and monitoring of AI systems,
- Drawing comparisons between multiple AI systems.

Machine-readable specification: The majority of the participants believed that the machine-readable representation is useful or to some extent useful for developing automated tools, establishing a common language, and structuring the EU registry of AI systems. Among these, establishing a common language around AI and risks rated useful by a higher proportion of participants, compared to the other tasks (see the details in Figure E.9). Further, the participants indicated that machine-readable specifications could assist in compliance tasks, including determining risk category, reporting, building AI governance tools, and achieving interoperability.

The overall framework: According to the participants, the three key objectives of the AI cards are *transparency* (80%), *accountability* (62%), and *comprehensibility* (62%). The following objectives were also mentioned by the participants: human oversight, terminological alignment, and fairness.

In terms of potential target users, participants found the AI Cards particularly beneficial for AI providers and AI deployers (see the results in Figure E.11). Civil society, researchers, and the general public were among the stakeholders identified by participants as potential beneficiaries of the AI Cards.

SUS scores: The results of SUS are shown in Figure 7.3 using a box plot, where due to the mixed tone of the items, a higher score for odd-numbered items and a lower score for even-numbered items are desired. The overall score in the range of 0 to 100 is calculated using the equation for standard SUS score proposed in [204]:

$$SUS = 2.5(20 + SUM(SUS01, SUS03, SUS05, SUS07, SUS09) - SUM(SUS02, SUS04, SUS06, SUS08, SUS10)) \quad (7.1)$$

There were 8 instances of missing values where a participant did not provide an answer to a SUS question. These missing values, that were distributed across 5 questions, were imputed by the median score calculated for the corresponding question. After missing data handling, the scores had a mean of 62.95 and median of 61.25, which is interpreted as a fairly good score, according to [204] and given that the AI Cards has been provided as a static visual framework without an interactive user interface. The predominantly favorable responses to frequency of use, ease of use, and ease of learning (SUS 01, 03, and 07) are indicators of the high perceived usability of AI Cards, which can be further enhanced by providing an interactive user interface to overcome technical barriers noticed by participants (SUS 04 and 10).

Another key aspect in evaluation of effectiveness and usefulness of an AI documentation framework is the extent of its adoption by the AI community,

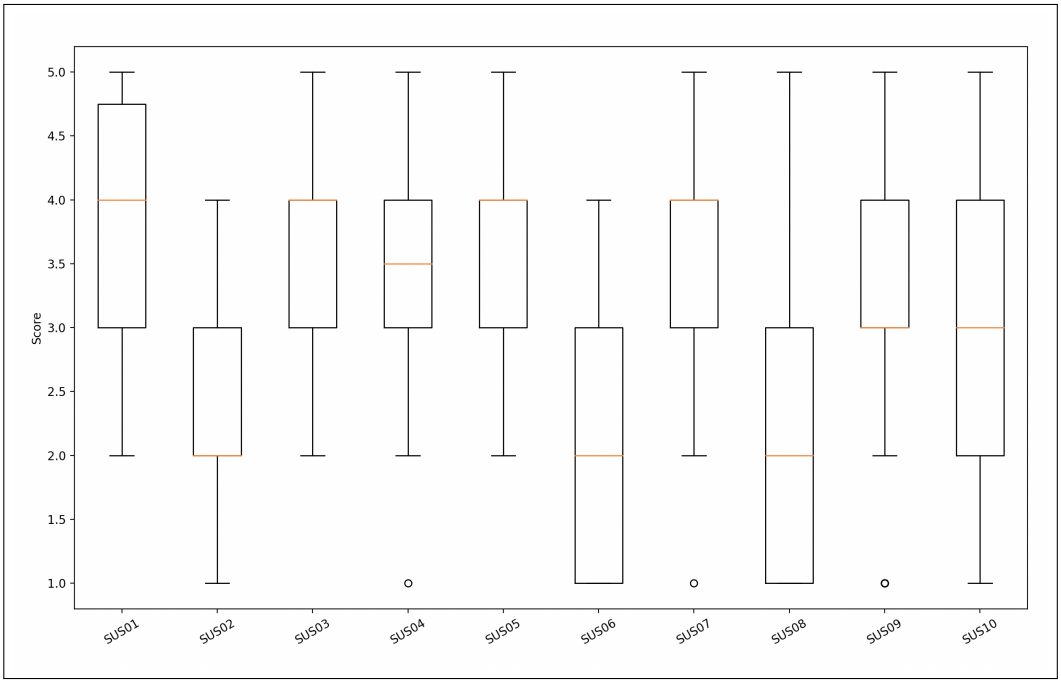


Figure 7.3: Box plot representing summary of SUS evaluating the AI Cards'. The scores are in a Likert 5 points scale where for odd-numbered items a higher score and for even-numbered a lower score are desired.

as evidenced in widely-adopted documentation approaches, e.g. Datasheets [134], Model Cards [147], and Factsheets [149]. Given that the AI Act was only published recently, the extent of AI Cards' adoption will need to be assessed over time. As indicated by one of the participants, *"The real world application would be interesting to see and may require changing based on different Member State authorities requiring other information be made available for documentation purposes"*.

Assessing the ease with which human users can fill out the AI Cards is relevant and further illustrates the usability and usefulness, however, this aspect of assessment is not included in the evaluation. This is due to the fact that the primary focus for generating AI Cards has been on automation of information retrieval using SPARQL queries, as illustrated in Section 5.3.

7.5 Evaluation of the Applicability of the Thesis Artefacts

As mentioned earlier, the applicability of AIRO and VAIR, the Semantic Web-based compliance mechanisms, and the AI Cards was demonstrated through proof-of-concept implementations in the previous chapters. Table 7.9 provides an overview of these proof-of-concept implementations. Given that the evaluation ultimately aimed to discover the extent to which Semantic Web technologies can be utilised, limited examples were used to demonstrate the application of the artefacts using at least one AI use case.

Table 7.9: An overview of the proof-of-concept implementations that demonstrate applicability of the thesis artefacts

Artefact	**Applicability in**	**proof-of-concept presented in**
AIRO and VAIR	Modelling AI use cases and their associated risks	Section 4.4
Annex III high-risk AI determinator	Determining high-risk AI systems	Section 5.1
AIUP	Using AIUP for expressing intended purpose within an `aiup:UsePolicy`	Subsection 5.2.2
Queries for generating AI and risk documentation	Retrieving AI and risk information	Section 5.3
AICat	Using AICat for providing metadata of AI use cases and their components in an `aicat:Catalog`	Subsection 5.4.2
AI Cards	Providing a summarised view of an AI use case	Section 6.4

CHAPTER

EIGHT

CONCLUSION

This chapter concludes the work presented in this thesis by revisiting the research question and objectives in Section 8.1 to demonstrate the extent to which they have been achieved. In Section 8.2, a summary of the contributions is provided. Section 8.3 refers to the impact the contributions have already made. Section 8.4 outlines potential directions for future work based on the research presented in this thesis. The thesis concludes with final remarks in Section 8.5.

8.1 Addressing the Research Question through Fulfilment of the Objectives

As a reminder, the research question of this thesis was:

To what extent can Semantic Web technologies facilitate compliance with the EU AI Act's risk management, documentation, and registration requirements?

The *extent* question was mainly determined through fulfilment of RO2 and RO3, with influence from RO4. Achieving these objectives was not possible without fulfilling RO1.

Fulfilment of RO1

The first research objective (RO1) concerned analysing the AI Act to identify the information requirements related to risk management, documen-

tation, and registration. To address this objective, the analysis (presented in Chapter 3), covered the following aspects:

- To fulfil RO1(a), an analysis of the risk-based classification for AI systems was presented in Section 3.2. The analysis covered Article 6 high-risk AI systems and Article 5 prohibited AI practices. Regarding high-risk AI systems, while all the AI applications specified in Annex III were investigated, only two examples of the Union harmonisation legislation (Annex I) were briefly discussed, given that a complete assessment was not feasible due to the depth of Annex I legislation. Although this analysis was limited, it showed that the overall approach in capturing the core concepts can be applied for determining Annex I high-risk AI systems. This can inform future efforts in structuring Annex I analysis and in turn lead to enhancing AIRO, VAIR, and the SHACL shapes by inclusion of Annex I conditions.

- To fulfil RO1(b), the intended purpose was conceptualised using the concepts identified for determination of high-risk and prohibited AI systems in Section 3.3.

- To fulfil RO1(c), the AI Act's risk management system requirements, outlined in Article 9, were analysed in Subsection 3.4.1 and technical documentation requirements, referred to in Article 11 and Annex IV, were detailed in Subsection 3.4.2. In this, relevant ISO/IEC standards, including ISO/IEC 42001 [37] and ISO/IEC 23894 [39], which are candidates for harmonisation, were used to guide the analysis.

- To fulfil RO1(d), registration obligations for high-risk AI providers and deployers (Article 49 and Annex VIII) and the characteristics of the EU database (Article 71) were detailed in Section 3.5.

This analysis only covered a subset of the AI Act, which was previously defined within the research scope (Subsection 1.2.1). Despite this, it is sufficient for addressing the research question since it investigated the key articles related to risk-based classification of the Act, risk management, technical documentation, and registration. It is worth noting that the obligations for general-purpose AI models were excluded from the scope, given that these recently-introduced obligations seem to be separate from the EU product safety framework (specified in NLF), adopted for AI systems in the Act. This suggests that analysing such requirements for developing semantic models requires additional guidelines, which are not currently available.

Fulfilment of RO2

To meet RO2, Chapter 4 introduced AIRO and VAIR as ontologies for modelling AI use cases and their associated risk management information in a FAIR manner. The reliance of the ontologies on the interpretation of the AI Act and the relevant standards might limit their use. However, being implemented as open knowledge graphs, both ontologies can be adopted and enhanced freely through inheritance or concept specialisations. They can also be reused in other ontologies within separate namespaces. While this reuse might be limited due to the specific conceptualisation of risk in the ontologies, the adoption of the widely-used ISO 31000 series in shaping this conceptualisation could enhance the expected broad applicability and reuse of the ontologies.

Fulfilment of RO3

AIRO and VAIR provided the basis upon which the Semantic Web-based approaches were developed in Chapter 5 to fulfil RO3. These FAIR approaches are listed in the following:

- RO3(a) was fulfilled by developing a set of SHACL shapes for determining Annex III conditions for high-risk AI systems (Section 5.1),
- RO3(b) was fulfilled by AIUP as a policy profile that extends ODRL to describe intended, precluded, and conditions of AI use as described by the AI Act (Section 5.2),
- RO3(c) was fulfilled by implementing SPARQL queries for information retrieval to generate documentation, particularly the AI Cards (Section 5.3),
- RO3(d) was fulfilled by implementing AICat as an extension of DCAT for cataloguing AI systems and their components in AI registries. AICat was designed based on the information elements that should be provided upon registration into the EU database (Section 5.4).

A caveat regarding these artefacts is that their applicability has been evaluated using only a limited set of examples, rather than comprehensive real-world use cases. However, it is important to note that such real-world use cases demonstrating compliance with the AI Act have not yet materialised and may not do so for several years to come.

Fulfilment of RO4

RO4 was accomplished by introducing AI Cards in Chapter 6. The AI Cards framework was designed in alignment with the requirements of the AI Act in regard to technical and risk documentation. It consists of a human-readable representation, as well as an open, interoperable machine-readable specification, which is empowered by AIRO and VAIR.

Addressing the Research Question

In the light of the accomplishment of the research objectives, the research question can be evaluated. As mentioned earlier, fulfilment of RO2 and RO3 directly correspond to the *extent* aspect of the research question by providing FAIR ontologies and approaches, which enable modelling AI and risk information and support implementation of specific tasks to facilitated compliance with the AI Act. Therefore, the *extent* question has been addressed by demonstrating the capabilities and benefits of different Semantic Web technologies in modelling AI and risk information, validating this information for determining Annex III high-risk AI systems, expressing the intended purpose within AI use policies, querying the information in particular for generating documentation, and cataloguing AI systems for registration into the EU database. Though this covers the scope of the research question, there are numerous tasks within the AI Act landscape that can potentially be addressed using Semantic Web technologies, including determining compliance obligations to which a system or entity is subjected, compliance checking, mapping the forthcoming harmonised standards to the Act's requirements, and conducting fundamental rights impact assessments, to name a few. Another caveat, as mentioned earlier, is that the thesis artefacts have only been evaluated against a limited number of use cases, therefore their applicability in real-world use cases, which are independently assessed against the AI Act, is yet to be evaluated as these use cases become available. Nevertheless, this thesis provides the foundation for such higher-function applications. It is also important to acknowledge that this thesis only offers one possible route to address the research question, and future research may lead to alternative approaches.

Supported by the evidence provided throughout this thesis, especially in Chapter 4, Chapter 5, and Chapter 6, the thesis concluded that Semantic Web technologies offer promising solutions towards compliance with requirements of the EU AI Act regarding risk management, documentation, and registration. It also demonstrated the impact of knowledge engineering approaches in alleviating compliance complexities and in turn reducing the

considerable amount of effort the implementation and enforcement of the AI Act require. Moreover, the thesis demonstrated a high degree of adaptability in the proposed approaches—an essential feature that is required to facilitate implementation of the necessary updates arising from future delegated and implementing acts, case laws, policies, standards, and guidelines.

8.2 Contributions

This research yielded two major and one minor contributions, described in the following.

8.2.1 Major Contribution: A Set of FAIR Artefacts to Assist with Compliance with the AI Act

In the absence of structured, open, interoperable, and extensible RegTech solutions to support the AI Act, this thesis introduces the *first* set of FAIR semantic models and approaches to assist with the AI Act's compliance tasks.

AIRO and VAIR advance the state of the art by providing novel AI risk ontologies and vocabularies based on the EU AI Act and ISO/IEC standards related to AI risks, including ISO/IEC 22989 [36], ISO/IEC 23894 [39] and ISO 31073 [42].

Compared to the state of the art AI risk ontologies, reviewed in Subsection 2.3.3, AIRO and VAIR are currently the only open semantic models developed as per the AI Act requirements. From the existing ontologies that model AI risk-related information, reviewed in Subsection 2.3.3, DocBiasO Ontology [114] and AIO [115] are the closest to this work, however both lack the holistic view of modelling AI use cases wherein technical information, context of use, and risk management information are provided. In addition to inclusion of the aforementioned aspects in AIRO and VAIR, they set up the foundation for implementing RegTech solutions for the AI Act, as demonstrated in Chapter 5. Provided under open licenses and in a modular form, AIRO and VAIR can be enhanced and extended to meet the sector-specific requirements and provisions of the forthcoming guidelines and standards. Further, they can be expanded as per national AI policies and guidelines that might be issued by Member States. These extensions can co-exist under separate namespaces reusing the open, consistent, and standardised representation of information offered by AIRO and/or VAIR. In this case, a key advantage of utilising open knowledge modelling standards, as opposed to proprietary solutions, is in providing an open and transparent way to conduct mappings, assess alignment with the AI Act, and check their

consistency. This is specially helpful in supporting the coordination among Member States and market surveillance authorities to tackle differences in interpretations of the AI Act—in particular those related to fundamental rights, that might disrupt the functioning of the EU digital single market.

The overall approach towards the design of AIRO and VAIR is also a *minor contribution*, which provides a manageable and modular approach in constructing consistent and integrable vocabularies for AI risks. This is particularly of importance for building the specific AI risk ontologies and taxonomies; such as sector- or technology-specific ones.

The thesis further demonstrated the application of AIRO and VAIR in implementation of a SHACL-based rule-checking tool for determination of Annex III high-risk AI systems. The SHACL shapes, as well as the tool, are limited in determining prohibited AI practices, Annex I high-risk AI systems, and exceptions to high-risk AI systems (Article 6(3)). However, using SHACL for modelling the high-risk conditions provides a transparent and codified approach that can be adopted for other categories and updated as per the forthcoming "*comprehensive list of practical examples of use cases of AI systems that are high-risk and not high-risk*" that should be provided by the Commission before 2 February 2026 (Article 6(5)). In addition, the semantic models of AI use cases can assist in discovering new patterns of high-risk AI systems that will emerge as the field progresses, which can be a helpful resource for the European Commission in updating the list of Annex III high-risk AI applications. Providing the classification rules using SHACL shapes also allows expressing additional rules that might be introduced by Member States, similar to the additional GDPR's DPIA requirements that have been introduced by EU Member States. Further, it enables comparing these additional rules, if any, to identify nuances that could potentially fragment the digital single market.

AIUP, the ODRL profile for declaring AI use policies, enables describing and sharing the intended purpose of an AI system, the precluded uses, and conditions of use in a machine-readable and open format. AIUP enables expressing legally binding statements regarding usage of AI systems in a consistent and transparent manner. This consistency is helpful in comparing multiple offers made by different AI providers, especially when procuring AI solutions. Moreover, expressing agreements between providers and deployers using AIUP ensures transparency in adoption of AI solutions, particularly by the public sector. Utilising AIUP for agreements also can foster better auditability in investigation of incidents or non-compliance, and in turn can promote accountability. Although the profile includes limited parties and actions, its extensibility feature allows further enhancements to express usage policies related to AI systems, general-purpose AI models, and datasets, as

required by the AI Act.

Information retrieval using SPARQL queries provides a dynamic and automated approach in generating AI documentation. While modelling AI use cases using AIRO and VAIR assist in maintaining the information and ensuring the consistency, SPARQL queries enable extracting information to create customised documentation for multiple purposes. This thesis provided the evidence on how a subset of the AI Act's technical documentation, in the form of AI Cards, can be created using SPARQL. It also showed how interoperability can be achieved through the use of the proposed semantic models and queries to integrate information from multiple sources. In practice, this approach reduces the resources required for managing the extensive information about AI systems and their risks that needs to be maintained and updated constantly. Moreover, it can be a helpful tool for auditors, enabling them to run complex queries to uncover non-compliance with the Act, as well as wider AI trustworthiness issues related to an AI system. Using a consistent and standardised model can also significantly facilitate joint investigations, that are performed through collaboration of multiple market surveillance authorities or the joint efforts of the Commission with market surveillance authorities (Article 74(11)).

Ultimately, AICat was proposed as an extension of DCAT to provide information about AI systems and their components in catalogues to support data governance in AI registries, including the EU database of high-risk AI. AICat enhances management of open metadata regarding AI systems in a machine-readable, searchable, and interoperable manner—the desired characteristics for the EU database of high-risk AI systems (Article 71). Furthermore, the metadata of the upcoming Commission-issued lists of high-risk and not high-risk AI use cases, which was mentioned earlier, can be delivered through AICat catalogues. AICat can be also expanded to be used as a mechanism for sharing the insights gained through the regulatory learning process, including those from regulatory sandboxes, between authorities involved in enforcement of the AI Act, such as market surveillance authorities within a Member State or across Member States. Using AICat and AIUP together facilitates discovery, integration, and sharing information and policies associated with AI systems and components amongst the stakeholders involved in the AI value chain based on the existing proven mechanism of (open) data portals.

This major contribution advances the state of art by providing a set of open, common, interoperable, and extensible artefacts, upon which researchers can build specific risk ontologies and tools for conducting, documenting, and sharing AI risk and impact assessments. AIRO and VAIR can be leveraged for annotating AI use cases and incidents to classify, collate,

and compare AI risks and risk assessment activities over time. These ontologies therefore can be utilised as an open and transparent instrument for comparing risk and impact assessments performed by participation of various stakeholders in the public interest. The Semantic Web-based RegTech solutions proposed in this thesis can be adopted by actors across the AI value chain, in particular AI providers and deployers who can enjoy the benefits of these solutions in alleviating compliance tasks and further reducing the compliance costs. The tool for determining Annex III high-risk AI can be beneficial in empowerment of AI end-users, AI subjects, and potentially affected stakeholders regarding the AI Act and therefore contribute to increasing public legal awareness. As discussed so far, the artefacts of the thesis can be adopted to facilitate the enforcement of the AI Act. In addition, they can be leveraged in development of harmonised standards related to risk management and documentation by the European standardisation bodies, notably CEN/CLC JTC 21, as state of the art approaches.

8.2.2 Major Contribution: the AI Cards Framework

The AI Cards framework serves as the second major contribution of this thesis, offering the first holistic framework for documenting a *given use* of an AI system that encompasses information regarding technical specifications, context of use, and risk management in both human- and machine-readable formats, based on the requirements of the AI Act. In addition, as discussed in Subsection 7.4.2, the AI Cards framework distinguishes itself from the state of the art by providing a summarised view of AI and risk information that can be automatically retrieved through querying the machine-readable specification of an AI use case. Supported by open formats of data representation, the AI Cards framework provides the adaptability required for addressing the future changes to the technical documentation information requirements, which will be introduced by the Commission through delegated acts (Article 11(3)). This format of representation also allows automation, as a much-needed feature to reduce the efforts and resources needed for generating and maintaining documentation [153, 14].

Co-design and evaluation of documentation frameworks have not been a common practice until recently—well-known documentation approaches such as Datasheets [134] and Model Cards [147] are neither developed through stakeholder engagement, nor formally evaluated. However, some recently-proposed documentation frameworks, which were mentioned earlier in Subsection 2.5.3, including [64, 155], consider stakeholder engagement and formal evaluation of their work. To actively involve stakeholders, the AI Cards has been developed in collaboration with JRC researchers, who were involved in

EU digital policymaking. The framework was further evaluated by 50 survey participants with expertise in AI and law.

In terms of benefits to stakeholders, the AI Cards assists AI providers in document generation, AI deployers in comparison of multiple AI systems for purposes such as AI procurement, and conformity assessment bodies in compliance checking tasks. In addition, the AI Cards as a framework has the potential to be adopted and used by various stakeholders, as the user study in Subsection 7.4.3 showed. While the AI Cards framework is characterised as a structured and adaptable AI documentation template that can be used for multiple purposes by variety of stakeholders, it is not a *one-size-fit-all* artefact. It rather offers a customisable solution that can be modified in accordance with the needs of specific stakeholders or particular sectors. This feature is rooted in provision of AI and risk information in a knowledge graph, upon which a variety of queries can be run to generate a specialisation of the AI Cards. Thus, the AI Cards can serve as an insightful resource for the Commission in development of the "*simplified technical documentation form*" for SMEs, as required by Article 11(1).

