
# Piloting a Maturity Model for Responsible Artificial Intelligence: A Portuguese case study

Rui Miguel Frazão Dias Ferreira , António GRILO , Maria MAIA

PII: S2666-6596(25)00013-7
DOI: https://doi.org/10.1016/j.jrt.2025.100117
Reference: JRT 100117

To appear in: *Journal of Responsible Technology*

Received date: 8 July 2024
Revised date: 18 February 2025
Accepted date: 21 March 2025

Please cite this article as: Rui Miguel Frazão Dias Ferreira , António GRILO , Maria MAIA , Piloting a Maturity Model for Responsible Artificial Intelligence: A Portuguese case study, *Journal of Responsible Technology* (2025), doi: https://doi.org/10.1016/j.jrt.2025.100117

## Highlights

- Frameworks to assist trustworthiness Artificial Intelligence can be very time consuming to apply.
- A practical Maturity Model for Responsible AI.
- Piloting Responsible AI in the field, companies and research centres.
- Gaps and needs to foster a trustworthy approach to the development and deployment of AI.

# Piloting a Maturity Model for Responsible Artificial Intelligence: A Portuguese case study

Authors:

- Rui Miguel Frazão Dias Ferreira, FCT NOVA School of Science and Technology, 2829-516 Caparica, Portugal, rmf.ferreira@campus.fct.unl.pt (corresponding author)
- António GRILO, FCT NOVA School of Science and Technology, 2829-516 Caparica, Portugal, antonio.grilo@fct.unl.pt, ORCID: https://orcid.org/0000-0002-6045-9994
- Maria MAIA, Institute for Technology Assessment and Systems Analyses (ITAS), Karlsruhe Institute of Technology (KIT), Karlstr. 11, 76133 Karlsruhe. Germany, maria.maia@kit.edu, ORCID: https://orcid.org/0000-0002-3501-6876



## Abstract

Recently, frameworks and guidelines aiming to assist trustworthiness in organizations and assess ethical issues related to the development and use of Artificial Intelligence (AI) have been translated into self-assessment checklists and other instruments. However, such tools can be very time consuming to apply. Aiming to develop a more practical tool, an Industry-Wide Maturity Model for Responsible AI was piloted in 3 companies and 2 research centres, in Portugal. Results show that organizations are aware of requirements (44%) to deploy a responsible AI approach and have a reactive response to its implementation, as they are willing to integrate other requirements (33%) into their business processes. The proposed Model was welcomed and showed openness from companies to consistently use it, since it helped to identify gaps and needs when it comes to foster a more trustworthy approach to the development and deployment of AI.

Key-words: Responsible Artificial Intelligence, Artificial Intelligence Regulation, Maturity Model, Ethical, Legal and Social Issues, Responsible Research and Innovation.

## 1. INTRODUCTION

According to the OECD, an Artificial Intelligence (AI) system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions, that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment (https://oecd.ai/en/ai-principles, 2024).

Due to its applicability, AI technologies have the potential to disrupt many aspects of human life, bringing countless benefits in areas such as climate action, sustainable infrastructure, health and well-being, quality education, and digital transformation (European Commission, 2019). However, the use of AI with its specific characteristics (namely opacity, complexity, bias and a certain degree of unpredictability) can adversely affect several fundamental rights enshrined in the EU Charter of Fundamental Rights (European Commission, 2021).

It is therefore important to understand how organisations and people involved in the development of AI, especially in the Research and Innovation (R&I) processes, are considering and dealing with those threats and risks, especially the ethical, legal and social issues (ELSI). In fact, it is in those processes, at the earlier stage, that citizens, organisations, and other stakeholders, can help to avoid technologies failing, and ensure that their positive and negative impacts are better governed and exploited (von Schomberg, 2011).

The main objective of this paper is to understand which frameworks and instruments are available to help the organizations dealing with the ethical, legal and social issues (ELSI) in the development of AI. A second objective is to understand how practical and industry-wide these frameworks and instruments are to fit the profile of the more dynamic AI developing organizations. As a result, three research questions arise: 1) Which frameworks and instruments addressing the ELSI are available to increase the responsibility of the organizations involved in the design, development and deployment of AI technologies? 2) How practical and industry-wide are these frameworks and instruments to fit the profile of the more dynamic AI developing organizations, especially small ones, startups, scaleups and research centres? Assuming that existing frameworks

and instruments addressing ELSI regarding AI development are complex and difficult to implement, this paper proposes a new practical Industry-Wide Responsible AI Maturity Model, for an in-depth self-assessment and easy identification of key improvement actions and best practices to the deployment of Trustworthy AI. From that assumption, a third research question arise: 3) How organizations from several sectors and in different AI development maturity stages performed in the use of that Industry-Wide Responsible AI Maturity Model?

To address the third research question, the model was piloted in 3 companies of different sizes and types and 2 research centres, in Portugal. Because it is a country characterized by a recent wave of AI and innovation development (OECD, 2023, p.30) and has several unicorns (recent tech companies with a valuation of more than one billion US dollars), the authors have chosen Portugal as a case-study.

The paper is organized as follow. An overview of theories, approaches and recent developments related to the application of AI are presented in Chapter 2. Chapter 3 outlines the methodology adopted with a description of each step and a justification for the use of the selected research technique. The Maturity Model for Responsible AI is presented in chapter 4, including the Model itself and the Assessment. Chapter 5 presents and discuss the results and the case studies conducted in the organisations where the Maturity Model was piloted, including a cross-case analysis. Chapter 6 summarizes the key findings and the implications and offer perspectives for future developments and possibilities.



## 2 BACKGROUND

An overview of the most recent and relevant frameworks and instruments available to organisations involved in the processes of design, development and deployment of AI, regarding the trustworthiness of those processes, is presented in the following sub-chapters.

### 2.1. Responsible Research and Innovation

von Schomberg R. (2019, p. 9) defines Responsible Research and Innovation (RRI) as "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products".

In recent years, the European Union, via its Horizon 2020 programme, funded several projects to foster RRI in the industry sector. Among those projects is the PRISMA project[1]: Piloting Responsible Research and Innovation in Industry. Working with eight companies, most of them SMEs, the project conducted case studies of good practices in RRI and help those companies to better integrate it in their innovation processes and business practices.

To implement these innovation activities two different methods were applied. In some cases, it was applied the "external approach", which means external support must be found in academic or consulting organisations. The alternative is the "embedded ethicist approach" which is a specialist recruited by the company to takes responsibility for the RRI policy and the PRISMA project. In both approaches the PRISMA RRI roadmap implementation demands considerable commitment and resources, especially in SMEs and start-ups (PRISMA Responsible innovation in practice: experiences from industry, 2020, p. 8).

### 2.2. Responsible Artificial Intelligence

Responsible Artificial Intelligence (RAI) is specifically focused on the responsible development and use of AI. Those AI systems are called Trustworthy AI systems, which means it should be lawful, complying with all applicable laws and regulations, it should be ethical, ensuring adherence to ethical principles and values; and it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm (EU, 2019). In practice, RAI and RRI often involve similar practices and approaches, such as involving a diverse set of stakeholders in the development and deployment of technology, and ensuring that technology development and deployment is transparent, accountable, and respects privacy and human rights.

In the global AI governance, there are a plethora of principles and ethics codes applied to AI technologies namely the following:

[1] For more information on the PRISMA project: https://www.rri-prisma.eu/ (accessed February 2025)

- Ethics Guidelines for Trustworthy AI, from the High-Level Expert Group on Artificial Intelligence (HLEG-AI), European Commission, 2019.
- Artificial Intelligence Risk Management Framework, from the National Institute of Standards and Technologies (NIST), US Department of Commerce, January 2023.
- Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, The Institute of Electrical and Electronics Engineers (IEEE), 2017.
- Artificial Intelligence at Google: Our Principles, 2018.
- Everyday Ethics for Artificial Intelligence, IBM, 2019.
- Artificial Intelligence & Responsible Business Conduct. OECD, 2019.