8.2.3 Minor Contribution: Analysis of the AI Act

The EU AI Act is a relatively new piece of regulation, considering that it has been published in July 2024. Therefore, as mentioned multiple times, there are limited resources and no authoritative guidelines to assist with interpretation of the final version of the AI Act. In this void of related work, this thesis makes a contribution by providing detailed, transparent, and independent analysis of the following articles and annexes:

- Articles 3(12) - definition of intended purpose,
- Article 5 - Prohibited AI practices,
- Article 6 - Classification rules for high-risk AI systems, Annex I - List of Union harmonisation legislation, and Annex III - High-risk AI systems referred to in Article 6(2),
- Article 9 - Risk management system,
- Article 11 - Technical documentation and Annex IV- Technical documentation referred to in Article 11(1),
- Article 49 - Registration, Article 71 - EU database for high-risk AI systems listed in Annex III, and Annex VIII - Information to be submitted upon the registration of high-risk AI systems in accordance with Article 49.

The interpretation offers a streamlined, yet extensive, view of risk management and technical documentation requirements by endorsing guidance from key recently-published ISO/IEC AI standards, namely ISO/IEC 42001 on AI management system and ISO/IEC 23894 on risk management. Implementation of the AI Act is a subject of controversy as some are concerned that its reliance on international standards might result in significant influence from the large digital platforms who provide most of the technical experts to the subcommittees that shape these standards [8], while some believe that the wide discretion in setting technical parameters, e.g. Annex III high risk categorisations, is not undertaken with sufficient transparency. This work offers a detailed, transparent, and traceable independent analysis that further led to a baseline technical solution for documenting and exchanging AI and risk information. As of August 2024, to the best of the author's knowledge, there are no publicly-available studies that cover the aforementioned aspects of the AI Act with the degree of details and clarity provided in this thesis.

Overall, this analysis is a helpful resource for those who need to develop an understanding of the risk management and documentation requirements of the AI Act. To be more specific, the analysis mainly benefits AI providers and deployers in understanding their obligations under the EU AI Act regarding risk management, documentation, and registration. Given that the analysis is made **openly accessible**, it can significantly assist SMEs and less-resourced organisations, such as NGOs and the public sector, who aim to deploy AI or already are AI deployers, in interpreting the AI Act. At the European level, this analysis can be used as an independent reference for the AI Office and the European Commission in development and structuring guidelines, as well as delegated and implementing acts. Notably, the 5-concept structure for describing high-risk AI systems can be adopted for structuring the list of practical examples of high-risk and not high-risk AI, that the European Commission is obliged to provide (Article 6(5)).

A summary of the key compliance and conformity assessments tasks that are supported by the contributions of this thesis is depicted in Figure 8.1.

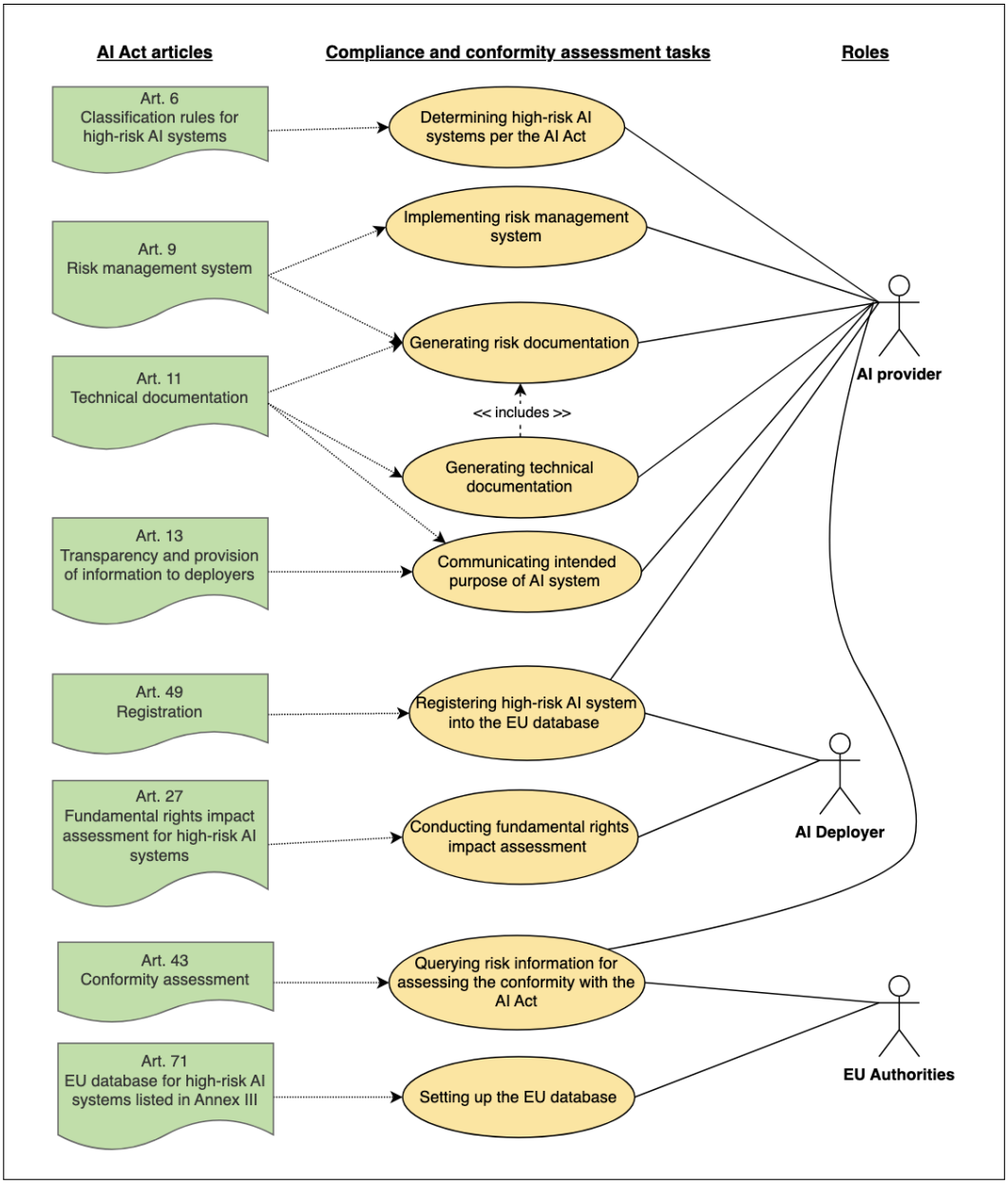


Figure 8.1: The key compliance and conformity assessments tasks supported by the thesis artefacts

8.3 Impact and Uptake of the Work

This research has been continually progressed with the AI Act's development process, till it was officially accepted as an EU law and published in the Official Journal of the European Union in July 2024. Consequently, the results of this work have increasingly attracted interest from the academics and wider audiences. This impact can be estimated through an analysis of the citations to the publications and the projects that reused AIRO and VAIR.

8.3.1 Analysis of Citations

As of August 2024, the 5 first-authored peer-reviewed publications related to this thesis have received a total of 25 citations (excluding self-citations) that the author is aware of. Table 8.1 and Table 8.2 provide an overview of the citations to the most influential papers stemmed from this work, which are "*AIRO: An Ontology for Representing AI Risks Based on the Proposed EU AI Act and ISO Risk Management Standards*" [70] and "*To Be High-Risk, or Not To Be—Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards*" [71]. In this overview, self-citations and non-English studies are excluded.

Table 8.1: Papers cited "AIRO: An Ontology for Representing AI Risks Based on the Proposed EU AI Act and ISO Risk Management Standards" [70] since its publication in 2022

Paper	Type	Year	Reason for citation
EA-Ontology: Ethical Assessment Ontology [110]	Conference paper	2022	Related work.
PaTrOnto, an ontology for patents and trademarks [205]	Workshop proceedings	2023	Related work.
OntoVAT, an ontology for knowledge extraction in VAT-related judgments [206]	Workshop proceedings	2023	Related work.
Using Ontological Knowledge and Large Language Model Vector Similarities to Extract Relevant Concepts in VAT-Related Legal Judgments [207]	Conference paper	2023	Related work.
An Approach to the Instantiation of the EU AI Act: A Level of Done Derivation and a Case Study from the Automotive Domain [208]	Conference paper	2023	Related work.
Towards Trustworthy-AI-by-Design Methodology for Intelligent Radiology Systems [209]	Conference paper	2023	Considers using AIRO in their future work.
Analysing and organising human communications for AI fairness assessment [210]	Journal paper	2024	Related work. Follows the iterative approach used in development of AIRO for building a conceptual model that captures information regarding AI fairness assessment processes in the public sector. Reuses definition provided in AIRO in their model.
SoK: How Artificial-Intelligence Incidents Can Jeopardize Safety and Security [211]	Conference paper	2024	Background information about AI incidents.
Leveraging Ontologies to Document Bias in Data [114]	Preprint	2024	Reuses AIRO in an ontology for documenting bias of ML (Doc-BiasO Ontology). Adopts the overall approach used in development of AIRO and VAIR.
A DSL for Testing LLMs for Fairness and Bias [212]	Conference paper	2024	Related work

Table 8.2: Papers cited "To Be High-Risk, or Not To Be—Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards" [71] since its publication in 2023

Paper	**Type**	**Year**	**Reason for citation**
Reading the drafts of the AI Act with a technical lens [213]	Workshop paper	2023	Related work.
Design of a Quality Management System based on the EU Artificial Intelligence Act [7]	Preprint	2024	Provides an implementation of the high-risk AI determinator using the 5-concept model.
Co-designing an AI Impact Assessment Report Template with AI Practitioners and AI Compliance Experts [155]	Preprint	2024	Uses the 5 concepts for describing use of an AI system within the *Impact Assessment Report Template*. Suggests utilising VAIR to guide filling out the AI system's use section of the report.
The Atlas of AI Incidents in Mobile Computing: Visualizing the Risks and Benefits of AI Gone Mobile [214]	Conference paper	2024	Uses the 5 concepts format to visualise AI use cases in order to communicate AI risks in a simplified and effective manner. Uses the 5 concepts to assist with determining the risk category under the AI Act.
ExploreGen: Large Language Models for Envisioning the Uses and Risks of AI Technologies [215]	Preprint	2024	Identifies the 5 concepts from AI use cases using an LLM-based framework and further uses the concepts to assist with determining the risk category under the AI Act.
Good Intentions, Risky Inventions: A Method for Assessing the Risks and Benefits of AI in Mobile and Wearable Uses [216]	Journal article	2024	Utilises the 5 concepts in LLM prompts to generate description of a use case and also uses them to assist with determining the risk category under the AI Act. Uses the 5 concepts for describing use of an AI system in the proposed *Risk Assessment Checklist for Mobile Computing*.
Leveraging Ontologies to Document Bias in Data [114]	Preprint	2024	Adopts the overall approach used in development of AIRO and VAIR
Information That Matters: Exploring Information Needs of People Affected by Algorithmic Decisions [217]	Preprint	2024	Background information about affected stakeholders.

8.3.2 Analysis of Ontology Reuse

Contribution to DPV

The Data Privacy Vocabulary (DPV), which is being developed and maintained by the W3C Data Privacy Vocabularies and Controls Community Group[1] (DPVCG), enables providing machine-readable specifications for personal data processing and use of technology based on legal requirements. While the first version of DPV [138] focused on modelling personal data activities, the scope of DPV 2.0 [57] has been expanded to cover the AI and digital regulatory landscape, partly owing to the author's work. DPV 2.0, in particular EU-AIAct[2] and TECH[3] extensions, includes concepts from AIRO and VAIR, due to the author's contributions to DPV. Thorough integration of AIRO and VAIR with DPV is underway.

Contribution to the PROTECT Project

In collaboration with Early Stage Researchers (ESR) across the PROTECT ITN network, a specialisation of AIRO for risks associated with the use of AI systems in the health sector was developed. The Health AI Risk Taxonomy (HART) provides a catalogue of known AI risks captured from real-world incidents indexed in the AIAAIC repository. An overview of HART is illustrated in Figure 8.2. Development of HART involved one legal researcher, two technology ethicists, and two knowledge engineers, including the author of this thesis. HART is available online at http://w3id.org/hart under the CC-BY-4.0 license.

In the compendium of PROTECT results, VAIR was used in A-ERAP[4]—a tool that automatically annotates AI use cases[5]. A-ERAP uses Named Entity Recognition (NER) and Named Entity Linking (NEL) techniques to annotate an input text according to VAIR and other ontologies developed by PROTECT researchers.

GitHub Projects Used AIRO

Two degree-required projects at Trinity College Dublin used AIRO, whose source code is available on GitHub. An MSc. project extended AIRO for

[1] https://www.w3.org/groups/cg/dpvcg/
[2] https://w3id.org/dpv/legal/eu/aiact
[3] https://w3id.org/dpv/tech
[4] https://tair.adaptcentre.ie/aerap.html
[5] See some use cases examples here: https://tair.adaptcentre.ie/use_cases/index-use.html

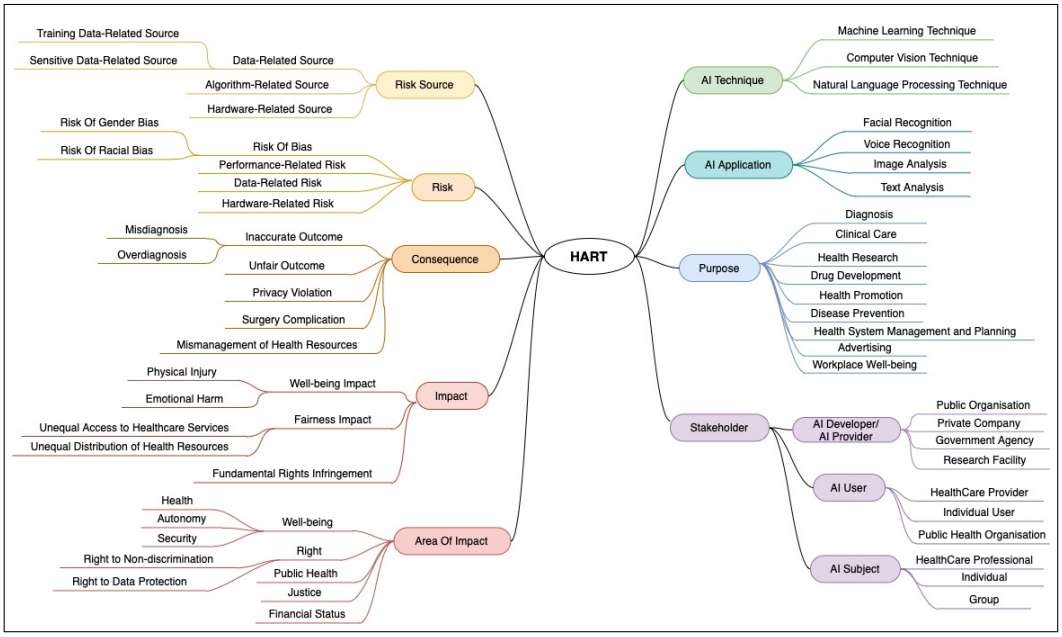


Figure 8.2: An overview of HART

AI bias, which contains different types of bias and enables documenting and reporting bias tests[6]. A final year project utilised LLMs to annotate AI incidents with AIRO concepts[7].

8.4 Directions for Future Work

8.4.1 Semantic Web Technologies for Compliance with Other Requirements of the AI Act

The AI Act is an extensive piece of legislation that has given rise to new challenges, a fraction of which are addressed in this thesis and in some of the related work. Yet, many of these challenges have remained unexplored. Two closely related requirements to the work presented in this thesis are requirements for fundamental rights impact assessments and general-purpose AI models. In the following, developing RegTech solutions for these two requirements is discussed as a possible direction for future work.

[6] https://github.com/drd00/msc-dissertation-files/blob/master/main.ttl
[7] https://github.com/paddyocallaghan/dissertation/tree/main/Annotated_incidents

Fundamental Rights Impact Assessment

The obligation for high-risk AI deployers to conduct *Fundamental Rights Impact Assessments (FRIA)*, set out in Article 27, has considerable overlaps with the risk management requirements, as illustrated in Table 8.3. In performing FRIA, AI deployers are permitted to use impact assessments that are already conducted by the AI provider (Article 27(2)). This implies a need for sharing risk and impact assessments among AI providers and deployers. As shown in this thesis, such exchange of information across the AI value chain can be supported by adoption of interoperable data formats.

Table 8.3: Summary of information requirements for FRIA

Article 27 clause	**Information requirement**
1a	Deployer's processes in which the system is used
1b	Period of time within which the system is used, Frequency of using the system
1c	Context of use, Categories of natural persons likely to be affected in a specified context, Groups likely to be affected in a specified context
1d	Risks of harm likely to impact the identified categories of persons or groups
1e	Information about implementation of human oversight measures
1f	Measures to address materialisation of risks

According to Article 27(5), one of the next steps for the AI Office is providing a template and an automated tool to simplify compliance with the FRIA requirements. The contributions of this work have great potential in supporting implementation of such tools. Furthermore, using this work alongside existing vocabularies, such as DPV [57], to support convergence of FRIA with other legally-required impact assessments, in particular the GDPR's DPIA, for conducting *shared impact assessments* [139] is a path worth exploring. With all these being said, the set of Semantic Web artefacts presented in this thesis can be extended for addressing the following research question:

To what extent can Semantic Web technologies facilitate compliance with the EU AI Act provisions regarding fundamental rights impact assessments and assist with the convergence of different impact assessments required by other digital regulations in the EU, in particular the GDPR's DPIA?

Obligations for General-Purpose AI Models

Under the AI Act, general-purpose AI models are subject to a set of obligations, listed in Article 53, regardless of their risk level. This suggests that these models are inherently high-risk. An additional set of obligations is set out in Article 55 for those general-purpose AI models that impose *systemic risks*[8]. Given the similarities between the obligations for general-purpose AI models, that include technical documentation and risk assessment, with those examined in this thesis, the outputs of this work, specifically AIRO, VAIR, AIUP, and AICat, can be customised as per the obligations outlined in Articles 53 and 55 to support compliance with general-purpose AI models. Therefore, this work can be used in addressing the following research question:

To what extent can Semantic Web technologies assist in addressing requirements of general-purpose AI models in regard to risk management and documentation and further facilitate information sharing between providers of general-purpose AI models and downstream providers, as required by the AI Act?

8.4.2 Navigating the Landscape of Digital Regulations within the EU and Beyond

Within the EU, a dense area of digital regulations has been established, wherein the pieces of legislation have overlapping scopes and mutual dependencies. Thus, it is highly likely that a given entity that incorporates AI, whether it is a smart toy or a chatbot or a social media platform, needs to comply with multiple EU regulations. While there have been studies looking into the interplay between these EU regulations, vide [218, 219, 220], navigating compliance with multiple EU regulations, and the harmonised standards associated with them, from a practical perspective is yet to be explored. In this, identification of the similar and overlapping requirements and further integrating them using a consistent and interoperable manner is essential. Examples of such overlapping requirements are risk management, impact assessments (as discussed earlier), transparency, documentation, and

[8] Article 3(65) defines systemic risk as "a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain".

reporting obligations.

In the global context, as a result of the EU's influence on technology-related rulemaking, developing AI regulations has become a priority for governments worldwide [4]. Thus, soon stakeholders will face a challenge in how to structure, document, and share information in the context of their AI systems or components such that this information assists with fulfilling different regulatory requirements of different legal jurisdictions, without impeding rapid progress in global markets. The EU-US Trade and Technology Council's report on trustworthy AI and risk management [221] indicates preliminary efforts from the political bodies, that govern AI within the EU and the US, to develop common terminology, measures, and thresholds, that can assist in addressing challenges of compliance with multiple AI regulations in different jurisdictions.

Against this background, a key question is how the artefacts proposed in this thesis can be reused and integrated with other RegTech solutions to address the challenges of compliance with multiple digital regulations in various jurisdictions. In a broader form, this question is stated as:

To what extent can Semantic Web-based artefacts presented in this thesis be used for compliance tasks related to risk and impact assessments, documentation, and sharing that stem from digital regulations within the EU or other jurisdictions?

8.4.3 Use of LLMs for Population of VAIR

AI risk is a fast evolving area of research—by way of evidence, during the final stages of writing this thesis, multiple AI risk taxonomies were published, including MIT's AI risk database[9] [1], Google DeepMind's taxonomy of GenAI misuse tactics [222], and the AI Risk Categorization Decoded (AIR 2024) [223]. Capturing and maintaining the current AI risk knowledge within VAIR, or any other risk taxonomies, is a resource-intensive task. Streamlining the taxonomy development process can be achieved through (semi-)automated ontology learning and generation. With the significant recent advancement in LLMs, novel knowledge engineering approaches that utilise LLMs have been proposed recently (vide [224, 225, 226, 227]). A notable work focused on AI risks is ExploreGen [215], mentioned earlier in Subsection 8.3.1, that uses LLMs to identify the 5 key concepts, proposed by this thesis, from AI use cases. While the idea of LLM-supported AI risk taxon-

[9] https://airisk.mit.edu/

omy generation is appealing, the risks of using LLMs, such as automation bias[10] and hallucination, can have severe consequences. This does not mean that LLMs should not be utilised for knowledge engineering, but it highlights the serious need for risk and impact assessments and quality assurance. This discussion leads to the following research question:

To what extent can Large Language Models (LLMs) facilitate population and maintenance of AI risk taxonomies?

This thesis concludes by stating these 3 paths for further work. By building the foundation for the future research, *the ending of this thesis is only a beginning in disguise.*

8.5 Final Remarks

With the publication of the EU AI Act and its subsequent enforcement, the AI value chain is now faced with a new set of requirements and obligations to comply with. This thesis only investigated a small, yet crucial, subset of the challenges arising from these requirements. With adopting an open, transparent, standardised, interoperable, and extensible approach towards addressing the challenges, this thesis established a foundation upon which a range of solutions for implementation and enforcement of the AI Act and, more broadly, for trustworthy implementation and use of AI can be developed.