The HLEG-AI Ethics Guidelines for Trustworthy AI is the most important code produced by the European Commission, and it was the basis for the EU AI Act. It was originally proposed by the European Commission on 21 April 2021, and politically agreed upon by all three EU institutions on 8 December 2023. The EU AI act was finally approved in May 2024. It starts with a definition of fundamental rights, then an identification of the ethical principles and their correlated values. It also lists the requirements for AI, according to Figure 1.


**Figure 1.** Realising Trustworthy AI throughout the system's entire life cycle, Ethics Guidelines for Trustworthy AI, from the High-Level Expert Group on Artificial Intelligence (HLEG-AI), European Commission, 2019, p20.

The seven requirements of the Ethics Guidelines for Trustworthy AI are: Human agency and oversight, Technical robustness and safety, Privacy and data governance, Transparency, Diversity, non-discrimination and fairness, Societal and environmental wellbeing, and Accountability.

In January 2023, the National Institute of Standards and Technologies (NIST) - US Department of Commerce, published the Artificial Intelligence Risk Management Framework (AI RMF 1.0), articulating the characteristics of trustworthy AI and offering

guidance for addressing them. The requirements are very aligned with the ones from EU Ethics Guidelines for Trustworthy AI.

More recently, in 2023, CEN issue the CEN/CLC/TR 17894 "Artificial Intelligence Conformity Assessment". This document sets out a review of the current methods and practices (including tools, assets, and conditions of acceptability) for conformity assessment as relevant for the development and use of AI systems. Among others, it addresses the conformity assessment for products, services, processes, management systems and organisations. It includes an industry horizontal (vertical agnostic) perspective and an industry vertical perspective.

Table 2.1 presents the results of matching the rights, principles, values and requirements for Trustworthy AI of the HLEG-AI with the main normative issues referenced in the other main codes.

**Table 2.1** Matching of HLEG-AI guidelines with the other most prominent global codes and principles for AI.

| HLEG-AI | OECD | IEEE | NIST AI | IBM | CEN |
|---|---|---|---|---|---|
| **Ethical Principles** | | | | | |
| Respect for human autonomy | Right to life | Human Rights | | | Human Oversight |
| Prevention of harm | Personal security | Awareness of misuse | Safeness | | Accuracy Cybersecurity |
| Fairness | | | Fairness | | |
| Explicability | | Transparency | Explainability and interpretability | Explainability | Transparency |
| **Key requirements for Trustworthy AI** | | | | | |
| Human agency and oversight | | Human Rights | | Value Alignment | Human Oversight |
| Technical robustness and safety | | Awareness of misuse | Security and resilience | | Controllability Robustness Accuracy Cybersecurity |
| Privacy and data governance | Privacy | Personal Data Rights and Individual Access Control | Privacy-enhanced | User Data Rights | Data Governance |
| Transparency | | Transparency | Transparency | | Transparency |
| Diversity, non-discrimination and fairness | Non-discrimination | | Fairness with harmful bias managed | Fairness | |
| Societal and environmental well-being | | Well-being promoted by Economic Effects | Safeness | | |
| Accountability | | Accountability | Accountability | Accountability | Monitoring, Record keeping through logging, Quality Management |

As Table 2.1 demonstrates, all the principles, guidelines, issues and requirements for ethics in AI, as referenced in the codes and principles referred to in this research, are included in the EU Ethics Guidelines for Trustworthy AI.

### 2.3 RAI Certification

Another important development was made by the Responsible AI Institute (RAII), based in the USA, with the development of conformity assessments and certifications for AI systems support practitioners. The Responsible AI Organizational Maturity Assessment, OMA (June 2022), was developed across five dimensions: Policy & Governance, Strategy & Leadership, Tools & Processes, People & Training, and Procurement Practices. The overall score obtained across all five dimensions translates to one of the 5 levels of RAI organizational maturity: Ad Hoc, Emerging, Tactical, Strategic, Transformative.

The OMA exercise includes a series of workshops and interviews, after which a Responsible AI Road Map is then proposed to guide implementation of the recommendations. In October 2022, the RAI Institute launched its RAII Certification Program based on the maturity assessment that evaluates AI systems, referenced above. Tested and fined toned during 2022 in a financial organisation case study, the Certification Program is tailored to specific industries and functions, the firsts are finance, health care, human resources and procurement.

Presumably, certification constitutes a demanding and complicated approach, with too many instruments, assessments, metrics and recommendations, different for each industry and function. It also demands a significant dedication of third-party consultancy to drive the work inside the organisations. For most organisations, at least in Europe, it makes sense to use a simpler approach, which is agnostic to all industries and functions, that allows for self-assessment and easy identification of key improvement practices, capabilities and competences.

## 2.4 Maturity Models in AI

The concept of measuring maturity was introduced with the Capability Maturity Model (CMM) from the Software Engineering Institute – Carnegie Mellon, and has expanded across a multitude of domains (Ellefsen A. et al, 2019). It become a popular way of evaluating maturity, used to assess the competency, capability and level of sophistication of a specific domain based on a more or less comprehensive set of criteria (de Bruin T. et al., 2005).

In 2017, to offer a structured way for companies to evaluate the degree to which its practices align with RRI, Stahl B. et al. propose the development of a RRI Maturity Model (Stahl B. et al., 2017). Resulting from the EU-funded ETICA project, RRI MM demanded considerable effort, as it included 30 semi-structured interviews, five bottom-up case studies, a large-scale Delphi study, 15 focus groups, and 4 in-depth case studies.

Stahl B. at al. considered that the development of a specific self-assessment tool is a natural next step, and one suitable way to provide support and guide individuals in industry to a deeper understanding of RRI.

Later, the research made by Schuster T. et al. (2021) on Maturity Models for the Assessment of AI (AIMM), identified 15 AIMM approaches but focused only on the three the authors classify as comprehensive and scientifically developed AIMMs. The other approaches are dismissed by Schuster T. et al. because they "have no empirical basis, lack documentation and can be understood as consulting offers of AI solutions to companies" (Schuster T. et al., 2021, p. 29). As a result, Schuster et al. propose their own Maturity Model, as they conclude they were unable to identify an AIMM that has been proved in practice and has already successfully passed through an evaluation phase.

Schuster et al. encompasses the Ethics and Privacy dimensions in its model, but in a rather simple way. For example, the pioneer level for the ethics dimension is described only with the following sentence: "the AI optimized data collection and structuring enables the standardized used of AI applications across companies based on a fully compliant

application of data protection principles and an ethical code of conduct" (Schuster T. et al., 2021, p. 32).

In resume, the work of Schuster et al. is relevant as it allows the understanding of the state of the art regarding several maturity models to assess AI status in SMEs, but none of these models assess specifically the ethical and responsibility perspectives.

In May 2023, Michael Mylrea and Nikki Robinson published the "AI Trust Framework and Maturity Model (AI-TFMM): Improving Security, Ethics and Trust in AI". AI-TFMM takes a holistic people, process, and technology approach (Mylrea M. and Robinson N., 2023):

- Technology: AI trust principles are documented through their lifecycle to be explainable (XAI), repeatable, interpretable, and transparent.
- People: someone is assigned/accountable to implement these principles through the AI project lifecycles.
- Process: the lifecycle and technology are tested.

A major challenge in this holistic approach is anticipated, as it is very unlikely that one organisation will achieve the same maturity level in the three vectors: Technology, People and Process, for the same domain. In fact, Mylrea M. et al. wrote in their conclusions "a holistic people, process and technology approach bolsters the contextual understanding of its application, but also introduces challenges for repeatability for different use cases" (Mylrea M. and Robinson N., 2023, p. 14).



## 2.5 Remarks

As described above, the most important ethic codes, frameworks and guidelines for Responsible AI are not simple to use for most of the organisations, especially the small ones, startups, scaleups and research centres. Table 2.2 presents a comparative analysis on the contribution of frameworks and instruments to address ELSI in the context of AI.

**Table 2.2** Comparative analysis on the contribution of frameworks and instruments to address ELSI in the context of AI.

| **Frameworks and instruments to address ELSI in the context of AI (main references)** | **Summary of the reasons why they are not sufficiently practical and simple to be used for SMEs, startups and scaleups and research centers** |
|---|---|
| **TA**<br>(Rip, 2018)<br>(Grunwald, 2009) | Only provided by experts.<br>The idea that technological developments can be predicted by extrapolation or other means is over-simplistic. |
| **RRI**<br>(Schomberg, 2019)<br>(van de Poel et al., 2017)<br>(Owen et al., 2021)<br>(European Commision, 2020) | PRISMA RRI roadmap implementation demands considerable commitment and resources, especially in SMEs and start-ups (PRISMA Responsible innovation in practice: experiences from industry, 2020, p. 8). |
| **The EU Ethical Guidelines for Trustworthy AI**<br>(European Commission, 2019) | Over-long assessment with 133 questions and does not offer a roadmap for improvement. |
| **RAII Organizational Maturity Assessment**<br>(Responsible AI Institute, 2022) | Do not highlight Human Agency and Oversight, nor Privacy and Data Governance. |
| **RAII Certification**<br>(Responsible AI Institute, 2022) | Is tailored to specific industries and functions and is not industry-wide. |
| **Stahl et al. RRI Maturity Model**<br>(2017) | Demands considerable effort, as it includes semi-structured interviews, case studies, a Delphi study, and focus groups.<br>Does not include a self-assessment tool. |
| **Schuster et al. Maturity Models in AI**<br>(2021). | It does not assess specifically the ethical and responsibility perspectives. |
| **Ellefsen et al. AI maturity model framework**<br>(2019) | Not oriented to help organizations with how responsible they are when using AI in their activities. |
| **AI Act**<br>(European Commission, 2024) | How will the instruments identified be designed to take into consideration the perspectives of the SMEs, startups and research centres? |

Even the EU Ethics Guidelines for Trustworthy AI, the most important code produced from the European Commission, needs simplification. In fact, for the AI industry

stakeholders, who's time is scarce, an assessment with 133 questions is too long, and the guidelines do not offer a simple framework for organisations to improve. There is a clear need to develop a simpler and easier self-assessment tool to fill in this gap.

The present research concluded that maturity models are an effective way of evaluating maturity, to assess the competency, capability and level of sophistication of a specific domain, Responsible Artificial Intelligence in this case.

An Industry Wide Maturity Model for Responsible AI, inspired by the EU Guidelines for Trustworthy AI and other principles and codes of conduct, allowing for self- assessment and easy identification of key improvement practices, capabilities and competences, was developed aiming for a practical and simpler tool, to fill that gap.

## 3. METHODOLOGY

The objective of the present research is to assess the frameworks and processes available to the organisations involved in the design, development and deployment of AI technologies, to address the values, needs and expectations of society regarding its trustworthy, and the way those instruments are suitable to them. Additionally, it also aims to fulfil the eventual gap of those instruments.

To better frame the knowledge on the way organization deal with ELSI values and threats in the AI development processes, an online questionnaire with closed-ended questions using a four Likert scale combined with and open question for comments was used. The Likert scale is one of the most common tools used for measuring attitudes in the social sciences, created by the sociologist Rensis Likert in 1932. It is a type of psychometric scale used in questionnaires to measure a person's preferences, degree of agreement, or a wide range of attitudes, such as satisfaction, importance or likelihood of behavior Tanujaya B. et al. (2023). A Likert scale typically includes a series of statements or questions, each with

a set of response options, such as “strongly agree”, “somewhat agree”, “neutral”, “somewhat disagree”, and “strongly disagree”.

Tanujaya B. et al. (2023) states that with an odd number of response options, the middle option (i.e., neither agree nor disagree) has an ambiguous meaning. They state that this could increase measurement error if respondents use that option in ways that do not reflect their perceived standing on the characteristic being measured.

Sometimes a 4-point (or other even-numbered) scale is used to produce an ipsative (forced choice) measure where no indifferent option is available (Bertram, 2007). An example of 4-point Likert scale uses in social science and attitude research projects is the one developed by Pornel and Saldaña (2013) to measure teachers’ attitudes towards research. Hence, the authors use a 4-point Likert scale with the following response options: “strongly agree”, “somewhat agree”, “somewhat disagree” and “strongly disagree”.

The questionnaire was developed in the SurveyMonkey platform and the link was sent via email to 19 respondents in 3 companies and 2 research centres, in Portugal.

In addition, to corroborate and complement information received from the previous mentioned questionnaire, group interviews were conducted in person. The online surveys were completed between June and November 2023, and the face-to-face group interviews / case studies occurred between September 2023 and January 2024 in several locations in Portugal.

The primary goal of the proposed Responsible AI Maturity Model is to enhance maturity levels and formulate a strategic roadmap for organizations. Thus, the aim is to promote a positive impact on the way companies design and implement AI systems, responsibly.

For the development of the Maturity Model, Bruin, T. et al. (2005) framework was adopted, following the steps: Step, Design, Populate, Test, Deploy and Maintain.

In the initial phase the scope or the focus of the maturity model is defined: Responsible Artificial Intelligence. Additionally, the Development Stakeholders are identified. These are the stakeholders for whom the model was developed and tested and the most relevant

organizations evolved in the development of AI in Portugal: the industry, large companies, SMEs and scale-ups (former startups that achieve a certain size) and research centres.

The second phase is the design and architecture of the model, which forms the basis for further development and application. To design the model is necessary to define how maturity stages can be reported to the audience. In this step the maturity is represented as a series of one-dimensional linear stages, a widely accepted form that has formed the basis for assessment in many existing tools. Regarding stages, for simplicity a model of four levels was used, namely: Unaware, Exploratory/reactive, Proactive and Strategic.

Select intuitive, clear and convincing levels is key to a good maturity model, by opposition of defining continuous boundaries between, rather than discrete ones (Stahl B. et al., 2017). Is important that the final stages are distinct and well-defined, and that there is a logical progression through stages that should be named with short labels that give a clear indication of their intent.

The third phase is to Populate the model. In this step is determined how maturity measurement can occur i.e., the inclusion of appropriate questions and measures within this instrument. To measure maturity, this research developed an assessment with a total of 57 questions. Those questions are related to the seven requirements from the EU Guidelines for Trustworthy AI (Human agency and oversight; Technical robustness and safety; Privacy and data governance; Transparency, Diversity, Non-discrimination and fairness; Societal and environmental well-being; and Accountability, and his twenty sub-requirements.

When the EU Guidelines for Trustworthy AI was first published, it does not reference potential threats created by Generative AI tools, such as ChatGPT. To address those issues, this research adds one additional sub-requirement named “Respect and fairness regarding IP and copyright” under the “Societal and environmental well-being” requirement.

Finally, to offer a roadmap for improvement, Methods to ensure Trustworthy AI and key-practices are generated by the Maturity Model developed. The next chapter, will present in detail all the components of the Maturity Model for Responsible AI, developed in this research.

The next phase is related to the steps Collect Data, Analyse Results and Interpret results of the Scientific Research Method.

Phase four addresses the test of the model developed by the stakeholders. The process used to test and implement the model in the organizations took the following steps.

- A 20-minute remote video call with all the respondents lead by the researcher to brief them about the context and objectives of the exercise.
- After receiving an invitation via email, the respondents fill the online survey at the SurveyMonkey application. It is important to note that respondents answer the survey without knowing the algorithm that determines the maturity level in each requirement, not to be influenced by that.