While the contributions of this work have already received positive feedback from academics, industry actors, policymakers, and standardisers, it is to be hoped that what is proposed in this thesis contributes to shaping an ecosystem wherein AI systems are not only legal, but —above all— are trustworthy and ethical.

[10] Uncritically accepting the outputs of an algorithm.

BIBLIOGRAPHY

[1] Peter Slattery, Alexander K Saeri, Emily A C Grundy, Jess Graham, Michael Noetel, Risto Uuk, James Dao, Soroush Pour, Stephen Casper, and Neil Thompson. *The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence*. URL: https://cdn.prod.website-files.com/669550d38372f33552d2516e/66bc918b580467717e194940_The%20AI%20Risk%20Repository_13_8_2024.pdf (visited on 08/15/2024).

[2] National Institute of Standards and Technology (NIST). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. 2023. DOI: 10.6028/NIST.AI.100-1.

[3] Spyros Makridakis. "The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms". In: *Futures* 90 (2017), pp. 46–60. ISSN: 0016-3287. DOI: 10.1016/j.futures.2017.03.006.

[4] Nathalie A. Smuha. "From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence". In: *Law, Innovation and Technology* 13.1 (2021), pp. 57–84. DOI: 10.1080/17579961.2021.1898300.

[5] European Parliament, Council of the European Union. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. June 2024. URL: http://data.europa.eu/eli/reg/2024/1689/oj.

[6] Rostam J Neuwirth. "Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA)". In: *Computer Law & Security Review* 48 (2023).

[7] Henryk Mustroph and Stefanie Rinderle-Ma. *Design of a Quality Management System based on the EU Artificial Intelligence Act*. 2024. arXiv: 2408.04689 [cs.SE]. URL: https://arxiv.org/abs/2408.04689.

[8] Michael Veale and Frederik Zuiderveen Borgesius. "Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach". In: *Computer Law Review International* 22.4 (2021), pp. 97–112. DOI: doi:10.9785/cri-2021-220402.

[9] European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs. *Commission notice The 'Blue Guide' on the implementation of EU product rules 2022*. 2022/C 247/01. 2022. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2022.247.01.0001.01.ENG (visited on 08/17/2024).

[10] Gabriele Mazzini and Salvatore Scalzo. "The proposal for the Artificial Intelligence Act: considerations around some key concepts". In: *Camardi (a cura di), La via europea per l'Intelligenza artificiale* (2023).

[11] Alessio Tartaro. "Towards European Standards Supporting the AI Act: Alignment Challenges on the Path to Trustworthy AI". In: *Proceedings of the AISB Convention*. 2023, pp. 98–106.

[12] *Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council Text with EEA relevance*. 2012. URL: http://data.europa.eu/eli/reg/2012/1025/oj.

[13] *C(2023)3215 – Standardisation request M/593 COMMISSION IMPLEMENTING DECISION of 22.5.2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence*. May 2023. URL: https://ec.europa.eu/growth/tools-databases/enorm/mandate/593_en.

[14] Laura Lucaj, Patrick van der Smagt, and Djalel Benbouzid. "AI Regulation Is (not) All You Need". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '23.

Chicago, IL, USA: Association for Computing Machinery, 2023, pp. 1267–1279. DOI: 10.1145/3593013.3594079.

[15] Eleni-Maria Kalogeraki and Nineta Polemi. "A taxonomy for cybersecurity standards". In: *Journal of Surveillance, Security and Safety* 5.2 (2024). ISSN: 2694-1015. DOI: 10.20517/jsss.2023.50.

[16] Paul Fehlinger. *Enabling the responsible use of technology at scale – Why Europe needs a regulatory technology innovation ecosystem*. Tech. rep. Sitra, 2023.

[17] Mark D. Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3.1 (2016). ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18.

[18] Christian Bizer, Tom Heath, and Tim Berners-Lee. "Linked Data: The Story so Far". In: *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. IGI Global, 2011, pp. 205–227. ISBN: 978-1-60960-593-3. DOI: 10.4018/978-1-60960-593-3.ch008.

[19] Luis-Daniel Ibáñez, Ian Millard, Hugh Glaser, and Elena Simperl. "An Assessment of Adoption and Quality of Linked Data in European Open Government Data". In: *The Semantic Web – ISWC 2019*. Ed. by Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon. Springer International Publishing, 2019, pp. 436–453. ISBN: 978-3-030-30796-7.

[20] Cleyton Mário de Oliveira Rodrigues, Frederico Luiz Gonçalves de Freitas, Emanoel Francisco Spósito Barreiros, Ryan Ribeiro de Azevedo, and Adauto Trigueiro de Almeida Filho. "Legal ontologies over time: A systematic mapping study". In: *Expert Systems with Applications* 130 (2019), pp. 12–30. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2019.04.009.

[21] *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)*. 2016. URL: http://data.europa.eu/eli/reg/2016/679/oj.

[22] Harshvardhan J. Pandit. "Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance". PhD thesis. School of Computer Science and Statistics, Trinity College Dublin, 2020.

[23] Tek Raj Chhetri, Anelia Kurteva, Rance J. DeLong, Rainer Hilscher, Kai Korte, and Anna Fensel. "Data Protection by Design Tool for Automated GDPR Compliance Verification Based on Semantically Modeled Informed Consent". In: *Sensors* 22.7 (2022). ISSN: 1424-8220. DOI: 10.3390/s22072763.

[24] Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. "PrOnto: Privacy Ontology for Legal Compliance". In: *Proceedings of the 18th European Conference on Digital Government (ECDG)*. 2018, pp. 142–151.

[25] *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)*. 2022. URL: http://data.europa.eu/eli/reg/2022/868/oj.

[26] Beatriz Esteves, Víctor Rodríguez Doncel, Harshvardhan J Pandit, and Dave Lewis. "Semantics for Implementing Data Reuse and Altruism Under EU's Data Governance Act". In: *Knowledge Graphs: Semantics, Machine Learning, and Languages*. Vol. 56. IOS Press, 2023, pp. 210–226. DOI: 10.3233/SSW230015.

[27] The EU Member States. *Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union Consolidated version of the Treaty on European Union*. 2016/C 202/01. 2016.

[28] European Commission, Directorate-General for Communications Networks, Content and Technology. *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts) and amending certain Union legislative acts*. Apr. 2021. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 (visited on 08/18/2024).

[29] Council of the European Union. *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach*. Nov. 2022. URL: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf (visited on 08/18/2024).

[30] Council of the European Union. *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artifi-*

cielle) et modifiant certains actes législatifs de l'Union - Text de compromis de la présidence - Version consolidée. June 2022. URL: https://artificialintelligenceact.eu/wp-content/uploads/2022/06/AIA-FRA-Consolidated-Version-15-June.pdf (visited on 08/18/2024).

[31] European Parliament. *Artificial Intelligence Act Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) (Ordinary legislative procedure: first reading)*. June 2023. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html (visited on 08/18/2024).

[32] European Parliament, Council of the European Union. *PROVISIONAL AGREEMENT RESULTING FROM INTERINSTITUTIONAL NEGOTIATIONS - Proposal for a regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts 2021/0106(COD) (COM(2021)0206 – C9-0146(2021) – 2021/0106(COD))*. Feb. 2024. URL: https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf (visited on 08/18/2024).

[33] European Parliament, Council of the European Union. *CORRIGENDUM to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ..... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) $P9_TA(2024)0138(COM(2021)0206–C9–0146/2021–2021/0106(COD))$*. Apr. 2024. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf (visited on 08/18/2024).

[34] David Fernández-Llorca, Emilia Gómez, Ignacio Sánchez, and Gabriele Mazzini. "An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI". In: *Artificial Intelligence and Law* (2024). ISSN: 1572-8382. DOI: 10.1007/s10506-024-09412-y.

[35] *Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence*. Dec.

2022. URL: https://ec.europa.eu/docsroom/documents/52376 (visited on 08/18/2024).

[36] ISO/IEC. *ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*. International Organization for Standardization (ISO), 2022. URL: https://www.iso.org/standard/74296.html.

[37] ISO/IEC. *ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system*. International Organization for Standardization (ISO), 2023. URL: https://www.iso.org/standard/81230.html.

[38] Josep Soler Garrido, Delia Fano Yela, Cecilia Panigutti, Henrik Junklewitz, Ronan Hamon, Tatjana Evas, Antoine-Alexandre André, and Salvatore Scalzo. *Analysis of the preliminary AI standardisation work plan in support of the AI Act*. Tech. rep. Joint Research Centre (Seville site), 2023.

[39] ISO/IEC. *ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management*. International Organization for Standardization (ISO), 2023. URL: https://www.iso.org/standard/77304.html.

[40] ISO. *ISO 31000:2018 Risk management — Guidelines*. Last reviewed and confirmed in 2023. International Organization for Standardization (ISO), 2018. URL: https://www.iso.org/standard/65694.html.

[41] ISO. *ISO Guide 73:2009 Risk management — Vocabulary*. Status: Withdrawn. International Organization for Standardization (ISO), 2009. URL: https://www.iso.org/standard/44651.html.

[42] ISO. *ISO 31073:2022 Risk management — Vocabulary*. International Organization for Standardization (ISO), 2022. URL: https://www.iso.org/standard/79637.html.

[43] "IEEE Standard Model Process for Addressing Ethical Concerns during System Design". In: *IEEE Std 7000-2021* (2021), pp. 1–82. DOI: 10.1109/IEEESTD.2021.9536679.

[44] ISO/IEC. *ISO/IEC DIS 12792 Information technology — Artificial intelligence — Transparency taxonomy of AI systems*. Status: Under development. International Organization for Standardization (ISO), 2024. URL: https://www.iso.org/standard/84111.html (visited on 09/16/2024).

[45] ISO/IEC. *ISO/IEC TR 24029-1:2021 Artificial Intelligence (AI) — Assessment of the robustness of neural networksPart 1: Overview*. International Organization for Standardization (ISO), 2021. URL: https://www.iso.org/standard/77609.html.

[46] ISO/IEC. *ISO/IEC TR 24027:2021 Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making*. International Organization for Standardization (ISO), 2021. URL: https://www.iso.org/standard/77607.html.

[47] Pascal Hitzler. "A review of the semantic web field". In: *Communications of the ACM* 64.2 (2021), pp. 76–83.

[48] Olaf Hartig, Pierre-Antoine Champin, Gregg Kellogg, and Andy Seaborne. *RDF 1.2 Concepts and Abstract Syntax*. W3C Working Draft. URL: https://www.w3.org/TR/rdf12-concepts/.

[49] W3C OWL Working Group. *OWL 2 Web Ontology Language*. W3C Recommendation. 2012. URL: https://www.w3.org/TR/owl2-overview/.

[50] Alistair Miles and Sean Bechhofer. *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation. 2009. URL: http://www.w3.org/TR/skos-reference.

[51] Steve Harris and Andy Seaborne. *SPARQL 1.1 Query Language*. W3C Recommendation. 2013. URL: https://www.w3.org/TR/sparql11-query/.

[52] Holger Knublauch and Dimitris Kontokostas. *Shapes Constraint Language (SHACL)*. W3C Recommendation. 2017. URL: https://www.w3.org/TR/shacl/.

[53] Renato Iannella, Michael Steidl, Stuart Myles, and Víctor Rodríguez-Doncel. *ODRL Version 2.2 Ontology*. W3C Recommendation. 2017. URL: http://www.w3.org/ns/odrl/2/.

[54] Riccardo Albertoni, David Browning, Simon J D Cox, Alejandra Gonzalez Beltran, Andrea Perego, and Peter Winstanley. *Data Catalog Vocabulary (DCAT) - Version 3*. W3C Recommendation. 2024. URL: https://www.w3.org/TR/vocab-dcat-3/.

[55] Martin Ebers. *Truly Risk-Based Regulation of Artificial Intelligence - How to Implement the EU's AI Act*. 2024. DOI: 10.2139/ssrn.4870387.

[56] María Poveda-Villalón, Alba Fernández-Izquierdo, Mariano Fernández-López, and Raúl García-Castro. "LOT: An industrial oriented ontology engineering framework". In: *Engineering Applications of Artificial Intelligence* 111 (2022).

[57] Harshvardhan J. Pandit, Beatriz Esteves, Georg P. Krog, Paul Ryan, Delaram Golpayegani, and Julian Flake. "Data Privacy Vocabulary (DPV) – Version 2". In: (2024). arXiv: 2404.13426 [cs.CY]. URL: https://arxiv.org/abs/2404.13426.

[58] Diego Berrueta and Jon Phipps. *Best Practice Recipes for Publishing RDF Vocabularies*. W3C Recommendation. 2008. URL: https://www.w3.org/TR/swbp-vocab-pub/.

[59] Farias Lóscio Bernadette, Caroline Burle, and Newton Calegari. *Data on the Web Best Practices*. W3C Recommendation. 2017. URL: https://www.w3.org/TR/dwbp/.

[60] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. "OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation". In: *International Journal on Semantic Web and Information Systems (IJSWIS)* 10.2 (2014), pp. 7–34.

[61] María Poveda-Villalón, Paola Espinoza-Arias, Daniel Garijo, and Oscar Corcho. "Coming to Terms with FAIR Ontologies". In: *Knowledge Engineering and Knowledge Management*. Ed. by C. Maria Keet and Michel Dumontier. Springer International Publishing, 2020, pp. 255–270. ISBN: 978-3-030-61244-3.

[62] Daniel Garijo and María Poveda-Villalón. "Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web". In: *Applications and Practices in Ontology Design, Extraction, and Reasoning*. Vol. 49. Studies on the Semantic Web. IOS Press, 2020, pp. 39–54.

[63] Cecilia Panigutti, Ronan Hamon, Isabelle Hupont, David Fernandez Llorca, Delia Fano Yela, Henrik Junklewitz, Salvatore Scalzo, Gabriele Mazzini, Ignacio Sanchez, Josep Soler Garrido, et al. "The role of explainable AI in the context of the AI Act". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 2023, pp. 1139–1150.

[64] Isabelle Hupont, David Fernández-Llorca, Sandra Baldassarri, and Emilia Gómez. "Use case cards: A use case reporting framework inspired by the european AI act". In: *Ethics and Information Technology* 26.2 (2024).

[65] José Luis González Álvarez, Juan José López Ossorio, Carlota Urruela, and Marina Rodríguez Díaz. "Integral Monitoring System in Cases of Gender Violence VioGén System". In: *Behavior & Law Journal* 4.1 (2018).

[66] John Bulles, Hennie Bouwmeester, and Anouschka Ausems. "A Best Practice for the Analysis of Legal Documents". In: *On the Move to Meaningful Internet Systems: OTM 2019 Workshops*. Ed. by Christophe Debruyne, Hervé Panetto, Wided Guédria, Peter Bollen, Ioana Ciuciu, George Karabatis, and Robert Meersman. Springer International Publishing, 2020, pp. 106–116. ISBN: 978-3-030-40907-4.

[67] Delaram Golpayegani, Joshua Hovsha, Leon W. S. Rossmaier, Rana Saniei, and Jana Mišić. "Towards a Taxonomy of AI Risks in the Health Domain". In: *2022 Fourth International Conference on Transdisciplinary AI (TransAI)*. 2022, pp. 1–8. DOI: 10.1109/TransAI54797.2022.00007.

[68] Gavin Abercrombie, Djalel Benbouzid, Paolo Giudici, Delaram Golpayegani, Julio Hernandez, Pierre Noro, Harshvardhan Pandit, Eva Paraschou, Charlie Pownall, Jyoti Prajapati, Mark A. Sayre, Ushnish Sengupta, Arthit Suriyawongkul, Ruby Thelot, Sofia Vei, and Laura Waltersdorfer. *A Collaborative, Human-Centred Taxonomy of AI, Algorithmic, and Automation Harms*. 2024. arXiv: 2407.01294 [cs.LG]. URL: https://arxiv.org/abs/2407.01294.

[69] European Commission and Directorate-General for Communications Networks, Content and Technology. *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment*. Publications Office of the European Unio, 2020. DOI: 10.2759/791819.

[70] Delaram Golpayegani, Harshvardhan J. Pandit, and Dave Lewis. "AIRO: An ontology for representing AI risks based on the proposed EU AI Act and ISO risk management standards". In: *Towards a Knowledge-Aware AI*. Vol. 55. IOS Press. 2022, pp. 51–65.

[71] Delaram Golpayegani, Harshvardhan J Pandit, and Dave Lewis. "To Be High-Risk, or Not To Be–Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 2023, pp. 905–915.

[72] Delaram Golpayegani, Isabelle Hupont, Cecilia Panigutti, Harshvardhan J. Pandit, Sven Schade, Declan O'Sullivan, and Dave Lewis. "AI Cards: Towards an Applied Framework for Machine-Readable AI and

Risk Documentation Inspired by the EU AI Act". In: *Privacy Technologies and Policy*. Ed. by Meiko Jensen, Cédric Lauradoux, and Kai Rannenberg. Springer Nature Switzerland, 2024, pp. 48–72. ISBN: 978-3-031-68024-3.

[73] Delaram Golpayegani, Beatriz Esteves, Harshvardhan J. Pandit, and Dave Lewis. "AIUP: an ODRL Profile for Expressing AI Use Policies to Support the EU AI Act". In: *Joint Proceedings of Posters, Demos, Workshops, and Tutorials of the 20th International Conference on Semantic Systems co-located with 20th International Conference on Semantic Systems (SEMANTiCS 2024)*. 2024.

[74] Delaram Golpayegani, Harshvardhan J. Pandit, and Dave Lewis. "Comparison and Analysis of 3 Key AI Documents: EU's Proposed AI Act, Assessment List for Trustworthy AI (ALTAI), and ISO/IEC 42001 AI Management System". In: *Artificial Intelligence and Cognitive Science*. Ed. by Luca Longo and Ruairi O'Reilly. Springer Nature Switzerland, 2023, pp. 189–200. ISBN: 978-3-031-26438-2.

[75] Dave Lewis, David Filip, and Harshvardhan J. Pandit. "An Ontology for Standardising Trustworthy AI". In: *Factoring Ethics in Technology, Policy Making, Regulation and AI*. Ed. by Ali G. Hessami and Patricia Shaw. IntechOpen, 2021. Chap. 5. DOI: 10.5772/intechopen.97478.

[76] Martin Ebers, Veronica R. S. Hoch, Frank Rosenkranz, Hannah Ruschemeier, and Björn Steinrötter. "The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)". In: *J* 4.4 (2021), pp. 589–603. ISSN: 2571-8800. DOI: 10.3390/j4040043.

[77] Nathalie A Smuha, Emma Ahmed-Rengers, Adam Harkens, Wenlong Li, James MacLaren, Riccardo Piselli, and Karen Yeung. *How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act*. 2021. DOI: 10.2139/ssrn.3899991.

[78] Juan Pablo Bermúdez, Rune Nyrup, Sebastian Deterding, Laura Moradbakhti, Céline Mougenot, Fangzhou You, and Rafael A. Calvo. "What Is a Subliminal Technique? An Ethical Perspective on AI-Driven Influence". In: *2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*. 2023, pp. 1–10. DOI: 10.1109/ETHICS57328.2023.10155039.

[79] Matija Franklin, Philip Moreira Tomei, and Rebecca Gorman. *Strengthening the EU AI Act: Defining Key Terms on AI Manipulation*. 2023. arXiv: 2308.16364 [cs.AI]. URL: https://arxiv.org/abs/2308.16364.

[80] Daria Bulgakova. "The Prohibited Artificial Intelligence Practice". In: *Theory and Practice of Forensic Science and Criminalistics* 32.3 (2023), pp. 89–112.

[81] Mark Leiser. "Psychological Patterns and Article 5 of the AI Act:AI-Powered Deceptive Design in the System Architecture and the User Interface". In: *Journal of AI Law and Regulation* 1.1 (2024). DOI: 10.21552/aire/2024/1/4.

[82] Sebastian Felix Schwemer, Letizia Tomada, and Tommaso Pasini. "Legal AI Systems in the EU's proposed Artificial Intelligence Act". In: *Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2021), held in conjunction with ICAIL*. 2021.

[83] Gijs van Dijck. "Predicting Recidivism Risk Meets AI Act". en. In: *European Journal on Criminal Policy and Research* 28.3 (2022), pp. 407–423. ISSN: 1572-9869. DOI: 10.1007/s10610-022-09516-8.

[84] Isabelle Hupont and Emilia Gómez. "Documenting use cases in the affective computing domain using Unified Modeling Language". In: *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE. 2022, pp. 1–8.

[85] Jonas Schuett. "Risk Management in the Artificial Intelligence Act". In: *European Journal of Risk Regulation* (2023), pp. 1–19. DOI: 10.1017/err.2023.1.

[86] Simon Tjoa, Peter Kieseberg Marlies Temper, Marlies Temper, Jakob Zanol, Markus Wagner, and Andreas Holzinger. "AIRMan: An Artificial Intelligence (AI) Risk Management System". In: *2022 International Conference on Advanced Enterprise Information System (AEIS)*. 2022, pp. 72–81. DOI: 10.1109/AEIS59450.2022.00017.

[87] Paolo Giudici, Mattia Centurelli, and Stefano Turchetta. "Artificial Intelligence risk measurement". In: *Expert Systems with Applications* 235 (2024). ISSN: 0957-4174. DOI: 10.1016/j.eswa.2023.121220.

[88] Claudio Novelli, Federico Casolari, Antonino Rotolo, Mariarosaria Taddeo, and Luciano Floridi. "Taking AI risks seriously: a new assessment model for the AI Act". In: *AI & Society* (2023). DOI: 10.1007/s00146-023-01723-z.

[89] Jessica Kelly, Shanza Ali Zafar, Lena Heidemann, João-Vitor Zacchi, Delfina Espinoza, and Núria Mata. "Navigating the EU AI Act: A Methodological Approach to Compliance for Safety-critical Products". In: *2024 IEEE Conference on Artificial Intelligence (CAI)*. 2024, pp. 979–984. DOI: 10.1109/CAI59869.2024.00179.