  Respondents can skip the statements they do not want to answer and could write comments at the end of each requirement statement block. Most of the respondents took between 30 and 40 minutes to fill the survey. Some of them did it for several days and the fastest took only 8 minutes.
- Once all the answers for a specific organization were done, the author calculated the RAI Maturity result using an Excel spreadsheet, including a radar diagram and produced the final report.
- For each organization, a one-and-a-half-hour case study face-to-face meeting was conducted, with all the respondents. Those meetings followed a strictly structured agenda:
    - High level presentation of the AI development projects in the organization, for the researcher to better understand the context.
    - How clear is the Assessment for RAI?

    - Discussion.
    - Use a 1 to 4 scale: “very clear”, “somewhat clear”, “somewhat not clear”, “not clear”.
- Presentation and discussion of the RAI Maturity Report.
- How clear is the RAI Maturity Model for Responsible AI for the organization?
    - Use a 1 to 4 scale: “very clear”, “somewhat clear”, “somewhat not clear”, “not clear”.
- Are the different hierarchical perspectives clear?
- Is the organization willing to do the exercise again? With which frequency?

Finally, a Case Study Report was prepared for each organization, to document the conclusions.

First the RAI MM was pre-tested in two organizations, one large research centre and an early-stage startup, both working in AI in Portugal. Based on the results of the pre-test, the model was tuned, complemented, and sophisticated in terms of is comprehensiveness and readiness for application. In particular, several survey statements were re-written, the YES/NO answer were replaced by a Likert scale and several key actions of the improvement roadmap were re-written.

The improved model was then implemented in a structured way in five other organizations in Portugal, chosen among the most relevant ones, evolved in the development of AI in Portugal, one large size private company, one medium size company (SME), one scale-up company, means a fast-growing tech company founded less than 10 years ago, and two research centres. Those organizations were selected among a list of ten ones, including large companies and public organizations that demonstrate less interests in participating, because they do not have relevant AI development experience. The strategy employed was a nonprobability sampling method,

selected on the basis of convenience. This approach entailed the selection of a purposive sample, with a relatively small number of decision-makers chosen to provide information of particular relevance to the research questions under examination (Tashakkori and Teddie 2009).

Participants were selected on the basis of their role and involvement in the decision-making process of the SME or start-up.

# 4. The Industry-Wide Maturity Model for RAI

## 4.1 Structure of the Maturity Model for Responsible AI

The Maturity Model for RAI structure, inspired by the Capability Maturity Model (CMM) are organised by Requirements for Trustworthy AI, these are related to Methods to Ensure Trustworthy AI, which contain Key Practices.

The proposed model adopts a 4-level model for reasons of simplicity and the relative novelty and complexity of the RAI subject. Table 4.1 presents the level definitions, inspired at Ellefsen et. al. (2019) and Stahl et al. (2017).

**Table 4.1.** Maturity Model stages adopted in this research.

| **Stage name** | **Level definition (for each sub-requirement)** |
|---|---|
| Unaware: | The organisation is not aware of RAI or its components and does not incorporate it in its processes. |
| Exploratory/reactive: | The organisation has a reactive response to external pressures concerning aspects of RAI. |
| Proactive: | The organisation realises the benefits of RAI and increasingly integrates these into its business processes. |
| Strategic: | The organisation has adopted RAI as a component of its strategic framework and aims to ensure that all R&D activities consider the RAI components. |

The Maturity Model includes the seven requirements inspired by the EU Ethics Guidelines for Trustworthy AI (2019), presented in chapter 2. One new sub-requirement, "Respect and fairness regarding IP and copyright", was created under the "Societal and environmental well-being" requirement, to address the issues arising from the Large Language Models or Foundation Models that surged into hype by November 2022, with the public launching of ChatGPT.

The requirements for Trustworthy AI and sub-requirement are the following:

- Human agency and oversight address the sub-requirements Human agency and autonomy and Human oversight.
- Technical robustness and safety deals with Resilience to attack and security, General safety, Accuracy and Reliability, fall-back plans and reproducibility.
- Privacy and data governance includes Privacy and Data Governance.
- Transparency addresses the sub-requirements Traceability, Explainability and Communication.
- Diversity, non-discrimination and fairness includes Unfair bias avoidance, Accessibility and universal design and Stakeholder engagement.
- Societal and environmental well-being addresses the sub-requirements Environmental well-being, Impact on work and skills, Impact on society at large or democracy and Respect and fairness regarding IP and copyright.
- Accountability includes Auditability and Risk management.

The model provides several technical and non-technical methods that can be employed to implement the requirements to ensure Trustworthy AI. Technical methods focus on developing and implementing specific technical mechanisms or algorithms to ensure trustworthy AI, such as Architectures for Trustworthy AI and Testing and Validating. Non-technical methods are broader in scope and encompass various organisational, legal, and societal measures such as regulation to govern the use and deployment of AI systems.

For each sub-requirement and depending on the maturity level, the model presents several key actions and related best practices that could be implemented and institutionalised to allow the organisation to evolve to the next maturity level and effectively minimise the risks while maximising the benefit of AI.

For example, for the requirement "Respect and fairness regarding IP and copyright", the following key actions and best practices will be suggested depending on the maturity level achieved by the organisation in this requirement:

- Unaware Level: Your organisation did not develop real awareness of the impact of the Al system in the eventual infringement on copyrights, trademarks, patents, and other intellectual property (IP) rights. To improve, the organisation must increase specific competencies and establish a strategy for this requirement.
- Exploratory Level: Your organisation developed some awareness related to the impact of the AI systems in the eventual infringement on copyrights, trademarks, patents, and other IP rights, but has not taken proactive actions to evaluate it. To improve, the organisation should implement the following processes:
    - Implemented mechanisms to assure that IP sources used to train the current AI system are always quoted and credited with a fair economic value (If this is the case).
    - Disclose any copyrighted material used to train the models.
- Proactive Level: Your organisation, to a reasonable degree, assesses the impact of the AI system in the eventual infringement on copyrights, trademarks, patents, and other IP rights. To improve, the organisation should implement the following processes:
    - Implemented mechanisms to assure that IP sources used to train the current AI system are always quoted and credited with a fair economic value (If this is the case).
    - Disclose any copyrighted material used to train the models.

- Strategic Level: Your organisation systematically assesses the impact of the AI system in the eventual infringement on copyrights, trademarks, patents, and other IP rights. The organisation should consolidate a continuous improvement cycle in what concerns this requirement.

### 4.2 The Responsible AI Assessment

As explained in the Methodology Chapter, the main method used for collecting data was through an online survey. A Responsible AI Assessment with 57 entries was developed, using the Assessment List of Trustworthy AI (ALTAI) from the HLEG-AI, the Ethical OS toolkit checklist, and others as an inspiration guideline to populate this RAI MM for each of its 7 requirements and sub-requirements.

It was adopted a 4-point Likert scale with the following response options: "strongly agree", "somewhat agree", "somewhat disagree" and "strongly disagree".

Valid answers do not include questions to which there was no response.

## CHAPTER 5 – RESULTS AND DISCUSSION

### 5.1. Results in the Organizations

This section presents and discuss the maturity levels achieved by each of the 5 organisation piloted, the comments made by the respondents, and the observations made during the case-study meetings.

Company One

Company One is a large corporation and part of a multinational group established in 1990. It develops IT solutions for the financial services sector, employing 650 persons, with €50 million revenue in Portugal. It is in a process of piloting the first AI development projects, trying to productise some of those experiences such as:

- Data Machine Learning.
- RPA – process automation.
- Voice interaction and chatbot.
- Predicting "next best action", which means helping the salespersons by suggesting the best financial product to promote to the client.
- Clustering clients, meaning to explore similitudes using data science.

This organization enrolled three respondents in the Maturity Model exercise: CO1, the Chief Technical Officer and member of the Board of Directors; CO2, the Technical Team Leader, and CO3, a Data Analyst.

Company One achieves the Exploratory level in 9 requirements and is Not Aware in the other 10 requirements. In the Traceability requirement, the organisation reaches the Proactive level. Figure 5.1 shows a radar diagram as a graphical method of displaying multivariate data representing the maturity levels achieved for Company One. The far away from the centre the more maturity the organization shows.

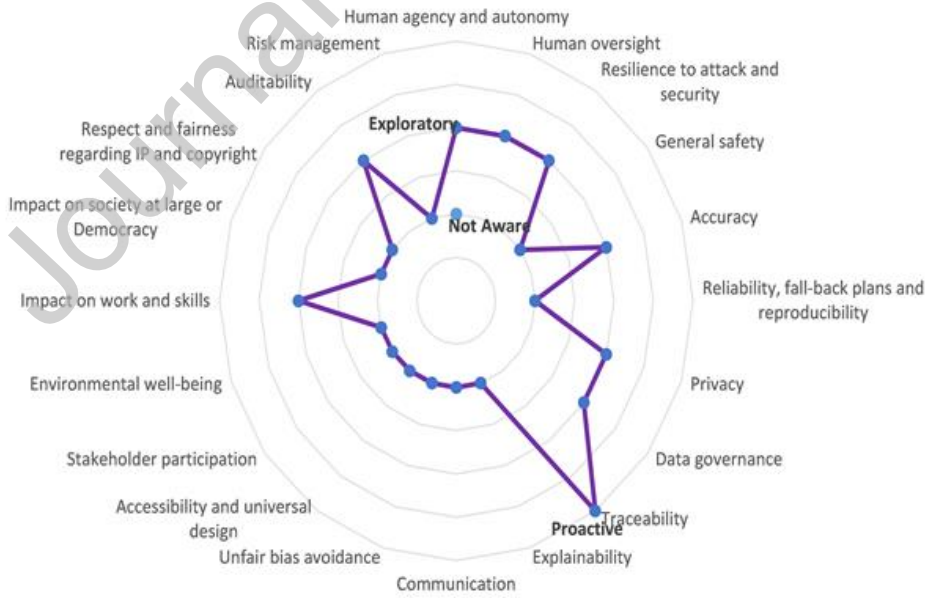


**Figure 5.1.** Radar diagram representing the maturity levels achieved for Company One

All the respondents of Company One wrote abundant comments which substantially enrich the report:

- One responded that the organisation is Not Aware of the all the 20 sub-requirements and respective sub-sections, and responds "Somewhat Agree" in some statements, writing in all the comments, "we are just taking the first steps in AI, we haven't reached this stage yet".
- The other respondents seem substantially aligned in most of the statements.
- One responds "Agree" and "Strongly Agree" with most of the sentences and thus considers the organisation Aware of most of the requirements' sub-sections. This is not the case for the sub-sections Unfair bias avoidance, Accessibility and universal design, Impact on work and skills, Impact on society at large or Democracy, and Risk management, where this respondent considers consider the organisation Not Aware.
- Other responds "Strongly Agree" to all the statements in the "Accessibility and universal design section", but considers the organisation Not Aware in this requirement, which seems contradictory.

Company Two

Company Two is a medium-size corporation, established in 2002, employing 130 people, with approximately €10 million revenue/year. The field of activity is industrial automation, industrial machines for manufacturing industries.

Regarding AI projects, Company Two develops quality inspection systems using vision-based technologies and data processing for machine learning. They enrolled four respondents in the Maturity Model exercise: CT1, the Chief Executive Officer (CEO); CT2, the Innovation Manager; CT3, a Senior Researcher, and CT4, the Innovation Project Manager.

Company Two reaches the Proactive level in 4 requirements, the Exploratory level in 11 requirements and the Not Aware level of the Explainability requirement. The requirements Privacy, Data Governance and Unfair Bias Avoidance are considered Not Fully Applicable in this research because Company Two's AI systems do not collect any information about users. Also, the Accessibility and Universal Design requirement is Not Fully Applicable to Company Two because its AI systems are not to be used by all people, but rather by specialised and full trained personnel.

Figure 5.2 shows a radar diagram representing the maturity levels achieved for Company Two.


**Figure 5.2**. Radar diagram representing the maturity levels achieved for Company Two

Statements 4.3.1, 6.3.2 and 6.4.3 are considered not fully applicable to Company Two and are thus removed for the calculation:

- 4.3.1 In cases of interactive AI systems (e.g., chatbots, robo-lawyers), my organisation communicates to users that they are interacting with an AI system instead of a human.

- 6.3.2 My organisation takes measures that ensure that the AI system does not negatively impact democracy, if that system could be used, for instance, to allow for fraud or generate or spread misinformation to create political distrust or social unrest.
- 6.4.3 Procedures are being taken by my organisation to reduce the prevalence of the content, if there is potential for toxic materials like conspiracy theories and propaganda to drive high levels of engagement.

The results in general did not show substantial differences between the respondents, although the most senior is clearly more optimistic regarding the way the company performs in the context of RAI.

The exception is one respondent who skipped a substantial number of statements, explaining that "The systems developed at Company Two use artificial intelligence in automation tools that do not have direct contact with humans, so there are some questions that do not fit the type of product", which seems correct. However, it was explained during the case-study face-to-face meeting that in this context end-users means workers that will use the automation tools.

In the "Risk Management" requirement, two respondents skip all the questions and one respondent responds "Somewhat Disagree" to the 3 questions. Despite this, the overall result is 61% for "Exploratory", meaning that more awareness is recommended to this requirement.



Company Three

Company Three is a medium-sized corporation, with 350 employees, founded in 2013. They don't disclose their annual revenue. It is a fast-growing former start-up, or so called "scale-up". It provides custom computer programming services, a LangOps platform that combines the best blend of machine and human translation to provide a consistent

multilingual customer experience, grow to new markets and build trust around the world. Its main AI developments in the projects are:

- Machine translation - builds and trains customised engines on custom data sets to perform translation within a specific industry or even a specific brand.
- Natural Language Processing (NLP) tasks - builds and trains technology to perform tasks such as named entity recognition, anonymisation and localisation.
- Quality evaluation - builds and trains engines that can predict the quality of translations generated by humans or machines.
- Lang Ops is a centralised platform to allow a company to centralise all their translation needs.

Company Three enrolled four respondents: CD1, the Director of Legal and Compliance; CD2, the Chair of the Ethics Committee; CD3, a backend engineer and Engineering manager, and CD4, the Head of Product.

Company Three fits the Strategic level in 7 requirements, the Proactive level in another 8 requirements, and the Exploratory level in 5 requirements; the Not Aware level is not achieved in any requirement.

The following statement is considered Not Fully Applicable to Company Three and so is removed from the calculation: 1.1.3 "If the AI system could create risk of human attachment, stimulate addictive behaviour, or manipulate user behaviour, my organisation takes measures to deal with possible negative consequences for end-users or subjects in case they develop a disproportionate attachment to the AI System".

Figure 5.3 represents the maturity levels achieved for Company Three in the form of a radar diagram.


**Figure 5.3**. Radar diagram representing the maturity levels achieved for Company Three.

Regarding differences between respondents, one of them skipped 35 out of 57 statements of the survey and wrote that they are not aware if the organisation does what is written in the statements. In the case study face-to-face meeting, this person said they are new to the organisation and skipped those statements because they are not certain how the organisation performs in those questions.

Other responds "Strongly Disagree" in the following statements, mostly in opposition to the other respondents, justifying this by their position as Director of Legal and Compliance, which "demands consider something implemented only if it is formally written and controlled":

- "My organisation has in place a policy for what happens to customer data if your company is bought, sold, or shut down".
- "In cases of interactive AI systems (e.g., chatbots, robo-lawyers), my organisation communicates to users that they are interacting with an AI system instead of a human.
- "My organisation communicated to users the technical limitations and potential risks of the AI system, such as its level of accuracy and/or error rates".