[90] Georg Stettinger, Patrick Weissensteiner, and Siddartha Khastgir. "Trustworthiness Assurance Assessment for High-Risk AI-Based Systems". In: *IEEE Access* 12 (2024), pp. 22718–22745. DOI: 10.1109/ACCESS.2024.3364387.

[91] Claudio Novelli, Guido Governatori, and Antonino Rotolo. "Automating Business Process Compliance for the EU AI Act". In: *Legal Knowledge and Information Systems*. IOS Press, 2023, pp. 125–130. DOI: 10.3233/FAIA230955.

[92] Balint Gyevnar, Nick Ferguson, and Burkhard Schafer. "Bridging the Transparency Gap: What Can Explainable AI Learn from the AI Act?" In: *ECAI 2023*. Vol. 372. Frontiers in Artificial Intelligence and Applications. IOS Press, 2023, pp. 964–971. DOI: 10.3233/FAIA230367.

[93] Isabelle Hupont, Marina Micheli, Blagoj Delipetrev, Emilia Gómez, and Josep Soler Garrido. "Documenting high-risk AI: a European regulatory perspective". In: *Computer* 56.5 (2023), pp. 18–27.

[94] OECD. *OECD Framework for the Classification of AI systems*. 2022. DOI: 10.1787/cb6d9eca-en.

[95] Mia Hoffmann and Heather Frase. *Adding Structure to AI Harm: An Introduction to CSET's AI Harm Framework*. Tech. rep. Center for Security and Emerging Technology (CSET), 2023. DOI: 10.51593/20230022.

[96] Sean McGregor. "Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.17 (2021), pp. 15458–15463. DOI: 10.1609/aaai.v35i17.17817.

[97] Mia Hoffmann, Mina Narayanan, Ankushi Mitra, Yu-Jie Liao, and Heather Frase. *CSET AI Harm Taxonomy for AIID and Annotation Guide*. Tech. rep. Center for Security and Emerging Technology (CSET), 2023. URL: https://github.com/georgetown-cset/CSET-AIID-harm-taxonomy.

[98] Nikiforos Pittaras and Sean McGregor. "A Taxonomic System for Failure Cause Analysis of Open Source AI Incidents". In: *Proceedings of the Workshop on Artificial Intelligence Safety 2023 (SafeAI 2023)*. 2023.

[99] Bruce G. Buchanan and Jonathan Glick. "AI Topics". In: *AI Magazine* 23.1 (2002). DOI: 10.1609/aimag.v23i1.1611.

[100] Joshua Eckroth, Liang Dong, Reid G. Smith, and Bruce G. Buchanan. "NewsFinder: Automating an AI news service". In: *AI Magazine* 33.2 (2012), pp. 43–43.

[101] The AVID community. *AI Vulnerability Database (AVID)*. URL: https://avidml.org/ (visited on 09/18/2024).

[102] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. "Taxonomy of Risks posed by Language Models". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '22. Seoul, Republic of Korea: Association for Computing Machinery, 2022, pp. 214–229. ISBN: 9781450393522. DOI: 10.1145/3531146.3533088.

[103] Hiroshi Tanaka, Masaru Ide, Jun Yajima, Sachiko Onodera, Kazuki Munakata, and Nobukazu Yoshioka. "Taxonomy of Generative AI Applications for Risk Assessment". In: *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI*. CAIN '24. Lisbon, Portugal: Association for Computing Machinery, 2024. DOI: 10.1145/3644815.3644977.

[104] Hao-Ping (Hank) Lee, Yu-Ju Yang, Thomas Serban Von Davier, Jodi Forlizzi, and Sauvik Das. "Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. Honolulu, HI, USA: Association for Computing Machinery, 2024. DOI: 10.1145/3613904.3642116.

[105] Norberto Nuno Gomes de Andrade and Verena Kontschieder. *AI Impact Assessment: A Policy Prototyping Experiment*. 2021. DOI: 10.2139/ssrn.3772500.

[106] Apostol Vassilev, Alina Oprea, Alie Fordyce, and Hyrum Anderson. *Adversarial Machine Learning A Taxonomy and Terminology of Attacks and Mitigations*. Tech. rep. National Institute of Standards and Technology, 2024. DOI: 10.6028/NIST.AI.100-2e2023.

[107] Reva Schwartz, Apostol Vassilev, Kristen K. Greene, Lori Perine, Andrew Burt, and Patrick Hall. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. Tech. rep. National Institute of Standards and Technology, 2022. DOI: 10.6028/NIST.SP.1270.

[108] André Steimers and Moritz Schneider. "Sources of risk of AI systems". In: *International Journal of Environmental Research and Public Health* 19.6 (2022).

[109] Drew Roselli, Jeanna Matthews, and Nisha Talagala. "Managing Bias in AI". In: *Companion Proceedings of The 2019 World Wide Web Conference*. WWW '19. San Francisco, USA: Association for Computing Machinery, 2019, pp. 539–544. ISBN: 9781450366755. DOI: 10.1145/3308560.3317590.

[110] Karen Leticia Vázquez-Flores, Elena Montiel-Ponsoda, and María Poveda-Villalón. "EA-Ontology: Ethical Assessment Ontology". In: *EKAW'22: Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management*. 2022.

[111] Ajaya Adhikari, Edwin Wenink, Jasper Van Der Waa, Cornelis Bouter, Ioannis Tolios, and Stephan Raaijmakers. "Towards FAIR Explainable AI: a standardized ontology for mapping XAI solutions to use cases, explanations, and AI systems". In: *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*. 2022, pp. 562–568. DOI: 10.1145/3529190.3535693.

[112] Patricia Inoue Nakagawa, Luís Ferreira Pires, João Luiz Rebelo Moreira, Luiz Olavo Bonino da Silva Santos, and Faiza Bukhsh. "Semantic Description of Explainable Machine Learning Workflows for Improving Trust". In: *Applied Sciences* 11.22 (2021). ISSN: 2076-3417. DOI: 10.3390/app112210804.

[113] Izaskun Fernandez, Cristina Aceta, Eduardo Gilabert, and Iker Esnaola-Gonzalez. "FIDES: An ontology-based approach for making machine learning systems accountable". In: *Journal of Web Semantics* 79 (2023). ISSN: 1570-8268. DOI: 10.1016/j.websem.2023.100808.

[114] Mayra Russo and Maria-Esther Vidal. *Leveraging Ontologies to Document Bias in Data*. 2024. arXiv: 2407.00509 [cs.AI]. URL: https://arxiv.org/abs/2407.00509.

[115] Marcin P. Joachimiak, Mark A. Miller, J. Harry Caufield, Ryan Ly, Nomi L. Harris, Andrew Tritt, Christopher J. Mungall, and Kristofer E. Bouchard. *The Artificial Intelligence Ontology: LLM-assisted construction of AI concept hierarchies*. 2024. arXiv: 2404.03044 [cs.LG]. URL: https://arxiv.org/abs/2404.03044.

[116] Julio Hernandez, Delaram Golpayegani, and Dave Lewis. *An Open Knowledge Graph-Based Approach for Mapping Concepts and Requirements between the EU AI Act and International Standards*. 2024. arXiv: 2408.11925 [cs.AI]. URL: https://arxiv.org/abs/2408.11925.

[117] Andrew Harrison, Dayana Spagnuelo, and Ilaria Tiddi. *An Ontology for Ethical AI Principles*. Preprint. 2021. URL: https://www.semantic-web-journal.net/content/ontology-ethical-ai-principles.

[118] Jhon Masso, Félix García, César Pardo, Francisco J. Pino, and Mario Piattini. "A Common Terminology for Software Risk Management". In: *ACM Transactions on Software Engineering and Methodology* 31.4 (2022). DOI: 10.1145/3498539.

[119] David Haynes. "Understanding Personal Online Risk To Individuals Via Ontology Development". In: *Knowledge Organization at the Interface: Proceedings of the Sixteenth International ISKO Conference, 2020, Aalborg, Denmark*. Ergon Verlag, 2020, pp. 171–180.

[120] Lucy McKenna, Junli Liang, Natalia Duda, Nick McDonald, and Rob Brennan. "ARK-Virus: An ARK Platform Extension for Mindful Risk Governance of Personal Protective Equipment Use in Healthcare". In: *Companion Proceedings of the Web Conference 2021*. WWW '21. Ljubljana, Slovenia: Association for Computing Machinery, 2021. ISBN: 9781450383134. DOI: 10.1145/3442442.3458609.

[121] Tiago Prince Sales, Fernanda Baião, Giancarlo Guizzardi, João Paulo A. Almeida, Nicola Guarino, and John Mylopoulos. "The Common Ontology of Value and Risk". In: *Conceptual Modeling*. Ed. by Juan C. Trujillo, Karen C. Davis, Xiaoyong Du, Zhanhuai Li, Tok Wang Ling, Guoliang Li, and Mong Li Lee. Springer International Publishing, 2018, pp. 121–135. ISBN: 978-3-030-00847-5.

[122] Vivek Agrawal. "Towards the Ontology of ISO/IEC 27005:2011 Risk Management Standard". In: *Proceedings of the Tenth International Symposium on Human Aspects of Information Security Assurance (HAISA 2016)*. 2016, pp. 101–111.

[123] Hilmy Hanif, Jorge Constantino, Marie-Therese Sekwenz, Michel van Eeten, Jolien Ubacht, Ben Wagner, and Yury Zhauniarovich. "Tough Decisions? Supporting System Classification According to the AI Act". In: *Legal Knowledge and Information Systems*. IOS Press, 2023, pp. 353–358. DOI: 10.3233/FAIA230987.

[124] Hilmy Hanif, Jorge Constantino, Marie-Therese Sekwenz, Michel van Eeten, Jolien Ubacht, Ben Wagner, and Yury Zhauniarovich. "Navigating the EU AI Act Maze using a Decision-Tree Approach". In: *ACM Journal on Responsible Computing* (2024). DOI: 10.1145/3677174.

[125] Renato Iannella and Serena Villata. *ODRL Information Model 2.2*. W3C Recommendation. 2018. URL: https://www.w3.org/TR/odrl-model/.

[126] Beatriz Esteves, Harshvardhan J Pandit, and Víctor Rodríguez-Doncel. "ODRL profile for expressing consent through granular access control policies in solid". In: *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE. 2021, pp. 298–306.

[127] Sushant Agarwal, Simon Steyskal, Franjo Antunovic, and Sabrina Kirrane. "Legislative Compliance Assessment: Framework, Model and GDPR Instantiation". In: *Privacy Technologies and Policy*. Springer International Publishing, 2018, pp. 131–149. ISBN: 978-3-030-02547-2.

[128] Marina De Vos, Sabrina Kirrane, Julian Padget, and Ken Satoh. "ODRL Policy Modelling and Compliance Checking". In: *Rules and Reasoning*. Ed. by Paul Fodor, Marco Montali, Diego Calvanese, and Dumitru Roman. SpringerInternational Publishing, 2019, pp. 36–51.

[129] Soheil Roshankish and Nicoletta Fornara. "Exploration of Norms and Policies in Digital Fashion Domain Using Semantic Web Technologies". In: *Design, User Experience, and Usability: Design for Contemporary Technological Environments*. Ed. by Marcelo M. Soares, Elizabeth Rosenzweig, and Aaron Marcus. Springer International Publishing, 2021, pp. 384–395. ISBN: 978-3-030-78227-6.

[130] Harshvardhan J Pandit and Beatriz Esteves. "Enhancing Data Use Ontology (DUO) for health-data sharing by extending it with ODRL and DPV". In: *Semantic Web Journal* (2024). DOI: 10.3233/SW-243583.

[131] Ensar Hadziselimovic, Kaniz Fatema, Harshvardhan J Pandit, and Dave Lewis. "Linked Data Contracts to Support Data Protection and Data Ethics in the Sharing of Scientific Data". In: *SemSci@ ISWC*. 2017, pp. 55–62.

[132] Tobias Dam, Andreas Krimbacher, and Sebastian Neumaier. *Policy Patterns for Usage Control in Data Spaces*. 2023. arXiv: 2309.11289 [cs.CR]. URL: https://arxiv.org/abs/2309.11289.

[133] Anthony Cintron Roman, Jennifer Wortman Vaughan, Valerie See, Steph Ballard, Jehu Torres, Caleb Robinson, and Juan M. Lavista Ferres. *Open Datasheets: Machine-readable Documentation for Open Datasets and Responsible AI Assessments*. 2024. arXiv: 2312.06153 [cs.LG]. URL: https://arxiv.org/abs/2312.06153.

[134] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. "Datasheets for datasets". In: *Communications of the ACM* 64.12 (2021), pp. 86–92.

[135] Muhammad Tuan Amith, Licong Cui, Degui Zhi, Kirk Roberts, Xiaoqian Jiang, Fang Li, Evan Yu, and Cui Tao. "Toward a standard formal semantic representation of the model card report". In: *BMC bioinformatics* 23.6 (2022), pp. 1471–2105. DOI: 10.1186/s12859-022-04797-6.

[136] Andy Donald, Apostolos Galanopoulos, Edward Curry, Emir Muñoz, Ihsan Ullah, M. A. Waskow, Maciej Dabrowski, and Manan Kalra. "Towards a Semantic Approach for Linked Dataspace, Model and Data Cards". In: *Companion Proceedings of the ACM Web Conference 2023*. WWW '23 Companion. Austin, TX, USA: Association for Computing Machinery, 2023, pp. 1468–1473. ISBN: 9781450394192. DOI: 10.1145/3543873.3587659.

[137] Jan Philip Wahle, Terry Ruas, Saif M. Mohammad, Norman Meuschke, and Bela Gipp. "AI Usage Cards: Responsibly Reporting AI-Generated Content". In: *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. 2023. DOI: 10.1109/JCDL57899.2023.00060.

[138] Harshvardhan J. Pandit, Axel Polleres, Bert Bos, Rob Brennan, Bud Bruegger, Fajar J. Ekaputra, Javier D. Fernández, Roghaiyeh Gachpaz Hamed, Elmar Kiesling, Mark Lizar, Eva Schlehahn, Simon Steyskal, and Rigo Wenning. "Creating a Vocabulary for Data Privacy". In: *On the Move to Meaningful Internet Systems: OTM 2019 Conferences*. Ed. by Hervé Panetto, Christophe Debruyne, Martin Hepp, Dave Lewis, Claudio Agostino Ardagna, and Robert Meersman. Springer International Publishing, 2019, pp. 714–730.

[139] Harshvardhan J. Pandit. "A Semantic Specification for Data Protection Impact Assessments (DPIA)". In: *Towards a Knowledge-Aware AI: SEMANTiCS 2022—Proceedings of the 18th International Conference on Semantic Systems, 13-15 September 2022, Vienna, Austria*. Vol. 55. Studies on the Semantic Web. IOS Press, 2022, pp. 36–50. DOI: 10.3233/SSW220007.

[140] Harshvardhan J Pandit, Paul Ryan, Georg Philip Krog, Martin Crane, and Rob Brennan. "Towards a Semantic Specification for GDPR Data Breach Reporting". In: *Legal Knowledge and Information Systems*. Vol. 379. Frontiers in Artificial Intelligence and Applications. IOS Press, 2023, pp. 131–136. DOI: 10.3233/FAIA230956.

[141] Paul Ryan, Rob Brennan, and Harshvardhan J Pandit. "DPCat: Specification for an Interoperable and Machine-Readable Data Processing Catalogue Based on GDPR". In: *Information* 13.5 (2022).

[142] Abhishek Kumar, Benjamin Finley, Tristan Braud, Sasu Tarkoma, and Pan Hui. "Sketching an AI Marketplace: Tech, Economic, and Regulatory Aspects". In: *IEEE Access* 9 (2021), pp. 13761–13774. DOI: 10.1109/ACCESS.2021.3050929.

[143] Mubashara Akhtar, Omar Benjelloun, Costanza Conforti, Pieter Gijsbers, Joan Giner-Miguelez, Nitisha Jain, Michael Kuchnik, Quentin Lhoest, Pierre Marcenac, Manil Maskey, Peter Mattson, Luis Oala, Pierre Ruyssen, Rajat Shinde, Elena Simperl, Goeffry Thomas, Slava Tykhonov, Joaquin Vanschoren, Jos van der Velde, Steffen Vogler, and Carole-Jean Wu. "Croissant: A Metadata Format for ML-Ready Datasets". In: DEEM '24. Santiago, AA, Chile: Association for Computing Machinery, 2024, pp. 1–6. DOI: 10.1145/3650203.3663326.

[144] Bert Van Nuffelen. *DCAT-AP 3.0*. 2024. URL: https://semiceu.github.io/DCAT-AP/releases/3.0.0/.

[145] Fabian Kirstein, Benjamin Dittwald, Simon Dutkowski, Yury Glikman, Sonja Schimmler, and Manfred Hauswirth. "Linked Data in the European Data Portal: A Comprehensive Platform for Applying DCAT-AP". In: *Electronic Government*. Springer International Publishing, 2019, pp. 192–204. ISBN: 978-3-030-27325-5.

[146] Arthur Schiltz and Emidio Stani. *MLDCAT-AP*. 2024. URL: https://semiceu.github.io/MLDCAT-AP/releases/2.0.0/ (visited on 09/13/2024).

[147] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. "Model cards for model reporting". In: *Proceedings of the conference on fairness, accountability, and transparency*. 2019, pp. 220–229.

[148] Giada Pistilli, Carlos Muñoz Ferrandis, Yacine Jernite, and Margaret Mitchell. "Stronger Together: on the Articulation of Ethical Charters, Legal Tools, and Technical Documentation in ML". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '23. Chicago, IL, USA: Association for Computing Machinery, 2023, pp. 343–354. DOI: 10.1145/3593013.3594002.

[149] Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. "FactSheets: Increasing trust in AI services through supplier's declarations of conformity". In: *IBM Journal of Research and Development* 63.4/5 (2019). DOI: 10.1147/JRD.2019.2942288.

[150] Sarah Holland, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. "The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards". In: *Data Protection and Privacy* 12 (2020).

[151] Ben Hutchinson, Andrew Smart, Alex Hanna, Emily Denton, Christina Greer, Oddur Kjartansson, Parker Barnes, and Margaret Mitchell. "Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. Virtual Event, Canada: Association for Computing Machinery, 2021, pp. 560–575. ISBN: 9781450383097. DOI: 10.1145/3442188.3445918.

[152] Marina Micheli, Isabelle Hupont, Blagoj Delipetrev, and Josep Soler-Garrido. "The landscape of data and AI documentation approaches in the European policy context". In: *Ethics and Information Technology* 25.4 (2023).

[153] Florian Königstorfer and Stefan Thalmann. "Software documentation is not enough! Requirements for the documentation of AI". In: *Digital Policy, Regulation and Governance* 23.5 (2021), pp. 475–488.

[154] Stefan Arnold, Dilara Yesilbas, Rene Gröbner, Dominik Riedelbauch, Maik Horn, and Sven Weinzierl. *Documentation Practices of Artificial Intelligence*. 2024. arXiv: 2406.18620 [cs.DL]. URL: https://arxiv.org/abs/2406.18620.

[155] Edyta Bogucka, Marios Constantinides, Sanja Šćepanović, and Daniele Quercia. *Co-designing an AI Impact Assessment Report Template with AI Practitioners and AI Compliance Experts*. 2024. arXiv: 2407.17374 [cs.HC]. URL: https://arxiv.org/abs/2407.17374.

[156] Leon Derczynski, Hannah Rose Kirk, Vidhisha Balachandran, Sachin Kumar, Yulia Tsvetkov, M. R. Leiser, and Saif Mohammad. *Assessing Language Model Deployment with Risk Cards*. 2023. arXiv: 2303.18190 [cs.CL]. URL: https://arxiv.org/abs/2303.18190.

[157] Eli Sherman and Ian Eisenberg. "AI Risk Profiles: A Standards Proposal for Pre-deployment AI Risk Disclosures". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. 21. 2024, pp. 23047–23052. DOI: 10.1609/aaai.v38i21.30348.

[158] Tycho de Graaf and Gitta Veldt. "The AI Act and Its Impact on Product Safety, Contracts and Liability". In: *European Review of Private Law* 30 (5 2022). DOI: 10.54648/erpl2022038.

[159] Ankit Gupta. *Smart Toys Market Research Report Information By Type (Robots, Interactive Games, Educational Robots), By Technology (Wi-Fi, Bluetooth, RFID or NFC), By Distribution Channel (Online/Ecommerce Stores, Specialty Stores, Toy Shops), By End-user (Toddlers, Pre-schoolers, School-going, Stripling), And By Region (North America, Europe, Asia-Pacific, And Rest Of The World) – Market Forecast Till 2030*. Tech. rep. Market Research Future, 2024. URL: https://www.marketresearchfuture.com/reports/smart-toys-market-10813.

[160] Andrew McStay and Gilad Rosner. "Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies". In: *Big Data & Society* 8.1 (2021). DOI: 10.1177/2053951721994877.

[161] Daniel Nagel, Rafael Capurro, Johannes Britz, Thomas Hausmanninger, Michael Nagenborg, Makoto Nakada, and Felix Weil. "Ethical Issues of Networked Toys". In: *International Review of Information Ethics* 27 (2018).

[162] Shih-Ting Chu, Gwo-Jen Hwang, and Yun-Fang Tu. "Artificial intelligence-based robots in education: A systematic review of selected SSCI publications". In: *Computers and Education: Artificial Intelligence* 3 (2022). ISSN: 2666-920X. DOI: 10.1016/j.caeai.2022.100091.

[163] *Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys*. 2009. URL: http://data.europa.eu/eli/dir/2009/48/oj.

[164] Daniel Zhang, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault. *The AI Index 2021 Annual Report*. 2021. arXiv: 2103.06312 [cs.AI]. URL: https://arxiv.org/abs/2103.06312.

[165] Emilia Niemiec. "Will the EU Medical Device Regulation help to improve the safety and performance of medical AI devices?" In: *Digital Health* 8 (2022). DOI: 10.1177/20552076221089079.

[166] *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC*. 2017. URL: http://data.europa.eu/eli/reg/2017/745/oj.