- "My organisation tests diversity and representativeness of end-users for specific target groups or problematic use cases or subjects in the data, like instances of personal or individual bias".
- "My organisation assesses whether there could be groups who might be disproportionately affected by the outcomes of the AI system".
- "My organisation implemented mechanisms to assure that IP sources used to train the current AI system are always quoted and credited with a fair economic value (If this is the case)".
- "My organisation discloses any copyrighted material used to train the models".

The other two respondents mostly are aligned in their responses.

Research Centre One

Research Centre One was established in 1992 and employs 60 PhD researchers, plus 50 PhD students plus 70 Master students. It does R&D in Robotics, Computer Vision, Artificial Intelligent Systems, Cognitive Systems, Energy and Sustainability. The research group evolved in this research was the Visual Information Security - VIS team. Their main AI project is the design, implementation and building of tools applied to the field of biometrics, namely in Facial Recognition, Presentation Attack Detection, Morphing Attack Detection, Biometric Template Protection, Standard and ICAO Requirements Analysis and Verification, Synthetic Realities and others.

This Research Centre enrolled five respondents in the Maturity Model exercise: RO1, the Research Group Leader; RO2, a research associate and project manager; RO3, a Liveness Detection model developer; RO4, a Researcher, and RO5, a Researcher in Machine Learning, Vision and Graphics.

Research Centre One reaches the Proactive level in 6 requirements and the Exploratory level in 14 requirements. Strategic level and Not Aware level were not achieved in any requirement.

The following statements are considered Not Fully Applicable to Research Centre One due to the nature of the facial recognition AI technologies it develops, and is thus removed for the calculation:

- (1.1.3) If the AI system could create risk of human attachment, stimulate addictive behaviour, or manipulate user behaviour, my organisation takes measures to deal with possible negative consequences for end-users or subjects in case they develop a disproportionate attachment to the AI System.
- (4.3.1) In cases of interactive AI systems (e.g., chatbots, robo-lawyers), my organisation communicates to users that they are interacting with an AI system instead of a human.
- (6.3.1) My organisation assesses the societal impact of the AI system's use beyond the end-user, such as potentially indirectly affected stakeholders or society at large, for example to spread hate or spread ransomware.
- (6.3.3) Procedures are being taken by my organisation to reduce the prevalence of the content, if there is potential for toxic materials like conspiracy theories and propaganda to drive high levels of engagement.

Figure 5.4 represents the maturity levels achieved for Research Centre One in the form of a radar diagram.


**Figure 5.4**. Radar diagram representing the maturity levels achieved for Research Centre One

In general, there is no substantial difference between the respondents in Research Centre One. Despite that, the most senior respondents seem to have a more optimistic view regarding the organisation actions, than the less senior respondents.

One respondent considers the organisation is Not Aware of the following requirements, but it seems they consider the statements not fully applicable to Research Centre One:

- Human agency and autonomy.
- Human oversight.
- Resilience to attack and security.

One respondent skipped the following three statements in the section "Impact on society at large or democracy".

- 6.3.1 My organisation assesses the societal impact of the AI system's use beyond the end-user, such as potentially indirectly affected stakeholders or society at large, for example to spread hate or spread ransomware.

- 6.3.2 My organisation takes measures that ensure that the AI system does not negatively impact democracy, if that system could be used, for instance, to allow for fraud or generate or spread misinformation to create political distrust or social unrest.
- 6.3.3 Procedures are being taken by my organisation to reduce the prevalence of the content, if there is potential for toxic materials like conspiracy theories and propaganda to drive high levels of engagement.

When questioned about the section "Impact on society at large or democracy", during the case-study face-to-face meeting, respondents debated and concluded that the AI technologies developed by Research Centre One contribute decisively to social security and positively impact democracy.

Some of the respondents consider Risk Management statements to be Not Applicable to the organization.



Research Centre Two

Research Centre Two was established in 1993 and employs 230 permanent researchers, 340 temporary researchers, mostly PhD students and post-docs, with approximately 12 million €/year budget.

It does fundamental and applied research in telecommunications and related areas. Regarding AI development it is focused on communication networks, multimedia communications, signal processing, electronic design, and many others.

Research Centre Two enrolled three respondents: RT1, a Senior Researcher and the Coordinator of the Information and Data Sciences Thematic Area, working in medical imageology and satellite remote sensing; RT2, a Senior Researcher in Biomedical Instrumentation, Signal Processing, and Knowledge Extraction, doing human gestures classification; RT3, a Senior Researcher.

Research Centre Two performs at the Exploratory level in 10 requirements and at the Proactive level in another 9 requirements.

In the Risk Management requirement, the organisation reaches a little above the Not Aware level, but since two of the respondents expressly wrote that the organisation are Not Aware, it is registered at that level.

Figure 5.5 represents the maturity levels achieved for Research Centre Two in the form of a radar diagram.

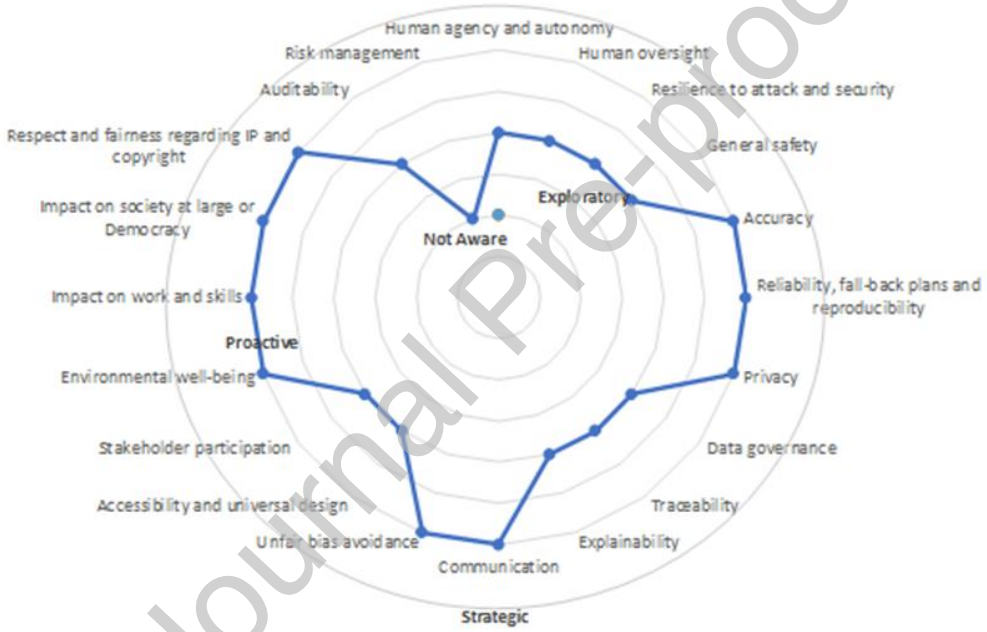


**Figure 5.5**. Radar diagram representing the maturity levels achieved for Research Centre Two.

There are no substantial differences in the results between the three respondents, except in the following statements, where two of the respondents differ substantially:

- 2.2.2 My organisation aligned the reliability/testing requirements of the AI system with the planned levels of stability and reliability.

- 4.1.2 My organisation implemented mechanisms to trace back which model or rules led to the decision(s) or recommendation(s) of the AI system.
- 4.2.2 My organisation implemented the following techniques to improve the explainability of the AI system such as model interpretability (saliency maps and layer-wise relevance propagation), fairness and bias assessment, causal inference or counterfactual analysis.
- 6.2.3 My organisation provides training opportunities and materials for re- and up-skilling.
- 6.3.4 My organisation implemented mechanisms to assure that IP sources used to train the current AI system are always quoted and credited with a fair economic value (If this is the case).

### 5.2 Results from the Case Studies meetings

Results from the face-to-face group meetings and case studies discussions regarding the clarity of the Assessment, the clarity of the Maturity Model Report and the organisation's willingness to repeat the exercise in the future are presented in Table 5.1:

**Table 5.1**. Feedback from the respondents on clarity

| | Company One | Company Two | Company Three | Research Center One | Research Center Two |
|---|---|---|---|---|---|
| **Feedback regarding how clear is the Assessment for RAI** | | | | | |
| **Very clear** | 2 | 2 | 3 | 3 | 3 |
| **Somewhat clear** | 1 | 1 | 1 | 1 | 0 |
| **Somewhat not clear** | 0 | 1 | 0 | 1 | 0 |
| **Not clear** | 0 | 0 | 0 | 0 | 0 |
| **Feedback regarding how clear is the RAI Maturity Report** | | | | | |
| **Very clear** | 3 | 3 | 4 | 5 | 2 |
| **Somewhat clear** | 0 | 1 | 0 | 0 | 1 |
| **Somewhat not clear** | 0 | 0 | 0 | 0 | 0 |
| **Not clear** | 0 | 0 | 0 | 0 | 0 |
| **With which frequency, in months, is the organisation willing to repeat the exercise?** | 12 | 36 | 12 | 18 to 24 | 24 |



Regarding the assessment, the results show that some respondents found it difficult to understand some of the statements, either because some of the concepts are somewhat new for them, or they did not have the opportunity to go through the glossary, although a link to it was provided at the beginning of the assessment. No relevant differences in the score were obtained in the different organisations, not even when comparing companies with research centres, which may be surprising due to the comments from research centre respondents.

The RAI Maturity Report was unanimously considered to be Very Clear or Clear. Seventeen out of nineteen respondents consider the report Very Clear and the other two respondents consider the report Somewhat Clear. This is because an oral presentation of

the report and a presentation discussion were held, where the respondents had the chance to clarify doubts.

Regarding the frequency with which the organisation is willing to repeat the exercise in the future, the results reflect the priority of AI development in each organisation.
Company Three expresses willingness to repeat the exercise in 12 months because AI development is at its core. Company One expresses the same because it wants to speed up their AI development. For Company Two, AI development is not so core, so they select 36 months. Research Centres willingness to repeat the exercise in the future lie between those periods, choosing to repeat the exercise in eighteen to twenty-four months in the future.
All the organisations confirmed the intention to start implementing some of the recommendations made.

### 5.3 Discussion

This section discusses the way the five organisations and the respondents performed towards the Maturity Model for RAI.
As expected, the five organisations result present substantial differences between them. Figure 5.6 presents a comparative bar chart to show the maturity levels achieved by each of the five organisations in each sub-requirement.

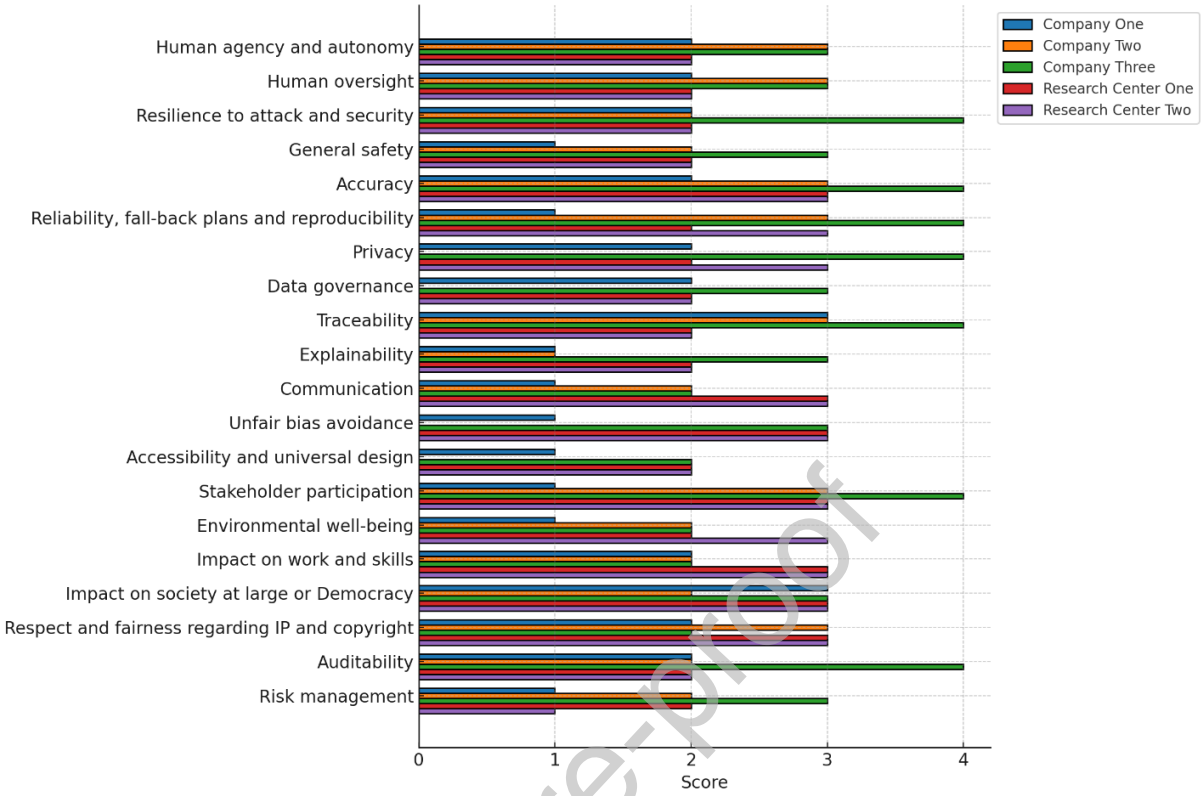


**Figure 5.6**. Comparative bar chart on the maturity levels achieved by each of the five organisations

Regarding how the size and type of the different organisations affects the way they deal with the development of Trustworthy A, the results show substantial differences between the organisations, but it is possible to understand some patterns. The more the company is dependent on AI for its business, or when AI is the core business of the organisation, the more RAI maturity model it shows. The opposite seems also true.

Company Three presents higher maturity in most of the requirements and is the only Company to achieve Strategic levels, in seven out of the 20 requirements. This result is expected as this is the only organisation fully dedicated to developing AI solutions, in this case for the machine translation global market, which means that they did not have other business than AI development.

Company Two shows exploratory and proactive levels in most of the requirements. This is also expected as this Company does a very contained and specific use of AI

development, in this case quality inspection systems using vision-based technologies and data processing for machine learning.

Company One shows mainly exploratory maturity and is Not Aware of a significant number of RAI requirements. This is also expected, as this organisation is a traditional software development company, starting to pilot its first experiences using AI.

Finally, we have the Research Centres, which showed similar, mostly exploratory and proactive results. Research Centre One achieves Exploratory level in 13 requirements and Proactive level for 7 requirements. For this organisation, the requirement of Privacy was considered Not Fully Applicable because in the AI systems they develop no user data is stored, at least in liveness detection.

Research Centre Two achieves Exploratory level in 10 requirements, Proactive level in 9 requirements and is Not Aware in the Risk Management requirement.

The results put Research Centres maturity in quite a challenging position: they are clearly aware of all the requirements, but they need to improve significantly to reach Strategic level.

Overall, in 44% of the requirements, the organisations showed a reactive response concerning several aspects of RAI, corresponding to the Exploratory/Reactive maturity level, which means that organisations are mostly aware of those RAI requirements and are starting to implement processes to deal with it.