[167] Elisabetta Biasin and Erik Kamenjašević. "Cybersecurity of medical devices: new challenges arising from the AI Act and NIS 2 Directive proposals". en. In: *International Cybersecurity Law Review* 3.1 (2022), pp. 163–180. DOI: 10.1365/s43439-022-00054-x.

[168] Matija Franklin, Hal Ashton, Rebecca Gorman, and Stuart Armstrong. "Missing Mechanisms of Manipulation in the EU AI Act". In: *The International FLAIRS Conference Proceedings* 35 (2022). DOI: 10.32473/flairs.v35i.130723.

[169] ISO/IEC. *ISO/IEC Directives, Part 1 Procedures for the technical work — Consolidated ISO Supplement — Procedures specific to ISO*. International Organization for Standardization (ISO). URL: https://www.iso.org/sites/directives/current/consolidated/index.html.

[170] *Commission communication in the framework of the implementation of Regulation (EC) No 765/ 2008 of the European Parliament and of the Council, Decision No 768/2008/EC of the European Parliament and of the Council, Regulation (EC) No 1221/2009 of the European Parliament and of the Council (Publication of titles and references of harmonised standards under Union harmonisation legisla-*

tion) (2018/C 209/02). Publications Office of the European Union, 2018. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017XC0310(05).

[171] Vidosav Majstorović and Valentina Marinković. "The Development of Business Standardization and Integrated Management Systems". In: *Journal of Medical Biochemistry* 30.4 (2011), pp. 334–345. DOI: 10.2478/v10011-011-0015-5.

[172] Natalya F Noy and Deborah L McGuinness. *Ontology development 101: A guide to creating your first ontology*. 2001.

[173] Mark A Musen and Protégé Team. "The Protégé Project: A Look Back and a Look Forward." In: *AI matters* 1.4 (2015), pp. 4–12.

[174] Daniel Garijo, Oscar Corcho, and Maria Poveda-Villalón. "FOOPS!: An Ontology Pitfall Scanner for the FAIR Principles". In: CEUR Workshop Proceedings 2980 (2021). URL: http://ceur-ws.org/Vol-2980/paper321.pdf.

[175] Daniel Garijo. "WIDOCO: A Wizard for Documenting Ontologies". In: *The Semantic Web – ISWC 2017*. Springer International Publishing, 2017, pp. 94–102. ISBN: 978-3-319-68204-4.

[176] ISO/IEC. *ISO/IEC TR 24028:2020 - Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*. International Organization for Standardization (ISO). URL: https://www.iso.org/standard/77608.html.

[177] DCMI Usage Board. *DCMI Metadata Terms*. DCMI Recommendation. 2020. URL: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/.

[178] Riccardo Albertoni and Antoine Isaac. "Introducing the Data Quality Vocabulary (DQV)". In: *Semantic Web* 12.1 (2021), pp. 81–97. DOI: 10.3233/SW-200382.

[179] Samoili S, Lopez Cobo M, Gomez Gutierrez E, De Prato G, Martinez-Plumed F, and Delipetrev B. *AI Watch. Defining Artificial Intelligence*. Publications Office of the European Union, 2020. ISBN: 978-92-76-17045-7 (online). DOI: 10.2760/382730.

[180] Samoili S, Lopez Cobo M, Delipetrev B, Martinez-Plumed F, Gomez Gutierrez E, and De Prato G. *AI Watch. Defining Artificial Intelligence 2.0*. Publications Office of the European Union, 2021. ISBN: 978-92-76-42648-6 (online). DOI: 10.2760/019901.

[181] Steven Melendez. "Uber driver troubles raise concerns about transgender face recognition". In: *Fast Company* (Sept. 8, 2018). URL: https://www.fastcompany.com/90216258/uber-face-recognition-tool-has-locked-out-some-transgender-drivers (visited on 07/30/2024).

[182] Chris Vallance. "Legal action over alleged Uber facial verification bias". In: *BBC* (Oct. 8, 2021). URL: https://www.bbc.com/news/technology-58831373 (visited on 07/30/2024).

[183] *External audit of the VioGén System*. 2022. URL: https://eticasfoundation.org/?audit-spotlight=the-adversarial-audit-of-viogen (visited on 07/30/2024).

[184] Harshvardhan J. Pandit, Declan O'Sullivan, and Dave Lewis. "Test-Driven Approach Towards GDPR Compliance". In: *Semantic Systems. The Power of AI and Knowledge Graphs*. Ed. by Maribel Acosta, Philippe Cudré-Mauroux, Maria Maleshkova, Tassilo Pellegrini, Harald Sack, and York Sure-Vetter. Springer International Publishing, 2019, pp. 19–33.

[185] Frances Gillis-Webber and C. Maria Keet. *A Review of Multilingualism in and for Ontologies*. 2022. arXiv: 2210.02807 [cs.AI]. URL: https://arxiv.org/abs/2210.02807.

[186] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. W3C Member Submission. 2004. URL: https://www.w3.org/submissions/2004/SUBM-SWRL-20040521/.

[187] William Van Woensel, Dörthe Arndt, Pierre-Antoine Champin, Dominik Tomaszuk, and Gregg Kellogg. *Notation3 Language*. W3C Community Group Draft Report. 2024. URL: https://w3c.github.io/N3/spec/.

[188] Eric Prud'hommeaux, Iovka Boneva, Jose Emilio Labra Gayo, and Gregg Kellogg. *Shape Expressions Language 2.1*. W3C Final Community Group Report. 2019. URL: http://shex.io/shex-semantics/.

[189] Holger Knublauch. *SHACL and OWL Compared*. 2017. URL: https://spinrdf.org/shacl-and-owl.html (visited on 09/16/2024).

[190] Serge Chávez-Feria, Raúl García-Castro, and María Poveda-Villalón. "Chowlk: from UML-Based Ontology Conceptualizations to OWL". In: *The Semantic Web*. Ed. by Paul Groth, Maria-Esther Vidal, Fabian Suchanek, Pedro Szekley, Pavan Kapanipathi, Catia Pesquita, Hala

Skaf-Molli, and Minna Tamper. Springer International Publishing, 2022, pp. 338–352.

[191] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Boris Villazón-Terrazas. "How to Write and Use the Ontology Requirements Specification Document". In: *On the Move to Meaningful Internet Systems: OTM 2009*. Ed. by Robert Meersman, Tharam Dillon, and Pilar Herrero. Springer Berlin Heidelberg, 2009, pp. 966–982. ISBN: 978-3-642-05151-7.

[192] Michael Steidl. *ODRL V2.2 Profile Best Practices*. W3C Community Group Final Report. URL: https://w3c.github.io/odrl/profile-bp/.

[193] Jeroen Naves and Pels Rijcken. *Proposal for standard contractual clauses for the procurement of Artificial Intelligence (AI) by public organisations*. 2023. URL: https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/eu-model-contractual-ai-clauses-pilot-procurements-ai (visited on 09/11/2024).

[194] European Commission, Joint Research Centre, Marina Manzoni, Rony Medaglia, Luca Tangi, Colin Van Noordt, Lorenzino Vaccari, and Dietmar Gattwinkel. *AI Watch, road to the adoption of artificial intelligence by the public sector – A handbook for policymakers, public administrations and relevant stakeholders*. Publications Office of the European Union, 2022. DOI: 10.2760/288757.

[195] European Commission and Joint Research Centre (JRC). *Selected AI cases in the public sector (JRC129301)*. Dataset. 2021. URL: http://data.europa.eu/89h/7342ea15-fd4f-4184-9603-98bd87d8239a (visited on 09/08/2024).

[196] *Regulation (EU) 2024/903 of the European Parliament and of the Council of 13 March 2024 laying down measures for a high level of public sector interoperability across the Union (Interoperable Europe Act)*. 2024. URL: http://data.europa.eu/eli/reg/2024/903/oj.

[197] ISO/IEC. *ISO/IEC 25059:2023 Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems*. International Organization for Standardization (ISO). URL: https://www.iso.org/standard/80655.html.

[198] UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. 2022. URL: https://unesdoc.unesco.org/ark:/48223/pf0000381137.locale=en.

[199] Amy K Heger, Liz B Marquis, Mihaela Vorvoreanu, Hanna Wallach, and Jennifer Wortman Vaughan. "Understanding Machine Learning Practitioners' Data Documentation Perceptions, Needs, Challenges, and Desiderata". In: *Proceedings of the ACM on Human-Computer Interaction* 6.CSCW2 (2022), pp. 1–29.

[200] European Commission and Directorate-General for Financial Stability, Financial Services and Capital Markets Union. *MRER proof of concept – Assessing the feasibility of machine-readable and executable reporting for EMIR*. Publications Office of the European Union, 2022. DOI: 10.2874/036007.

[201] European Commission and Directorate-General for Communications Networks, Content and Technology. *Ethics Guidelines for Trustworthy AI*. Publications Office of the European Union, 2019. DOI: 10.2759/177365.

[202] Asunción Gómez-Pérez. "Ontology Evaluation". In: *Handbook on Ontologies*. Ed. by Steffen Staab and Rudi Studer. Springer Berlin Heidelberg, 2004, pp. 251–273. ISBN: 978-3-540-24750-0. DOI: 10.1007/978-3-540-24750-0_13.

[203] John Brooke. "SUS-A "quick and dirty" usability scale". In: *Usability evaluation in industry* (1996), pp. 189–194.

[204] James R. Lewis. "The System Usability Scale: Past, Present, and Future". In: *International Journal of Human–Computer Interaction* 34.7 (2018), pp. 577–590. DOI: 10.1080/10447318.2018.1455307.

[205] Davide LIGA, Réka MARKOVICH, and Daniele Amitrano. "PaTrOnto, an ontology for patents and trademarks". In: *Proceedings of the Seventeenth International Workshop on Juris-Informatics 2023 (JURISIN 2023)*. Tokyo, Japan, 2023.

[206] Davide Liga, Réka MARKOVICH, and Alessia Fidelangeli. "Ontovat, an ontology for knowledge extraction in vat-related judgments". In: *Proceedings of the Seventeenth International Workshop on Juris-Informatics 2023 (JURISIN 2023)*. Tokyo, Japan, 2023.

[207] Davide Liga, Alessia Fidelangeli, and Réka Markovich. "Using Ontological Knowledge and Large Language Model Vector Similarities to Extract Relevant Concepts in VAT-Related Legal Judgments". In: *New Frontiers in Artificial Intelligence*. Ed. by Mayumi Bono, Yasufumi Takama, Ken Satoh, Le-Minh Nguyen, and Setsuya Kurahashi. Springer Nature Switzerland, 2024, pp. 115–131. ISBN: 978-3-031-60511-6.

[208] Fabian Hüger, Alexander Poth, Andreas Wittmann, and Roland Walgenbach. "An Approach to the Instantiation of the EU AI Act: A Level of Done Derivation and a Case Study from the Automotive Domain". In: *Systems, Software and Services Process Improvement*. Ed. by Murat Yilmaz, Paul Clarke, Andreas Riel, and Richard Messnarz. Springer Nature Switzerland, 2023, pp. 111–123. ISBN: 978-3-031-42307-9.

[209] Clotilde Braye, Jérémy Clech, Arnaud Gotlieb, Nadjib Lazaar, and Patrick Malléa. "Towards Trustworthy-AI-by-Design Methodology for Intelligent Radiology Systems". In: *Santé et IA séminaire' at Plate-Forme Intelligence Artificielle (PFIA 23)*. 2023.

[210] Mirthe Dankloff, Vanja Skoric, Giovanni Sileno, Sennay Ghebreab, Jacco van Ossenbruggen, and Emma Beauxis-Aussalet. "Analysing and organising human communications for AI fairness assessment". In: *AI & SOCIETY* (2024). ISSN: 1435-5655. DOI: 10.1007/s00146-024-01974-4.

[211] Richard May, Jacob Krüger, and Thomas Leich. "SoK: How Artificial-Intelligence Incidents Can Jeopardize Safety and Security". In: *Proceedings of the 19th International Conference on Availability, Reliability and Security*. ARES '24. Vienna, Austria: Association for Computing Machinery, 2024. DOI: 10.1145/3664476.3664510.

[212] Sergio Morales, Robert Clarisó, and Jordi Cabot. "A DSL for Testing LLMs for Fairness and Bias". In: *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems*. MODELS '24. Linz, Austria: Association for Computing Machinery, 2024, pp. 203–213.

[213] Tiphaine Viard, Melanie Gornet, and Winston Maxwell. "Reading the drafts of the AI Act with a technical lens". In: *NeurIPS 2023 Workshop on Regulatable ML*. 2023. URL: https://openreview.net/forum?id=pnvRy1VzJZ.

[214] Edyta Paulina Bogucka, Marios Constantinides, Julia De Miguel Velazquez, Sanja Scepanovic, Daniele Quercia, and Andrés Gvirtz. "The Atlas of AI Incidents in Mobile Computing: Visualizing the Risks and Benefits of AI Gone Mobile". In: *Adjunct Proceedings of the 26th International Conference on Mobile Human-Computer Interaction*. MobileHCI '24 Adjunct. Melbourne, VIC, Australia: Association for Computing Machinery, 2024.

[215] Viviane Herdel, Sanja Šćepanović, Edyta Bogucka, and Daniele Quercia. *ExploreGen: Large Language Models for Envisioning the Uses and Risks of AI Technologies*. 2024. arXiv: 2407.12454 [cs.HC]. URL: https://arxiv.org/abs/2407.12454.

[216] Marios Constantinides, Edyta Paulina Bogucka, Sanja Scepanovic, and Daniele Quercia. "Good Intentions, Risky Inventions: A Method for Assessing the Risks and Benefits of AI in Mobile and Wearable Uses". In: *Proceedings of the ACM on Human-Computer Interaction* 8.MHCI (2024). DOI: 10.1145/3676507.

[217] Timothée Schmude, Laura Koesten, Torsten Möller, and Sebastian Tschiatschek. *Information That Matters: Exploring Information Needs of People Affected by Algorithmic Decisions*. 2024. arXiv: 2401.13324 [cs.HC]. URL: https://arxiv.org/abs/2401.13324.

[218] Maximilian Grafenstein. *Reconciling Conflicting Interests in Data through Data Governance. An Analytical Framework (and a Brief Discussion of the Data Governance Act Draft, the Data Act Draft, the AI Regulation Draft, as well as the GDPR)*. 2022. DOI: 10.2139/ssrn.4104502.

[219] Artur Bogucki, Alex Engler, Clément Perarnaud, and Andrea Renda. *The AI Act and Emerging EU Digital Acquis Overlaps, gaps and inconsistencies*. CEPS (Centre for European Policy Studies), 2022.

[220] Matthias Artzt and Tran Viet Dung. "Artificial Intelligence and Data Protection: How to Reconcile Both Areas from the European Law Perspective". In: *Vietnamese Journal of Legal Sciences* 7.2 (2022), pp. 39–58. DOI: 10.2478/vjls-2022-0007.

[221] EU-US Trade and Technology Council. *TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management*. en. 2022. URL: https://digital-strategy.ec.europa.eu/en/library/ttc-joint-roadmap-trustworthy-ai-and-risk-management (visited on 09/05/2024).

[222] Nahema Marchal, Rachel Xu, Rasmi Elasmar, Iason Gabriel, Beth Goldberg, and William Isaac. *Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data*. 2024. arXiv: 2406.13843 [cs.AI]. URL: https://arxiv.org/abs/2406.13843.

[223] Yi Zeng, Kevin Klyman, Andy Zhou, Yu Yang, Minzhou Pan, Ruoxi Jia, Dawn Song, Percy Liang, and Bo Li. *AI Risk Categorization Decoded (AIR 2024): From Government Regulations to Corporate Policies*. 2024. arXiv: 2406.17864 [cs.CY]. URL: https://arxiv.org/abs/2406.17864.

[224] Hamed Babaei Giglou, Jennifer D'Souza, and Sören Auer. "LLMs4OL: Large Language Models for Ontology Learning". In: *The Semantic Web – ISWC 2023*. Ed. by Terry R. Payne, Valentina Presutti, Guilin Qi, María Poveda-Villalón, Giorgos Stoilos, Laura Hollink, Zoi Kaoudi, Gong Cheng, and Juanzi Li. Springer Nature Switzerland, 2023, pp. 408–427. ISBN: 978-3-031-47240-4.

[225] Patricia Mateiu and Adrian Groza. "Ontology engineering with Large Language Models". In: *2023 25th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. 2023, pp. 226–229. DOI: 10.1109/SYNASC61333.2023.00038.

[226] Vamsi Krishna Kommineni, Birgitta König-Ries, and Sheeba Samuel. *From human experts to machines: An LLM supported approach to ontology and knowledge graph construction*. 2024. arXiv: 2403.08345 [cs.CL]. URL: https://arxiv.org/abs/2403.08345.

[227] Mohammad Javad Saeedizade and Eva Blomqvist. "Navigating Ontology Development with Large Language Models". In: *The Semantic Web*. Ed. by Albert Meroño Peñuela, Anastasia Dimou, Raphaël Troncy, Olaf Hartig, Maribel Acosta, Mehwish Alam, Heiko Paulheim, and Pasquale Lisena. Springer Nature Switzerland, 2024, pp. 143–161. ISBN: 978-3-031-60626-7.

Appendices

APPENDIX

ONE

PREFIXES AND NAMESPACES

Table A.1: Prefixes and Namespaces

Prefix	Namespace	Description
aicat	https://w3id.org/aicat#	AI Catalogue vocabulary
airo	https://w3id.org/airo#	AI Risk Ontology
aiup	https://w3id.org/aiup#	AI Use Policy profile
dcat	https://www.w3.org/TR/vocab-dcat-3/	Data Catalog Vocabulary - Version 3
dct	http://purl.org/dc/terms/	Dublin Core Metadata Terms
dpv	https://w3id.org/dpv#	Data Privacy Vocabulary
dqv	http://www.w3.org/ns/dqv#	Data Quality Vocabulary
freq	http://purl.org/cld/freq/	Dublin Core Collection Description Frequency Vocabulary
odrl	https://www.w3.org/ns/odrl/2/	Open Digital Rights Language
owl	http://www.w3.org/2002/07/owl#	The OWL 2 Schema vocabulary (OWL 2)
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	The RDF Concepts vocabulary (RDF)
rdfs	http://www.w3.org/2000/01/rdf-schema#	The RDF Schema vocabulary (RDFS)
tech	https://w3id.org/dpv/tech#	DPV's technology extension
vair	https://w3id.org/vair#	Vocabulary of AI Risks
xsd	http://www.w3.org/2001/XMLSchema#	W3C XML Schema Definition Language (XSD)

APPENDIX

TWO

ANALYSIS OF PROHIBITED AI PRACTICES

This Appendix provides the analysis of Article 5 clauses wherein the conditions that make an AI system prohibited under the AI Act are outlined.

Table B.1: Analysis of prohibited AI practices listed in Article 5, Points (1a) to (1d)

Art. 5 clause	Domain	Purpose	Capability	Data processed	AI subject	Locality of use	Consequence	Impact
(1a)	Any	Any	Subliminal Capability, Manipulation, Deception	Any	Person Group of Persons	Any	Impaired Decision Making	(Significant) Harm
(1b)	Any	Any	Exploitation Of Vulnerability	Any	Vulnerable Person, Vulnerable Groups Of Persons	Any	Materially Distorting Behaviour, Exploiting Vulnerability	(Significant) Harm
(1c)	Any	Evaluation Of People, Classification Of People	Social Scoring	Social Behaviour Data, Known, Inferred or Predicted Personal Characteristics Known, Inferred or Predicted Personality Characteristics	Natural Person, Group of Persons	Any	Any	Discriminatory Treatment, Detrimental Treatment, Unfavourable Treatment
(1d)	Any	Assessing Risk of Committing a Criminal Offence, Predicting Risk of Committing a Criminal Offence	Profiling, Personality Traits Assessment, Personality Characteristics Assessment	Any	Natural Person	Any	Any	Any

Table B.2: Analysis of prohibited AI practices listed in Article 5, Points (1e) to (1h)

Art. 5 clause	Domain	Purpose	Capability	Data processed	AI subject	Locality of use	Consequence	Impact
(1e)	Any	Creating Facial Recognition Databases, Expanding Facial Recognition Databases	Web Scraping	Facial Images From The Internet Facial Images From CCTV Footage	Natural Person	Any	Any	Any
(1f)	Employment, Education	Any	Emotion Recognition	Any	Natural Person	Workplace, Education Institution	Any	Any
(1g)	Any	Any	Profiling	Special Category Data	Natural Person	Any	Any	Any
(1h)	Law Enforcement	Remote Identification	Real-Time Remote Biometric Identification	Biometric Data	Natural Person	Publicly Accessible Spaces	Any	Any

APPENDIX

THREE

ANALYSIS OF ANNEX III HIGH-RISK AI SYSTEMS

This Appendix illustrates how the 5 concepts, identified in Subsection 3.2.1, are applied to each clause in Annex III of the AI Act, which defines the conditions that make an AI system high-risk. In this, for each clause there is one corresponding 5-concept model, except for Points 4b and 5d (shown in Table C.4 and Table C.5), which are subdivided into two specific conditions. Additionally, the conditions pertaining to the use of polygraphs (Annex III, Points 6b and 7a) can be modelled in two different ways (see Table C.6 and Table C.7).

Table C.1: Analysis of Annex III, Point 1 high-risk AI systems

Clause	The 5 concepts
(1a)	1. Domain: Any 2. Purpose: Any 3. Capability: *Remote Identification* OR *Biometric Identification* 4. Deployer: Any 5. AI Subject: *Natural Person*
(1b)	1. Domain: Any 2. Purpose: Categorisation 3. Capability: *Biometric Categorisation* OR *Sensitive Attribute Inference* 4. Deployer: Any 5. AI subject: *Natural Person*
(1c)	1. Domain: Any 2. Purpose: *Recognising Emotions* OR *Detecting Emotional State* 3. Capability: *Emotion Sensing* OR *Biometrics-Based Emotion Sensing* 4. Deployer: Any 5. AI subject: *Natural Person*

Table C.2: Analysis of Annex III, Point 2 high-risk AI systems

Clause	The 5 concepts
(2)	**1. Domain**: *Critical Infrastructure* **2. Purpose**: *Serving Safety Function In Operation Of Critical Digital Infrastructure* OR *Serving Safety Function In Management Of Critical Digital Infrastructure* OR *Serving Safety Function In Operation Of Road Traffic* OR *Serving Safety Function In Management Of Road Traffic* OR *Serving Safety Function In Operation Of The Supply Of Water* OR *Serving Safety Function In Management Of The Supply Of Water* OR *Serving Safety Function In Operation Of The Supply Of Gas* OR *Serving Safety Function In Management Of The Supply Of Gas* OR *Serving Safety Function In Operation Of The Supply Of Heating* OR *Serving Safety Function In Management Of The Supply Of Heating* OR *Serving Safety Function In Operation Of The Supply Of Electricity* OR *Serving Safety Function In Management Of The Supply Of Electricity* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: Any

Table C.3: Analysis of Annex III, Point 3 high-risk AI systems

Clause	The 5 concepts
(3a)	**1. Domain**: *Education* **2. Purpose**: *Determining Access To Educational Institutions* OR *Determining Access To Vocational Training Institutions* OR *Determining Admission To Educational Institutions* OR *Determining Admission To Vocational Training Institutions* OR *Assigning Persons To Educational Institutions* OR *Assigning Persons To Vocational Training Institutions* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: *Natural Person*
(3b)	**1. Domain**: *Education* **2. Purpose**: *Evaluating Learning Outcomes* **3. Capability**: Any **4. Deployer**: *Educational Institution* OR *Vocational Training Institution* **5. AI subject**: *Natural Person* OR *Course Attendee*
(3c)	**1. Domain**: *Education* **2. Purpose**: *Assessing Level Of Education* OR *Assessing Accessible Level Of Education* **3. Capability**: Any **4. Deployer**: *Educational Institution* OR *Vocational Training Institution* **5. AI subject**: *Natural Person*
(3d)	**1. Domain**: *Education* **2. Purpose**: *Monitoring Prohibited Behaviour During Test* OR *Detecting Prohibited Behaviour During Test* **3. Capability**: Any **4. Deployer**: *Educational Institution* OR *Vocational Training Institution* **5. AI subject**: *Student*

Table C.4: Analysis of Annex III, Point 4 high-risk AI systems

Clause	The 5 concepts
(4a)	**1. Domain**: *Employment* **2. Purpose**: *Recruiting* OR *Job Candidate Selection* OR *Placing Targeted Job Advert* OR *Job Application Analysis* OR *Job Application Screening* OR *Job Application Filtering* OR *Evaluating Job Interview* OR *Evaluating Job Candidates* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: *Natural Person* OR *Job Applicant* OR *Potential Job Applicant* OR *Job Candidate*
(4b)-1	**1. Domain**: *Employment* **2. Purpose**: *Making Decision On Terms Of Work Related Relations* OR *Making Promotion Decision* OR *Making Contract Termination Decision* OR *Monitoring Employee Performance* OR *Evaluating Employee Performance* OR *Monitoring Employee Behaviour* OR *Evaluating Employee Behaviour* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: *Natural Person* OR *Employee*
(4b)-2	**1. Domain**: *Employment* **2. Purpose**: *Allocating Tasks* **3. Capability**: *Behaviour Analysis* OR *Personality Traits Analysis* **4. Deployer**: Any **5. AI subject**: *Natural Person* OR *Employee*

Table C.5: Analysis of Annex III, Point 5 high-risk AI systems

Clause	The 5 concepts
(5a)	**1. Domain**: *Public Service* **2. Purpose**: *Evaluating Eligibility For Public Assistance Services* OR *Granting Public Assistance Services* OR *Reducing Public Assistance Services* OR *Revoking Public Assistance Services* OR *Reclaiming Public Assistance Services* OR *Evaluating Eligibility For Healthcare Services* OR *Granting Healthcare Services* OR *Reducing Healthcare Services* OR *Revoking Healthcare Services* OR *Reclaiming Healthcare Services* **3. Capability**: Any **4. Deployer**: *Public Authority* OR *Public Authority Agent* **5. AI subject**: *Natural Person* OR *Family* OR *Public Services Applicant* OR *Potential Public Services Applicant* OR *Public Services Recipient*
(5b)	**1. Domain**: *Public Service* OR *Private Service* **2. Purpose**: *Assessing Creditworthiness* OR *Determining Credit Score* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: *Natural Person*
(5c)	**1. Domain**: *Public Service* OR *Private Service* **2. Purpose**: *Health Insurance Risk Assessment* OR *Health Insurance Pricing* OR *Life Insurance Risk Assessment* OR *Life Insurance Pricing* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: *Natural Person*
(5d)-1	**1. Domain**: *Public Service* OR *Private Service* **2. Purpose**: *Evaluating Emergency Call* OR *Classifying Emergency Call* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: *Natural Person*
(5d)-2	**1. Domain**: *Public Service* OR *Private Service* **2. Purpose**: *Dispatching Emergency Service* OR *Prioritisation Of Emergency Service* OR *Emergency Triage* **3. Capability**: Any **4. Deployer**: *Emergency Service Provider* OR *Emergency Healthcare Provider* OR *Police* Or *Fire Brigade* OR *Medical Aid Provider* **5. AI subject**: Any

Table C.6: Analysis of Annex III, Point 6 high-risk AI systems

Clause	The 5 concepts
(6a)	**1. Domain**: *Law Enforcement* **2. Purpose**: *Assessing Risk Of Becoming Victim Of Crime* **3. Capability**: Any **4. Deployer**: *Law Enforcement Authority* OR *Law Enforcement Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: *Natural Person*
(6b)	**1. Domain**: *Law Enforcement* [**2. Purpose**: Any **3. Capability**: *Lie Detection*] OR [**2. Purpose**: Detecting Lies **3. Capability**: *Emotion Recognition*] **4. Deployer**: *Law Enforcement Authority* OR *Law Enforcement Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: *Natural Person*
(6c)	**1. Domain**: *Law Enforcement* **2. Purpose**: *Evaluating Reliability Of Evidence In Investigation Of Criminal Offences* OR *Evaluating Reliability Of Evidence In Prosecution Of Criminal Offences* **3. Capability**: Any **4. Deployer**: *Law Enforcement Authority* OR *Law Enforcement Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: Any
(6d)	**1. Domain**: *Law Enforcement* **2. Purpose**: *Assessing Risk Of Offending* OR *Assessing Risk Of Re-offending* OR *Assessing Personality Traits* OR *Assessing Past Criminal Behaviour* **3. Capability**: Any **4. Deployer**: *Law Enforcement Authority* OR *Law Enforcement Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: *Natural Person* or *Group*
(6e)	**1. Domain**: *Law Enforcement* **2. Purpose**: *Detecting Criminal Offences* OR *Investigating Criminal Offences* OR *Prosecuting Criminal Offences* **3. Capability**: *Profiling* **4. Deployer**: *Law Enforcement Authority* OR *Law Enforcement Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: *Natural Person*

Table C.7: Analysis of Annex III, Point 7 high-risk AI systems

Clause	The 5 concepts
(7a)	**1. Domain**: *Migration Management* OR *Asylum Management* OR *Border Control Management* [**2. Purpose**: Any **3. Capability**: *Lie Detection*] OR [**2. Purpose**: Detecting Lies **3. Capability**: *Emotion Recognition*] **4. Deployer**: *Public Authority* OR *Public Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: *Natural Person*
(7b)	**1. Domain**: *Migration Management* OR *Asylum Management* OR *Border Control Management* **2. Purpose**: *Assessing People Related Risk* OR *Assessing Security Risk* OR *Assessing Health Risk* OR *Assessing Risk Of Irregular Immigration* **3. Capability**: Any **4. Deployer**: *Public Authority* OR *Public Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: *Natural Person* OR *Individual Intends To Enter State* OR *Individual Entered State*
(7c)	**1. Domain**: *Migration Management* OR *Asylum Management* OR *Border Control Management* **2. Purpose**: *Examining Asylum Application* OR *Examining Visa Application* OR *Examining Residence Permits Application* OR *Examining Migration Related Complaints* OR *Evaluating Reliability Of Evidence In Migration Related Applications* **3. Capability**: Any **4. Deployer**: *Public Authority* OR *Public Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: *Natural Person* OR *Asylum Seeker* OR *Residence Permit Applicant* OR *Visa Applicant*
(7d)	**1. Domain**: *Migration Management* OR *Asylum Management* OR *Border Control Management* **2. Purpose**: *Detecting Individuals* OR *Recognising Individuals* OR *Identifying Individuals* **3. Capability**: Any **4. Deployer**: *Public Authority* OR *Public Authority Agent* OR *EU Institution* OR *EU Agency* OR *EU Office* OR *EU Body* **5. AI subject**: *Natural Person*

Table C.8: Analysis of Annex III, Point 8 high-risk AI systems

Clause	The 5 concepts
(8a)	**1. Domain**: *Administration Of Justice* OR *Administration Of Democratic Processes* **2. Purpose**: *Researching Facts* OR *Interpreting Facts* OR *Researching Law* OR *Interpreting Law* OR *Applying The Law To Facts* OR *Settling Dispute* **3. Capability**: Any **4. Deployer**: *Judicial Authority* OR *Judicial Authority Agent* **5. AI subject**: Any
(8b)-1	**1. Domain**: *Administration Of Justice* OR *Administration Of Democratic Processes* **2. Purpose**: *Influencing Election Outcome* OR *Influencing Referendum Outcome* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: Any
(8b)-2	**1. Domain**: *Administration Of Justice* OR *Administration Of Democratic Processes* **2. Purpose**: *Influencing Voting Behaviour* **3. Capability**: Any **4. Deployer**: Any **5. AI subject**: *Natural Person*

APPENDIX

FOUR

ANALYSIS OF ANNEX IV ON TECHNICAL DOCUMENTATION

The following provides an analysis of Annex IV, as stated in the European Commission's proposed AI Act [28]. Within this, the ID of requirements explicitly mentioned in the legal text starts with **R** and ID of those extracted from standards or added with the purpose of structuring starts with **A**.

D.1 General Description of AI System

> *1. A general description of the AI system including: (a) its intended purpose, the persons developing the system the date and the version of the system;*

R1a-1. AI system's intended purpose,
AR1a-2 AI capability(ies),
R1a-3. AI developer(s),
AR1a-4. AI provider(s),
R1a-5. AI system's release date,
R1a-6. AI system's version,

> *(b) how the AI system interacts or can be used to interact with hardware or software that is not part of the AI system itself, where applicable;*

R1b-1. External[1] software the AI system interacts with,
R1b-2. Details of interaction with external software,
R1b-3. External hardware the AI system interacts with,
R1b-4. Details of interaction with external hardware,

[1]Not part of the AI system.

R1b-5. External software that can be interacted with using the AI system,
R1b-6. Details of interaction with external software through the AI system,
R1b-7. External hardware that can be interacted with using the AI system,
R1b-8. Details of interaction with external hardware through the AI system,

> *(c) the versions of relevant software or firmware and any requirement related to version update;*

AR1c-1. AI system's version release note,
R1c-1-1. AI version update's software requirements (dependencies),
R1c-1-1-1. Software the AI system is dependent on,
R1c-1-1-2. Version of the software the AI system is dependent on,
AR1c-1-2. AI version update's hardware requirements,
R1c-1-3. AI version update's firmware requirements,
R1c-1-3-1. Firmware the AI system is dependent on,
R1c-1-3-2. Version of the firmware the AI system is dependent on,
R1c-1-4. AI version update's additional requirements,

> *(d) the description of all forms in which the AI system is placed on the market or put into service;*

R1d-1. Form(s) (modalities) in which the AI system is placed on the market or put into service,
R1d-2. Description of each form in which the AI system is placed on the market or put into service,

> *(e) the description of hardware on which the AI system is intended to run;*

R1e-1. Hardware (components) required for running the AI system,
R1e-2. Description of the hardware (components) required for running the AI system,

> *(f) where the AI system is a component of products, photographs or illustrations showing external features, marking and internal layout of those products;*

R1f-1. Entities (e.g. products, photographs, and illustrations) of which AI system is a component,
R1f-2. External features of the entity of which the AI system is a component,

R1f-3. Marking of the entity of which the AI system is a component,
R1f-4. Internal layout of the entity of which the AI system is a component,

> *(g) instructions of use for the user and, where applicable installation instructions;*

R1g-1. Instruction for use,
R1g-2. Installation instruction,

D.2 Description of the AI Elements and Development Processes

> 2. A detailed description of the elements of the AI system and of the process for its development, including:
> *(a) the methods and steps performed for the development of the AI system, including, where relevant, recourse to pre-trained systems or tools provided by third parties and how these have been used, integrated or modified by the provider;*

R2a-1. Methods used for the development of the AI system,
R2a-2. AI system's development processes (steps),
R2a-3. **Third-party systems**, e.g. pre-trained systems, used in/for the development of the AI system,
R2a-4. Description of how third-party systems have been used, integrated, or modified by the AI provider,
R2a-5. **Third-part tools** used in/for the development of the AI system,
R2a-6. Description of how third-party systems have been used, integrated, or modified by the AI provider,

> *(b) the design specifications of the system, namely the general logic of the AI system and of the algorithms; the key design choices including the rationale and assumptions made, also with regard to persons or groups of persons on which the system is intended to be used; the main classification choices; what the system is designed to optimise for and the relevance of the different parameters; the decisions about any possible trade-off made regarding the technical solutions adopted to comply with the requirements set out in Title III, Chapter 2;*

R2b-1. AI system's design specifications
R2b-1-1. Overall logic of the AI system,

AR2b-1-2. AI system's algorithmic design,

AR2b-1-2-1. Algorithms used within the AI system,

R2b-1-2-2. Logic of the AI system's algorithms,

R2b-1-3. Description of AI design choices made during AI development,

R2b-1-3-1. Rationale of design decisions,

R2b-1-3-2. Assumptions made in regard to design designs,

R2b-1-3-3. AI subjects considered in design decisions,

R2b-1-4. Choices made in regard to classification tasks,

R2b-1-5. Optimisation purpose of the AI system, i.e. quality parameters the AI system is optimised for,

R2b-1-6. Relevance of AI parameters,

R2b-1-7. Trade-offs made in implementing technical solutions to comply with the AI Act's requirements for high-risk AI systems,

> *(c) the description of the system architecture explaining how software components build on or feed into each other and integrate into the overall processing; the computational resources used to develop, train, test and validate the AI system;*

R2c-1. AI system architecture illustrating the components incorporating the system and their relationships,

R2c-2. AI system architecture description (documentation),

R2c-2-1. Software components incorporating the AI system,

R2c-2-2. Description of software component development,

R2c-2-3. Description software components integration,

R2c-2-4. Description of how software components are integrated into the overall processing of the AI system,

R2c-3. Computational resources used in different stages of the AI lifecycle,

R2c-3-1. Computational resources used for AI development,

R2c-3-2. Computational resources used for training of the AI system,

R2c-3-3. Computational resources used for testing of the AI system,

R2c-3-4. Computational resources used for validation of the AI system,

> *(d) where relevant, the data requirements in terms of datasheets describing the training methodologies and techniques and the training data sets used, including information about the provenance of those data sets, their scope and main characteristics; how the data was obtained and selected; labelling procedures (e.g. for supervised learning), data cleaning methodologies (e.g. outliers detection);*

AR2d-1. Models trained,

R2d-1-1. Training methodology,

R2d-1-2. Training technique,

R2d-1-3. data requirements which include:

R2d-1-3-1. Information about the datasets used for training the model (training datasets),

R2d-1-3-1-1. Training dataset's provenance information,

R2d-1-3-1-2. Training dataset's scope,

R2d-1-3-1-3. Training dataset's characteristics,

R2d-1-3-1-4. Training data acquisition process (how the training dataset was obtained),

R2d-1-3-1-5. Training data selection process (how each dataset was selected),

R2d-1-3-1-6. Data labelling procedure,

R2d-1-3-1-7. Data cleaning methodologies,

> *(e) assessment of the human oversight measures needed in accordance with Article 14, including an assessment of the technical measures needed to facilitate the interpretation of the outputs of AI systems by the users, in accordance with Articles 13(3)(d);*

AR2e-1. Description of human oversight measure,

AR2e-1-1. Purpose of human oversight measure, e.g. interpretation of the output (see Article 14(4) for more examples) *AR2e-1-2.* The risks the human oversight measures aim to minimise (Article 14(2)),

AR2e-1-3. Type of human oversight measure, e.g. technical and organisational,

AR2e-1-4. Implementation type of human oversight measure, as described by Article 14(3): "measures (a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service; (b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user."

R2e-2. Assessment of Human oversight measures,

> *(f) where applicable, a detailed description of pre-determined changes to the AI system and its performance, together with all the relevant information related to the technical solutions adopted to ensure continuous compliance of the AI system with the relevant requirements set out in Title III, Chapter 2;*

R2f-1. Description of pre-determined changes to the AI system,

R2f-2. Description of pre-determined changes to the AI system's performance,

R2f-3. Technical solutions in place to ensure compliance with the requirements of high-risk AI systems in the aftermath of the pre-determined changes to the AI system and its performance,

> *(g) the validation and testing procedures used, including information about the validation and testing data used and their main characteristics; metrics used to measure accuracy, robustness, cybersecurity and compliance with other relevant requirements set out in Title III, Chapter 2 as well as potentially discriminatory impacts; test logs and all test reports dated and signed by the responsible persons, including with regard to pre-determined changes as referred to under point (f).*

R2g-1. AI system validation procedures,

R2g-2. Information about the datasets used for validating the model (validation datasets),

R2g-3. AI system testing procedures,

R2g-4. Information about the datasets used for testing the model (testing datasets),

R2g-5. AI system characteristics, including but not limited to accuracy, robustness, cybersecurity, compliance with the requirements of high-risk Act, and bias,

R2g-5-1. Metrics used to measure the characteristic,

AR2g-5-2. Tests, benchmarks and/or standards used for measuring the metric and their results,

AR2g-6. Test documentation,

R2g-6-1. Test log,

R2g-6-1-1. Test log date,

R2g-6-1-2. Responsible person(s) for test log,

R2g-6-1-3. Log of tests conducted in regard to the AI system's pre-determined changes,

R2g-6-2. Test report,

R2g-6-2-1. Test report date,

R2g-6-2-2. Responsible person(s) for test report,

R2g-6-2-3. Report of tests conducted in regard to the AI system's pre-determined changes,

D.3 Monitoring, Functioning, and Control

> *3. Detailed information about the monitoring, functioning and control of the AI system, in particular with regard to its capabilities and limitations in performance, including the degrees of accuracy for specific persons or groups of persons on which the system is intended to be used and the overall expected level of accuracy in relation to its intended purpose; the foreseeable unintended outcomes and sources of risks to health and safety, fundamental rights and discrimination in view of the intended purpose of the AI system; the human oversight measures needed in accordance with Article 14, including the technical measures put in place to facilitate the interpretation of the outputs of AI systems by the users; specifications on input data, as appropriate;*

R3-1. Information about monitoring of the AI system,
R3-1-1. Monitoring of the AI system's capabilities,
R3-1-1-1. Monitoring of the AI system's performance,
R3-1-1-2. Monitoring of the AI system's accuracy for intended AI subjects,
R3-1-1-3. Monitoring of the AI system's overall accuracy,
R3-1-2. Monitoring of the AI system's limitations,
R3-1-2-1. Monitoring of the AI system's performance limitations,
R3-1-2-2. Monitoring of the AI system's limitations in regard to the degree of accuracy for intended AI subjects,
R3-1-2-3. Monitoring of the AI system's limitations in regard to its overall degree of accuracy,
R3-1-3. Monitoring for foreseeable unintended outcomes,
R3-1-4. Monitoring for risk sources,
R3-1-4-1. Monitoring for sources of risk to health,
R3-1-4-2. Monitoring for sources of risk to safety,
R3-1-4-3. Monitoring for sources of risk to fundamental rights,
R3-1-4-4. Monitoring for sources of risk to non-discrimination, i.e. bias risk,
R3-1-5. Monitoring of human oversight measures,
R3-1-6. Monitoring of input data as per its specifications,
R3-2. Information about functioning of AI system,
R3-2-1. AI system's capabilities,
R3-2-1-1. AI system's performance,
R3-2-1-2. AI system's accuracy for intended AI subjects,
R3-2-1-3. AI system's overall accuracy,
R3-2-2. AI system's limitations,

R3-2-2-1. AI system's performance limitations,

R3-2-2-2. AI system's limitations in regard to the degree of accuracy for intended AI subjects,

R3-2-2-3. AI system's limitations in regard to its overall degree of accuracy,

R3-2-3. Functioning of the system in the event of foreseeable unintended outcomes,

R3-2-4. Functioning of the system in the event of materialisation of risk sources,

R3-2-4-1. Functioning of the system in the event of materialisation of sources of risk to health,

R3-2-4-2. Functioning of the system in the event of materialisation of sources of risk to safety,

R3-2-4-3. Functioning of the system in the event of materialisation of sources of risk to fundamental rights,

R3-2-4-4. Functioning of the system in the event of materialisation of sources of discrimination risk,

R3-2-5. Functioning of human oversight measures,

R3-2-6. Functioning of AI system in regard to input data specifications, *R3-3.* Information about control of AI system, *R3-3-1.* Controls in place to ensure the AI system's expected capabilities,

R3-3-1-1. Controls in place to ensure the AI system's expected level of performance,

R3-3-1-2. Controls in place to ensure the AI system's expected level of accuracy for intended AI subjects,

R3-3-1-3. Controls in place to ensure the AI system's expected overall accuracy,

R3-3-2. Controls in place in regard to the AI system's limitations,

R3-3-2-1. Control in place in regard to the AI system's performance limitations,

R3-3-2-2. Controls in place in regard to the AI system's limitations regarding the degree of accuracy for intended AI subjects,

R3-3-2-3. Controls in place in regard to the AI system's limitations regarding its overall degree of accuracy,

R3-3-3. Controls for foreseeable unintended outcomes,

R3-3-4. Controls in place regarding risk sources,

R3-3-4-1. Controls in place regarding sources of risk to health,

R3-3-4-2. Controls in place regarding sources of risk to safety,

R3-3-4-3. Controls in place regarding sources of risk to fundamental rights,

R3-3-4-4. Controls in place regarding sources of discrimination risk,

R3-3-5. Controls in place in regard to human oversight measures,

R3-3-6. Control of AI system in regard to input data specifications,

D.4 Risk Management System

4. A detailed description of the risk management system in accordance with Article 9;

R4-1. Description of AI risk management system,

AR4-1-1. Role of the organisation in relation to the AI system,

AR4-1-2. External context of AI system related to the AI risk management system,

AR4-1-3. Internal context of AI system related to the AI risk management system,

AR4-1-3-1. Intended purpose of the AI system,

AR4-1-4. Needs and expectations of stakeholders (interested parties) in regard to AI risk management,

AR4-1-5. Scope of AI risk management system,

AR4-1-6. AI risk management policies,

AR4-1-7. AI risk management system roles and responsibilities,

4-1-8. AI risk management information,

AR4-1-8-1. Scope of AI risk management,

AR4-1-8-2. AI risk management objectives,

AR4-1-8-3. AI risk management tools,

AR4-1-8-4. AI risk management techniques,

AR4-1-8-5. AI risk management resources,

AR4-1-8-6. AI risk management responsibilities,

AR4-1-8-7. Internal context of the AI system related to AI risk management,

AR4-1-8-8. External context of the AI system related to AI risk management,

AR4-1-8-9. AI Risk criteria (for evaluation of risk significance),

AR4-1-8-10. AI risk assessment information,

AR4-1-8-10-1. AI risk identification information,

AR4-1-8-10-1-1. Assets and their value,

AR4-1-8-10-1-2. Risk sources (events),

AR4-1-8-10-1-3. Entities associated with risk sources,

AR4-1-8-10-1-4. Risks,
AR4-1-8-10-1-5. Consequences,
AR4-1-8-10-1-6. Impacts,
AR4-1-8-10-2. AI risk analysis information,
AR4-1-8-10-2-1. Analysis of risk sources,
AR4-1-8-10-2-2. Analysis of risks,
AR4-1-8-10-2-3. Analysis of consequences,
AR4-1-8-10-2-4. Analysis of impacts,
AR4-1-8-10-3. AI risk evaluation information,
AR4-1-8-10-3-1. Evaluation of risk sources,
AR4-1-8-10-3-2. Evaluation of risks,
AR4-1-8-10-3-3. Evaluation of consequences,
AR4-1-8-10-3-4. Evaluation of impacts,
AR4-1-8-11. AI risk treatment information,
AR4-1-8-11-1. Statement of applicability,
AR4-1-8-11-1-1. Control measures,
AR4-1-8-11-1-2. Objectives of the measures,
AR4-1-8-11-1-3. Residual risk,
AR4-1-8-12. AI system impact assessment,
AR4-1-9. AI quality objectives,
AR4-1-10. AI management system change plan,
AR4-1-11. AI risk management system resources,
AR4-1-12. Information regarding competence,
AR4-1-13. Information regarding operation of AI risk management system processes,
AR4-1-14. Results of AI risk assessments,
AR4-1-15. Results of AI risk treatments,
AR4-1-16. Results of AI impact assessments,
AR4-1-17. Results of monitoring, measurement, analysis and evaluation of the AI risk management system,
AR4-1-18. Information regarding implementation of the audit programme,
AR4-1-19. Audit results,
AR4-1-20. Information regarding AI risk management system review,
AR4-1-21. Information regarding non-conformity and corrective actions,

D.5 Changelog

> *5. A description of any change made to the system through its lifecycle;*

R5-1. Description of the changes made to the AI system,
AR5-2. Description of the changes made to the components incorporating the AI system,

D.6 Harmonised Standards

> *6. A list of the harmonised standards applied in full or in part the references of which have been published in the Official Journal of the European Union; where no such harmonised standards have been applied, a detailed description of the solutions adopted to meet the requirements set out in Title III, Chapter 2, including a list of other relevant standards and technical specifications applied;*

R6-1. **Harmonised standards** applied,
R6-1-1. Level of conformity, i.e. full or in part,
AR6-1-2. Type of conformity assessment, e.g. self-assessment, third-party assessment,
R6-2. **Other standards** applied,
R6-2-1. Level of conformity, i.e. full or in part,
R6-2-2. The AI Act's high-risk AI requirements met by applying the standard,
AR6-2-3. Type of conformity assessment, e.g. self-assessment, third-party assessment,
R6-3. **Technical specifications** applied,
R6-3-1. Level of conformity, i.e. full or in part,
R6-3-2. The AI Act's high-risk AI requirements met by applying the technical specification,
AR6-3-3. Type of conformity assessment, e.g. self-assessment, third-party assessment,
R6-4. Description of **other solutions** adopted to meet the AI Act's high-risk AI requirements

D.7 EU Declaration of Conformity

> *7. A copy of the EU declaration of conformity;*

R7-1. **EU declaration of conformity**,

D.8 Post-Market Monitoring System

> *8. A detailed description of the system in place to evaluate the AI system performance in the post-market phase in accordance with Article 61, including the post-market monitoring plan referred to in Article 61(3).*

R8-1. Description of the **AI performance evaluation plan**, which is a part of the AI management system (see ISO/IEC 42001, point 9 on performance evaluation),

R8-2. **Post-market monitoring plan**, according to Article 61(3), an implementing act containing a template for the post-market monitoring plan will be adopted by the European Commission.

APPENDIX

FIVE

AI CARDS SURVEY

E.1 Informed Consent

The consent form, which is designed following the Trinity College Dublin's research ethics and data protection policies and guidelines, is shown at the beginning of the questionnaire as follows:

Important notes for the participants

- Individual results are anonymous and your data is not identifiable.
- The resulting data will be stored as summary tables with numeric ratings provided, the opinions provided in textual format, and the numeric answers of the usability test.
- The resulting data will be stored safely using the IT services called MyZone Google Drive which complies with GDPR rules.
- The lead researcher (Delaram Golpayegani) and the collaborators (Dr. Isabelle Hupont (JRC) , Dr. Cecilia Panigutti (JRC), Dr. Sven Schade (JRC), Prof. Harshvardhan Pandit (DCU), Prof. Declan O'Sullivan (TCD), and Prof. Dave Lewis (TCD)) will be the only people with access to the data until its publication in an open data repository.
- We will perform a quantitative analysis of the ranges and the usability scores using statistical summaries, reporting aggregated results. Opinion analysis will be conducted and documented manually by the lead researcher. None of your personal details will be recorded.
- You are free to withdraw your consent at any time until submitting the questionnaire.

Publication

- The result of this work would be published in relevant academic conferences and journals, as well as the PhD thesis of the lead researcher (Delaram Golpayegani) at Trinity College Dublin.

Conflict of interest

- The collaborators will not participate in the survey.

Declaration

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data is used for scientific purposes, and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that if the results of the research have been published, or my data has been fully anonymised so that it can no longer be attributed to me, then it will no longer be possible to withdraw.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation [request to] have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.

- I understand that my participation is fully anonymous and that no personal details about me will be recorded.
- I understand that if I or anyone in my family has a history of epilepsy then I am proceeding at my own risk.
- I understand that personal information about me, including the transfer of this personal information about me outside of the EU, will be protected in accordance with the General Data Protection Regulation.

☐ I consent to participate in this study, and consent to the data processing necessary to enable my participation and to achieve the research goals of this study.

E.2 Survey Questions and Results

E.2.1 Background Question

1. What is the sector of your current employment? Please select all that apply to you.

☐ Public organisation

☐ EU institution

☐ Industry

☐ Academia and research

☐ NGO

☐ Other

See the distribution of participant's sector of employment in Figure E.1.

2. Please select all the roles that apply to you:

☐ AI developer

☐ AI researcher (in any areas, such as computer science, law, and ethics)

☐ AI deployer (i.e. deploying AI in domains such as public sector and healthcare)

☐ AI auditor

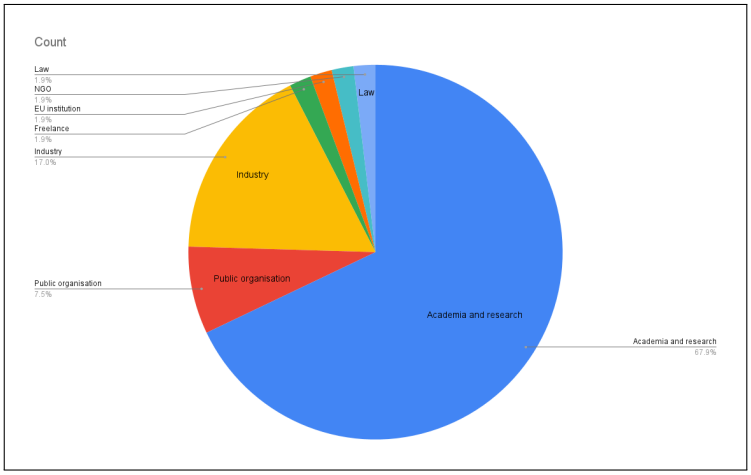


Figure E.1: Distribution of participants' sector of employment

☐ AI Policymaker

☐ AI Standardiser

☐ AI consultant

☐ Legal expert

☐ Technology ethicist

☐ Other:

See the distribution of participants' roles in Figure E.2.

3. What is your level of familiarity with the EU AI Act ? (Mark only one)

☐ Extremely familiar [I have read the AI Act and have knowledge about most of the provisions o the EU AI Act]

☐ Moderately familiar [I have read some parts of the EU AI Act and have knowledge about some of the provisions of the AI Act]

☐ Somewhat familiar [I am aware of the overall objectives, structures, and requirements of the EU AI Act]

☐ Slightly familiar [I am aware of the overall objectives of the EU AI Act]

☐ Not familiar at all [I have no knowledge or awareness of the EU AI Act]

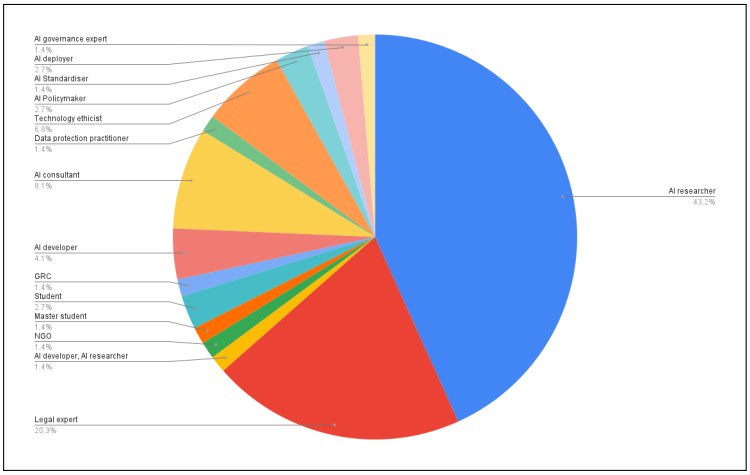


Figure E.2: Distribution of participants' roles

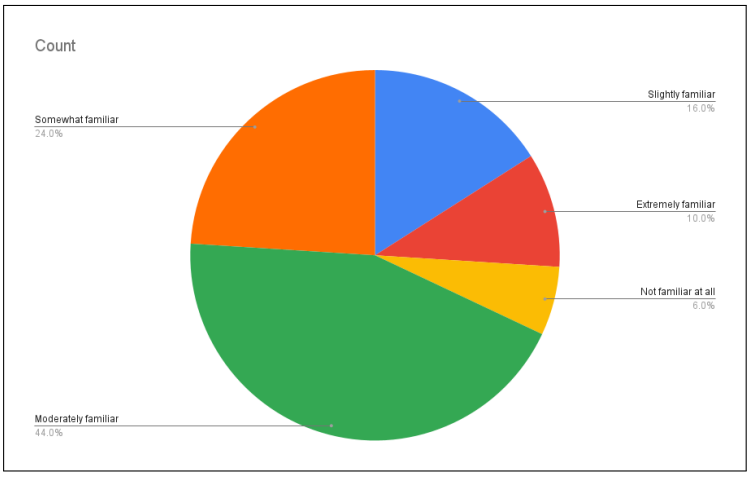


Figure E.3: Participants' level of familiarity with the AI Act

See participants' level of familiarity with the AI Act in Figure E.3.

4. If there are any specific articles or aspects of the EU AI Act that you are particularly familiar with, please select them:

☐ AI risk management (Art. 9)

☐ Data and data governance (Art. 10)

☐ Technical documentation (Art. 11 and Annex IV)

☐ Transparency (Art. 13)

☐ Human oversight (Art. 14)

☐ Accuracy, robustness and cybersecurity (Art. 15)

☐ Conformity assessment (Art. 43)

☐ Harmonised standards (Art. 40)

☐ Other:

For the participants' familiarity with AI Act articles see Figure E.4.

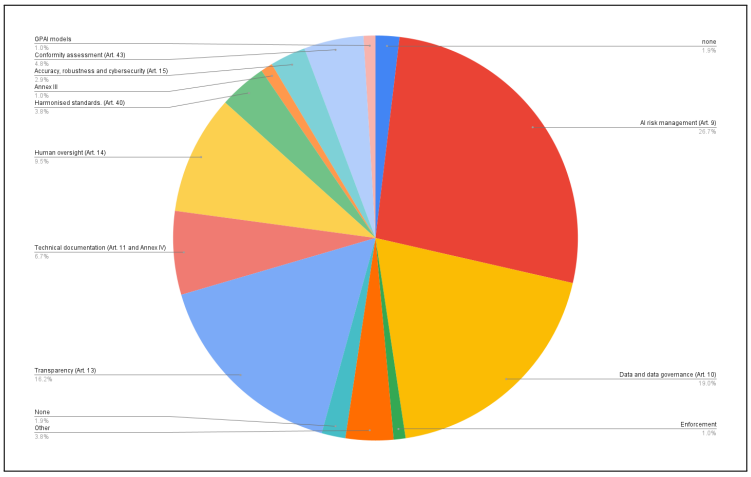


Figure E.4: AI Act Articles the participants were most familiar with

E.2.2 Visual Representation of AI Cards

This section asks your opinion about the visual representation of the AI Cards show below. Also, you can find an example of the visual representation of AI Cards for an AI-based student proctoring system here: https://delaramglp.github.io/aicards/example/.

1a. In general, how would you rate the usefulness of the AI Cards visualisation in representing key AI and risk information? (please don't limit yourself to any regulation)

☐ 0 [Not useful]

☐ 1

☐ 2

☐ 3

☐ 4 [Very useful]

For the results, refer to Figure E.5.

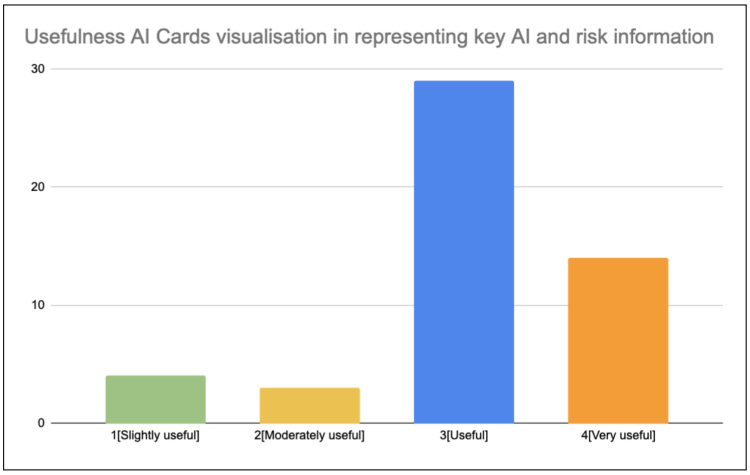


Figure E.5: Usefulness of the human-readable representation of AI Cards

1b. If there is any key information that you miss in the visualisation of AI Cards, please list them; noting that an AI Card acts as a summary card of an AI system and its risks?

See the participants' comments in Table E.1, Table E.2, and Table E.3. The participants' comments in these Tables, and the subsequent ones, have not been modified in any way.

1c. Is there any information within the visualisation of AI Cards that you believe should not be presented? If yes, please list them.

See the participants' comments in Table E.4 and Table E.5.

Table E.1: The information participants missed in the human-readable representation of AI Cards - part 1

Sector	Role	Level of familiarity with the Act	Comment
Academia and research	AI researcher	Slightly familiar	"1. Specifics about the deployer (which university, which department, contact point) 2. Link to further legal compliance documents (e.g. GDPR Privacy Notice)"
Industry	AI researcher, AI governance expert	Extremely familiar	"Details about the anomalies (which law article is not fulfilled, which quality requirement is violated, etc.)"
Academia and research	Legal expert	Moderately familiar	"method of risk assesment + other fundamental rights"
Industry, Academia and research	AI developer, AI researcher	Moderately familiar	"Descriptions of potential risks specific to the model"
EU institution	AI researcher	Moderately familiar	"As a summary, it is a very nice representation. But I cannot help thinking that all data visually marked on the card must be adequately substantiated, and for me that "substance" is the most relevant part when assessing risk mitigation. For example, "fundamental rights" includes many types of fundamental rights, and the AI card does not allow to identify which ones. Or, when referring to the measures, it is unclear if these measures are efficient and relevant or not. I guess that the AI Card has to be understood to be accompanied by additional information to support everything set out in it. I also miss the clear identification regarding the fact that the card focuses on "systems" rather than on models (note that a very important part of the AI Act now focuses on general-purpose AI models). "
Academia and research	AI researcher	Moderately familiar	"I think maybe in the impact section, it could be more specific about what kind of rights are meant. Individual rights? Fundamental rights?"
Academia and research	AI researcher	Moderately familiar	"There is no overall summary/traffic light that shows me the status, just many details"
Academia and research	Legal expert	Somewhat familiar	"Information related to the registry process"

Table E.2: The information participants missed in the human-readable representation of AI Cards - part 2

Sector	Role	Level of familiarity with the Act	Comment
Academia and research	AI researcher	Extremely familiar	"I think more information needs to be represented in the general information to assist with using the AI cards at later stages by other actors in the lifecycle. "
Academia and research	AI researcher	Moderately familiar	"Maybe the link to the GDPR can be useful in this visualization, and for full compliance maybe adding a short version of the requirements of data processing, that would make the cards useful specifically for enterprises "
Industry	AI consultant, Legal expert	Moderately familiar	"Heading: Controls/Limitations - retention period of data if personal data processed - last test/bug/issue and date "
Public organisation	Data protection practitioner in public sector	Somewhat familiar	"A legend/key to explain the abbreviations. (Could be a link to an online resource.)"
Academia and research	AI researcher	Slightly familiar	"They seem very confusing"
Public organisation	AI researcher	Somewhat familiar	"What is the role of the model? What is it actually detected? How do you ensure fairness? Such explainability is important - especially for the general public"
Academia and research	Technology ethicist	Moderately familiar	"It would be great to capture the as-of-yet unknown aspects and risks. The cards currently presuppose that all/main uses and impacts can already be identified."
Academia and research	AI researcher	Moderately familiar	"useful, but it does make it appear as if risk is a linear, uniform category. Perhaps it appears like that in law (even though it is not), but it certainly is a situation of comparing apples and pears when seen from an interpretive perspective. That means I feel a certain hesitation / resistance when looking at the cards. Not my notion of risk. I have the same feeling with risk in the AI act."
Academia and research	AI researcher	Somewhat familiar	"No it looks quite comprehensive to me"

Table E.3: The information participants missed in the human-readable representation of AI Cards - part 3

Sector	Role	Level of familiarity with the Act	Comment
Academia and research	AI developer, AI researcher	Somewhat familiar	"Where is AI is run (i.e. in house by the developer, on the cloud, etc) Whether 3r parties can see that AI, its data, or its output Where the data is held (i.e. in house by the developer, on the cloud, etc) How the data was sourced (for example, scraped from online, licenced, etc – and whether explicit permission was granted or whether there could be a copyright conflict)"
Academia and research	AI researcher	Moderately familiar	"It may be just the representation as an image but to be useful, the card needs to have links to explanations or illustrative examples so that the user can quickly understand the scope of each of the sections. Also, such assistance would provide more insight into the importance/impact/consequences of each section. Finally it's not really clear who the form/card is intended for. I had thought it was intended for the developers/implementers to track/document/manage risk and compliance. However, considering it again, is this intended for a 3rd party to explain/provide verification of compliance?"
Academia and research	AI researcher, AI Standardiser	Slightly familiar	"scores"

Table E.4: The information participants identified as additional in the human-readable representation of AI Cards - part 1

Sector	Role	Level of familiarity with the Act	Comment
Academia and research	AI researcher	Slightly familiar	"Versions of components: Will the AI Card version change (and will users have to review them then) with every update of any component, although the other elements of the card remain the same?"
Academia and research	Legal expert	Moderately familiar	"instead of checklist add more space for descriptive analysis"
Industry, Academia and research	AI developer, AI researcher	Moderately familiar	"Quality heptagon"
Academia and research	AI researcher	Moderately familiar	"I'm not sure about point 7. Quality. Maybe it is important to include, but I don't know where these principles come from? Are they included in the AI Act?"
Academia and research	AI researcher	Moderately familiar	"There does not seem to be a place to list input datasets that are not personal or sensitive. It may be important to know these as they could be combined with the other types within the system"
Law	Legal expert, Student	Moderately familiar	"Place in which the card was issued, as well as a declaration of good faith filling "
Academia and research	AI researcher	Slightly familiar	"I'm not sure what the predetermined changes section is for"

Table E.5: The information participants identified as additional in the human-readable representation of AI Cards -part 2

Sector	Role	Level of familiarity with the Act	Comment
Academia and research	AI researcher	Moderately familiar	"there are something that are unclear, but I don't run an AI so perhaps it's me. I don't get what "modality" refers to. From the perspective of "user studies" I would state that the input - output model is an abstraction while in reality it is not that simple. Everyone knows that, but the cards put the respondent in a position where there is no choice but to go along with it."
Academia and research	AI researcher	Moderately familiar	"I'm not exactly clear what you mean by 'components' in section 3. If you wish to identify data sets, (AI) models used, general purpose then why not ask for them directly? There are so many 'important' components in a system at different levels of abstraction etc. it's easy to get too complex (and therefore overwhelming for the reader & developer). It may be easier just to ask about the dataset used or generated in the system (indicating what components are responsible for them), AI models (indicating what components are responsible for their execution) etc., general purpose etc. Also it's a little surprising why you ask for general purpose after asking about individual models (cognitively this would normally come first)."

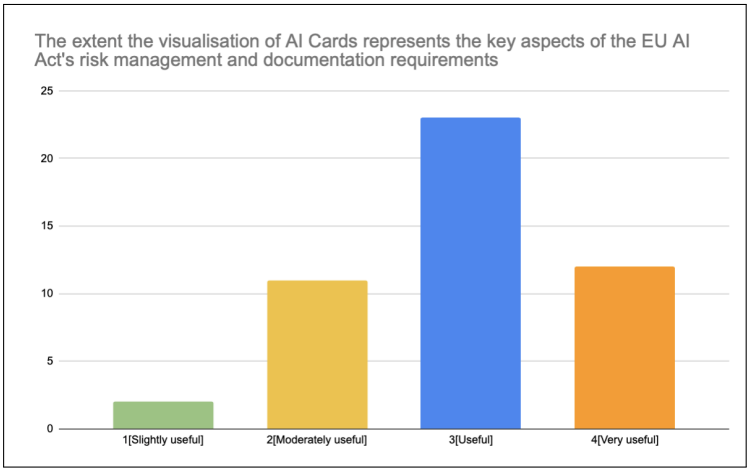


Figure E.6: Participants' view on the extent the visualisation of AI Cards represents the key risk and AI information as per the AI Act

2a. To what extent the visualisation of AI Cards represents the key aspects of the EU AI Act risk management and documentation requirements?

☐ 0 [No extent]

☐ 1

☐ 2

☐ 3

☐ 4 [Large extent]

2b.If you can, please elaborate on your response.
For comments refer to Table E.6.

Table E.6: Comments regarding the extent the visualisation of AI Cards represents the key aspects of the EU AI Act's risk management and documentation requirements

Comment
"I am not sure, which are the key aspects ot AI Act's requirements."
"as in the GDPR, risk management is an undefined notion leading to lower the assessment and thus the management."
"It's high level, and risk management is about processes, hence I don't think a high-level visualization would provide any actionable insight into the AI Act RM. "
"Sorry not to be more positive, but again, although it is a very interesting summary, the mere visualisation does not represent the key aspects in my opinion. For the same reason than before, I also miss the part on risk management and document requirements for GPAI models, which is now one of the most important parts of the EU AI Act. "
"I can see that many areas of consideration are covered."
"It closely resembles the requirements in the conformity assessments, making it extremely relevant."
"They give a concise first approach to the risk-based method upon which the AI Act was built. "
"As it is based ithe 5 pillars of the AI act, it contains the necessary information"
"It presents the main information that allows the AI agent (provider, deployer, user or regulatory bodies) to understand the key operations made by the AI system and the risks involved."
"These above components provide an overview of the AI system which makes it easier for auditing and standardisation across multiple AI systems within a company. The real world application would be interesting to see and may require changing based on different Memeber State authorities requiring other information be made available for documentation purposes."
"I think that the card provides a very useful overal indicative summary, but which would complement detailed documentation. For example, documentation which elaborates on each section of the card."
"The field of application will have its own specific requirements, updates at each level may be required to fulfill both transparency and data protection."
I chose "2" as there is no option "I don't know" and I am not familiar with the details of the act. You should test your survey with people outside your expert group before rolling it out to catch issues like this."
"I think it might be important to include a more detailed justification for why and how the risk level was determined, as well to outline specific risks introduced by the AI system."
"they are not the key aspects to me. But, I can imagine that the legal categorization of risk is very important to others. So the idea is very good. But it seems as if the cards are intended to have a decisive result, (e.g. what level of risk) The function of the cards seems to be a discussion, or reflection. Which is it? Let's elaborate (why not) The risk category in law has a very specific history in policy making (going back to Chernobyl and nuclear science). From the point of view of my field, this is the second time that "european law" takes shape around a category of risk. Naturally this risk category is alien to AI or computer science. Rather than normalizing its usage, through discussions and reflection, the process seems to be to formalize the category instantly. On the one hand, the cards could help, opening up the reflection on risk in AI, on the other hand, the legal conception of law is presented as fact, as a fixed category, as simply a matter of implementation. The cards are a bit of both, and i fear that will limit its utility."
"It looks great. I know it is a summary but some short descriptive text of the key words in general information would be useful."

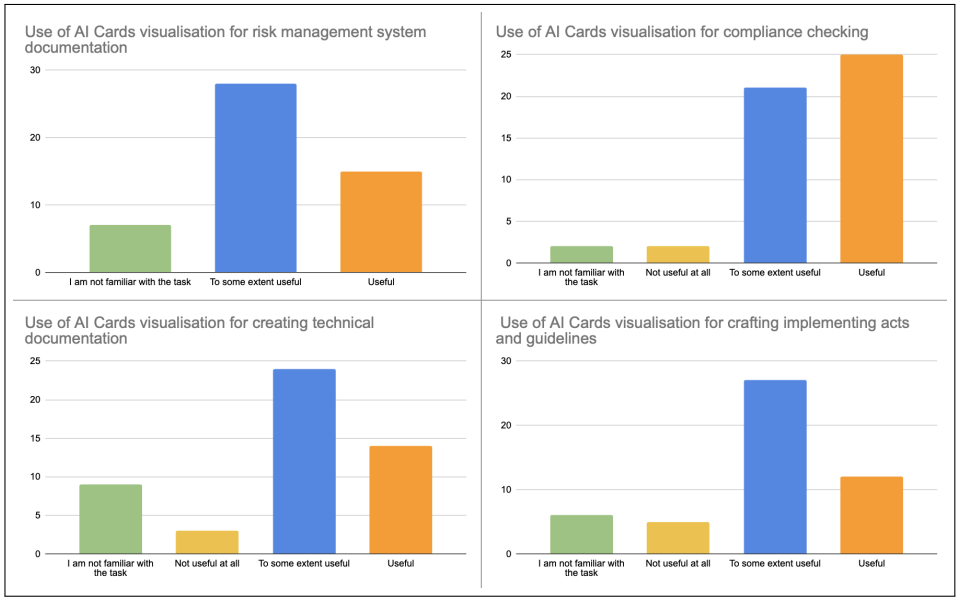


Figure E.7: Usefulness of the human-readable representation of the AI Cards for AI Act compliance and enforcement tasks

3a. In regard to implementation and enforcement of the EU AI Act, how do you rate the use of AI Cards visualisation to assist with the following tasks?

- Documenting risk management system
- Creating technical documentation
- Compliance checking
- Crafting implementing acts and guidelines

For results, see Figure E.7.

3b. If you can, please elaborate on the other potential uses of AI Cards visualisation that you foresee.
For results, see Table E.7.

4a. How would you rate the usefulness of AI Cards as a tool to facilitate exchange of information regarding an AI system and its risks?

- Within AI development eco-system
- For communication with authorities (for legal compliance)

Table E.7: Potential uses of the human-readable representation of AI Cards, identified by participants

Potential uses of human-readable representation of the AI Cards
"summarizing AI systems and scenarios, high-level information for natural persons. Long texts won't be read, probably."
"Dashboarding."
"it can be useful for the end user as well to understand quickly the system"
"Collectible trading cards"
"It is important to visualise many of these things as a time series as the system is often constantly evolving and it is hard to see how this fits with the snapshot provided. Often in Cybersecurity there is also a Plan of Actions and Milestones associated with the risk mitigations so a link to that would be good"
"Raising awareness"
"One of the main benefits I find in the EU cards is the traceability and transparency therefore, I believe they might be very useful for the development p fa process, additionally I believe that keeping them updated is doable"
"AI cards could also be useful for monitoring the system, review periods for pre-determined changes, and explaining the core concepts to clients who lack technical knowledge."
"The cards might be best for comparing different AI technologies, but not so much as a summary for a particular technology- they seem too simplistic but at the same time too busy"
"This system may be useful when examining new regulations and their compliance with existing law and regulations - or need for updating old regulations"
"Could the AI cards be made interactive? They provide a high-level overview of the application, but contextual boxes or links can offer more detailed where needed/required."
"I can imagine a set of cards for discussing AI and risk broadly, a set for categorizing how law would designate risk in relation to a certain AI system, and one that is aimed at communication with third parties."
"Public comms about various AIS and their safety – honeslty I think that's where the greatest contribution of AI cards may lie"

- For communication with the public

See the results in Figure E.8.

E.2.3 Machine-Readable Representation of AI Cards

This section asks your opinion regarding the machine-readable representation of AI Cards. An example of machine-readable representation of AI Cards for an AI-based proctoring system can be accessed here: https:

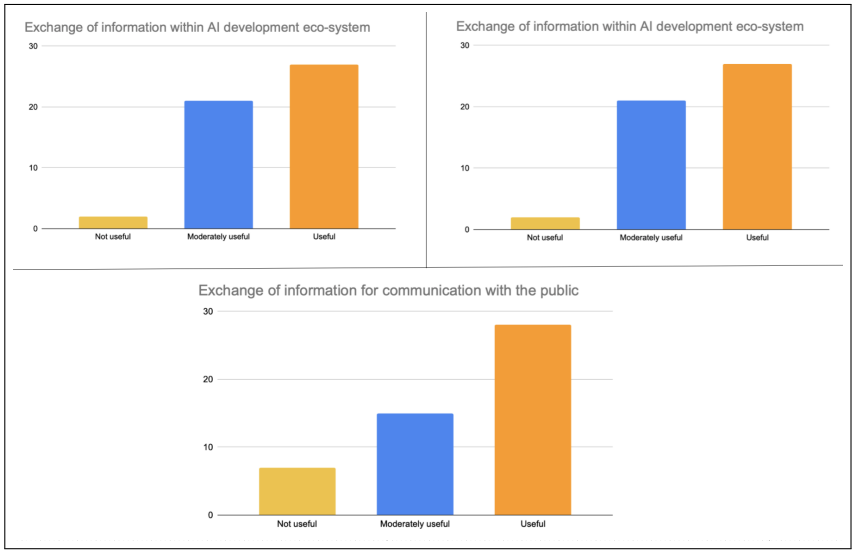


Figure E.8: Usefulness of of the human-readable representation of the AI Cards for information exchange

//delaramglp.github.io/aicards/example/

1a. How would you rate the usefulness of the machine-readable representation of AI Cards in the following tasks?

- Development of automated tools for legal compliance and conformity assessment
- Establishing a common language around AI and its risks
- Structuring databases of AI Systems, e.g. EU database of high-risk AI systems

See the results in Figure E.9.

1b. If you can, please elaborate on the other potential uses of the machine-readable representation of AI Cards that you foresee.

See the potential uses of machine-readable representation of AI Cards from the participants' perspective in Table E.8.

E.2.4 Overall Framework

This section seeks your opinion about the AI Card Framework which includes both human and machine readable formats of AI and risk documentation.

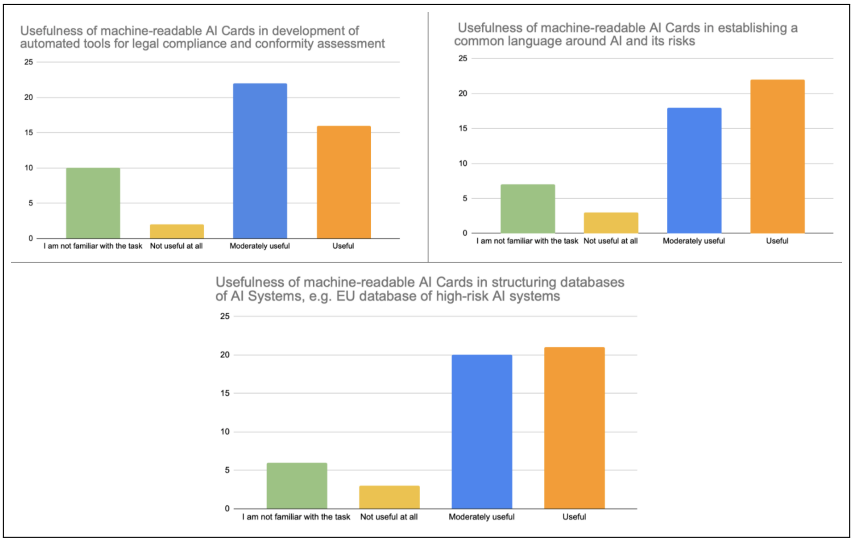


Figure E.9: Usefulness of the machine-readable representation of the AI Cards for tasks related to AI Act's compliance and enforcement

1. Overall, how would you describe the key objective(s) of the AI Cards framework?

- Transparency
- Interoperability
- Accountability
- Trustworthiness
- Comprehensibility
- Other:

See the results in Figure E.10. The following objectives also mentioned by the participants: human oversight, visualisation, terminological alignment, fairness, data protection, interpretation, and comprehension.

2a. How do you rate the degree each stakeholder could benefit from the AI Cards framework?

- AI developers
- AI providers
- AI deployers

Table E.8: Potential uses of the machine-readable representation of AI Cards, identified by participants

Potential uses of machine-readable representation of the AI Cards
"catalog"
"To be honest, I find it difficult to evaluate this part. I can't really tell you whether the selected format is the most suitable or can be improved in some way. "
"Development of cross-organisational AI governance tools within an AI provider toolchain"
"To understand under which risk category an AI system should be subject to"
"Building models with better compliance capabilities."
"I think this cards are very useful to give hace an idea of the models, further information might be required for a deeper understanding"
"May be used as a standard for checking with other AI regulations for example in the US"
"Regarding question 1a above (Establishing a common language around AI and its risks) - for the public this may set up a structured check but it is not educating the public about the AI risk."
"They would make sense as the default approach for ensuring and reporting AI act compliance. However, that would require significant buy-in from stakeholders, or a decision at the regulatory level."
"If i understand correctly it is suppose to be helpful that the cards are machine readable. In the context of establishing a common language that means that the interpretation that they facility is reduced. That's a problem. It's useful as a tool for communication of stakeholders, if that communication is shared fully, in its richness, and experience. Not, if it is brought back to data points or standardized language of policy making. This leads to an extra layer of bureaucracy (as we have seen with previous technologies), and not to greater involvement in the policy making by a variety of stakeholder. Interpassivity (where you have to engage) vs. interactivity (where you want to engage)."
"Machine-readable representation is very useful and detailed with a lot of properties which could be used to boost interoperability"
"I am unsure who would be writing the machine-readable version – if its the developers, and their claims are not audited, it would be quite easy for them to hide something – but that is not uniqu of an issue to AI Cards, of course."
"My larger concern would be the contology used to express their system (your ex:¡name¿). Since this system does not use a formal, defined ontology here, querying it from the end-user perspective will be very hard, as it will require a person to potentially query every AI system individually with different ontologies. I'd be concerned about how usable this system is in the absence of all components being fully-defined ontologyies, in all honesty." I put ""moderately useful"" above mostly because this goes a good way towards that final goal – but without more precise ontologies, I am unsure if this would be widely useful in practice."

- AI auditors, including conformity assessment bodies
- Lay users
- AI policymakers and regulators, including the EU AI office

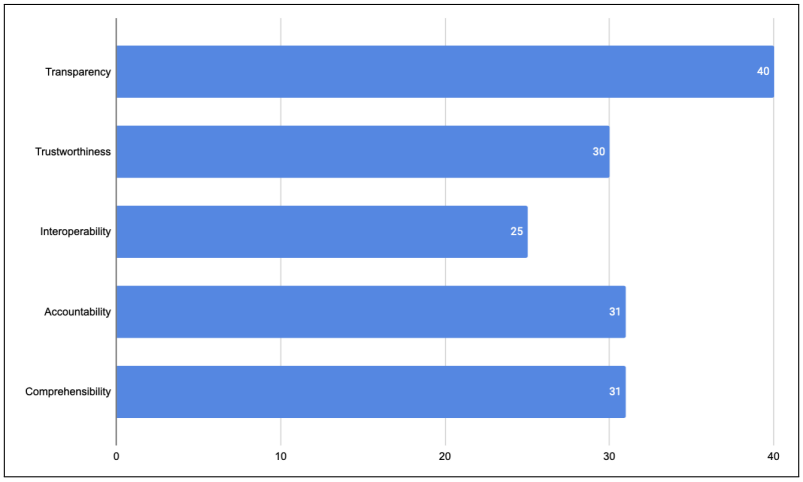


Figure E.10: Distribution of the AI Cards' objectives according to the participants' opinions

- AI standardisation bodies

See results in Figure E.11.

2b. Are there other stakeholders that you think might benefit from the AI Cards?

The participants mentioned the following stakeholders: civil society (3 mentions), researchers (3 mentions) organisations, the general public (2 mentions), AI subjects, end-users, AI brokers (for comparing AI Systems). Also, one of participants stated that: "The danger in this approach is trying to be 'all things to all people'. There may be abstrations of the AI Cards which are useful to different user types (e.g. developers, deployers, policy makers) but the level of depth/specificity is probably different."

AI Cards Framework Usability Test

1. I think that AI stakeholders would like to use the AI Cards framework frequently.

Mean of the results: 3.9, median of the results: 4

2. I find the AI Cards framework unnecessarily complex.

Mean of the results: 2.32, median of the results: 2

3. I think the AI Cards framework is easy to use.

Mean of the results: 3.7, median of the results: 4

4. I think that AI stakeholders would need the support of a technical person to be able to u the AI Cards framework.

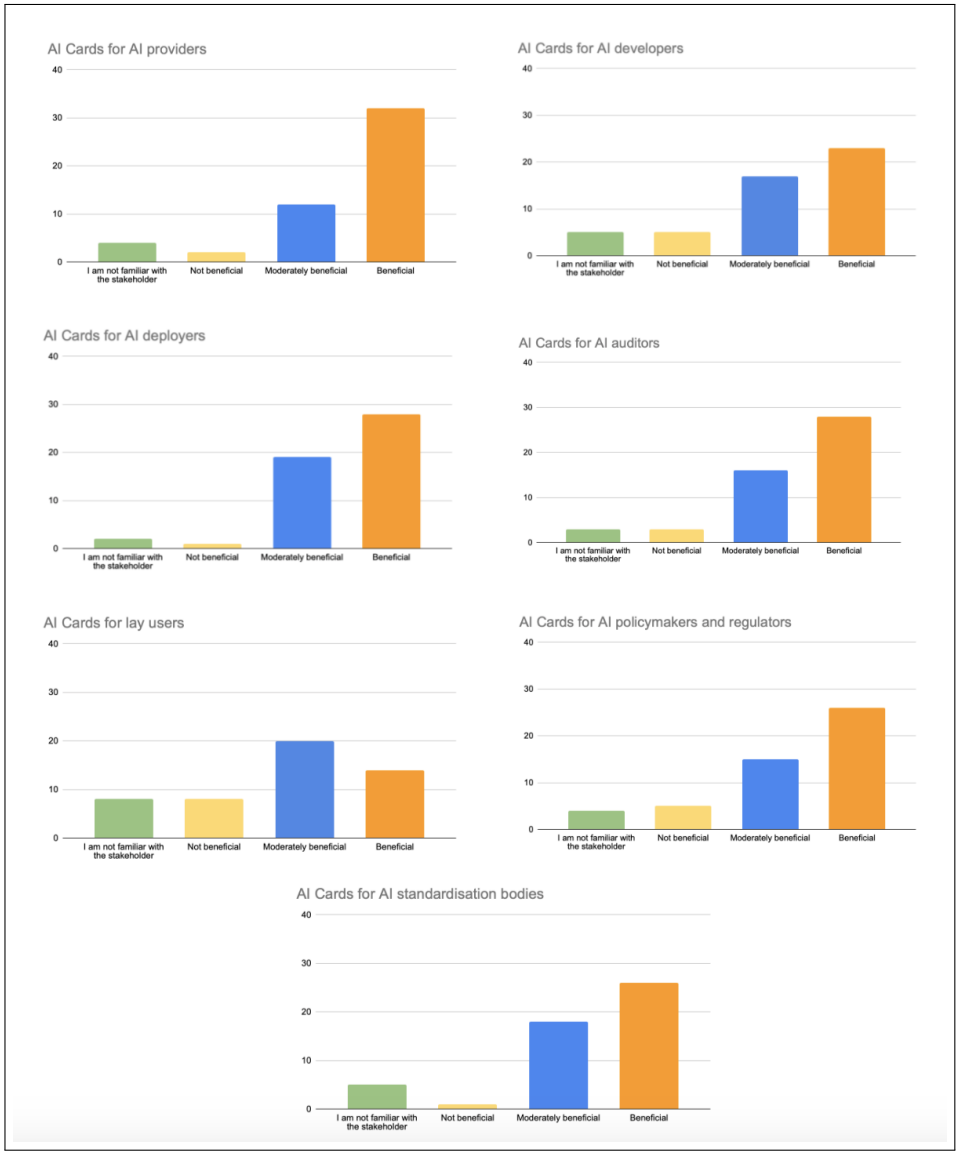


Figure E.11: Benefits of AI Cards for different stakeholders as perceived by participants

Mean of the results: 3.46, median of the results: 3.5

5. I find the various aspects (i.e. information elements and human- and machine-readable formats) in the AI Cards framework are well integrated.

Mean of the results: 3.88, median of the results: 4

6. I think there is too much inconsistency in the AI Cards framework.

Mean of the results: 2.04, median of the results: 2

7. I would imagine that most people would learn to use the AI Cards

framework very quick

Mean of the results: 3.72, median of the results: 4

8. I find the AI Cards framework very cumbersome to use.

Mean of the results: 2.24, median of the results: 2

9. I feel very confident using the AI Cards framework.

Mean of the results: 3.24, median of the results: 3

10. I need to learn a lot of things before I could get going with the AI Cards framework

Mean of the results:3.2, median of the results: 3