For 33% of the requirements, the organisations showed a Proactive response, seeming to realise the benefits of RAI and increasingly integrate these into their business processes. This means that in those cases, the organisations have a proactive attitude toward Ethical, Legal and Social Issues.

The requirements where the organisations demonstrated more maturity are Accuracy, Traceability, Stakeholder Participation and Impact on society at large or democracy. The first two are more technical and related to the concept of quality, which is usually well-

mastered in organisations dealing with IT. The last two requirements are more aspirational, which means organisations seem to understand that establishing enduring mechanisms for stakeholder engagement and incorporating their regular feedback is valuable to foster the development of Trustworthy AI, and that people in those organisations have a strong societal perspective.

The requirements where the organisations demonstrated less maturity are Explainability, Accessibility and Universal Design and Risk Management. This conclusion is not a surprise since explainability is a relatively novel and complex issue, especially in AI systems. Accessibility and Universal Design is relevant mainly in organisations developing solutions for consumers, which is not the case in all the organisations studied. Finally, Risk Management is an especially demanding requirement in this Maturity Model.

Research Centres respondents showed more difficulties in the assessment, expressing that they "do not interact with final users", and that "they investigate and create pieces of technology that would eventually be part of commercial products used by final users". This is especially common in the following statements of the assessment:

- (1.1.1) My organisation made the end-user aware that a decision, content, advice or outcome of the AI system is the result of an algorithmic decision (Human agency and autonomy requirement).
- (3.2.3) My organisation has in place a policy for what happens to customer data if your company is bought, sold, or shut down (Data Governance requirement).

Even so, they perform at Exploratory level for 55% of the requirements and at Proactive level in 35% of the requirements, well above the corporations.

Regarding how the hierarchical level and type of functions of the people involved in AI development processes affect the way they deal with ELSI values and threats, is not obvious to establish a pattern. Eventually the most senior respondents seem to have a

more optimistic view regarding the organisation's actions. Other potential reasons for differences are:

- The nature of the function of the respondent: legal functions are more demanding because they are bound to formalities? Technical functions sometimes are less aware of how others perform in the organisations?
- The culture of the respondents: more rigorous vs more tolerant with formal rules?

Seventeen out of nineteen respondents said that the RAI Maturity Report was very clear and useful, and all of them are willing to repeat the exercise in the future. All of them confirmed the intention to start implementing some of the recommendations made and to repeat the exercise in the future.

## 6. CONCLUSIONS

Competition in AI technologies is at its fiercest, pushing companies to move fast and sometimes to cut corners when it comes to risks to human rights and other societal impacts. Without simple methodologies and widely accepted tools, it is difficult for organisations to adopt a safe pace on how to develop and deploy AI in a trustworthy way. This paper presents the most relevant frameworks and instruments available to the organisations involved in the processes of design, development and deployment of AI technologies, regarding the trustworthy of those processes. Drained from the literature review, a comparative analysis on the contribution of frameworks and instruments to address ELSI in the context of AI, from several authors (von Schomberg R., Shuster, T., Responsible AI Institute, European Commission, etc.) was concluded the presumably complexity and time consumption that the more dynamic organizations face to utilize those frameworks. To fill that gap, this paper presents a practical and simpler tool, an Industry Wide Maturity Model for Responsible AI and discusses a pilot in 3 companies

of different sizes and types and 2 research centres, in Portugal. Results show that organizations are aware of requirements (44%) to deploy a responsible AI approach and have a reactive response to its implementation, as they are willing to integrate other requirements (33%) into their business processes. The proposed Model was welcomed and showed openness from companies to consistently use it, since it helped to identify gaps and needs when it comes to foster a more trustworthy approach to the development and deployment of AI.

The results demonstrate that AI practitioners that went through the Maturity Model for RAI presented in this paper, gained a better understanding of how their developments and processes could impact the stakeholders that will use their innovations.

In the context of the rules imposed by the EU AI Act, the present Maturity Model for Responsible AI could be a useful tool as it was specifically developed to be practical and industry-wide and oriented to SMEs, start-ups and organizations of a kind. The present Maturity Model could be an exemplar starting point to the development of such conformity assessments, codes of conduct and governance mechanisms, using the self-assessment developed in this research, to evaluate the level of maturity regarding the requirements demanded by the AI Act and then use the Technical and the Non-Technical Methods and the Key Actions to design an improvement plan.

For large enterprises: The model should be used to formalise AI governance structures and regulatory compliance efforts. For SMEs and startups: Emphasis should be placed on cost-effective AI ethics integration, such as external audits or shared compliance resources.

Several limitations were encountered in adapting the maturity model to varied organizational contexts. Piloting the model in more organizations and of different sizes can bring more consistency to the results. Further work with organizations from around Europe and abroad should also be pursued.

Potential next steps such as longitudinal studies to track AI maturity progression over time and cross-sector validation of the model, would enhance the study's contribution to the field.

As AI entrenches every day in every process, people other than those involved in AI planning, design, development and implementation could also bring different perspectives to overall trustworthy AI. In this research only those related to decision-making were included, which can be considered a limitation of the study. For future research it is recommended to extend the scope of the participants in the organizations.

## Declaration of Generative AI and AI-assisted technologies in the writing process'.

During the preparation of this work the author used DEEPL WRITE in order to improve the wording in a limited number of statements. After using this tool, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

## REFERENCES


- Bertram D. (2007). "Likert scale and the meaning of life".
- Borrego, M., Douglas, E. and Amelink, C. (2009). "Quantitative, Qualitative, and Mixed Research Methods Engineering Education". Journal of Engineering Education. Pages 53-56.
- Brannen, J. (2005). "Mixing methods: the entry of qualitative and quantitative approaches into the research process International". Journal of Social Research Methodology Vol 8 No 3 July 2005 Special Issue.
- Bryman A. (2012). "Integrating Quantitative and Qualitative Research: How It Is Done". Qualitative Research 6 (1): pages 97-113. https://doi.org/10.1177/1468794106058877.
- Burns Richard and Robert P. Burns (2008). "Business Research Methods ands Statiscs using SPSS. Sage.
- CEN, CEN/CLC/JTC 2, (2023). "Introduction to latest draft trustworthiness framework".
- CEN, prCEN/CLC/TR 17894, (2023)."Artificial Intelligence Conformity Assessment".
- Cornish, P., (2021). "Environmental, Social and Governance Risk", Hendersen Risk Limited.
- De Bruin, T. and Rosemann, M. (2005). "Understanding the main phases of developing a maturity assessment model", 16th Australasian Conference on information systems maturity assessment model, Sydney. https://eprints.qut.edu.au/25152/1/Understanding_the_Main_Phases_of_Developing_a_Maturity_Assessment_Model.pdf.
- Ellefsen, A., Oleśków-Szłapka, J., Pawłowski, G., Toboła, A., (2019). "Striving for excellence in AI implementation: AI maturity model framework and preliminary

research results", LogForum - Scientific Journal of Logistics, pages 363 -376. https://yadda.icm.edu.pl/baztech/contributor/c105b0ea4d9484e25abf930f3ad78a05"

- European Commission, (2021). EU AI Act, "Proposal for a regulation of the European Parliament and the Council laying down harmonized rules on artificial intelligence and amending certain union legislative acts".
- European Commission, (2020). "PRISMA Responsible innovation in practice: experiences from industry".
- European Commission, (2019). "Ethics guidelines for trustworthy AI from the High-Level Expert Group on Artificial Intelligence (HLEG-AI)".
- Google, (2018). "Artificial Intelligence at Google: Our principles".
- Grunwald, A. (2009), "Technology assessment: concepts and methods, Handbook of Philosophy of Science, Philosophy of Technology and Engineering Sciences", pp.1103-1146
- IBM, (2019). "Everyday ethics for Artificial Intelligence".
- IEEE, (2017). The Institute of Electrical and Electronics Engineers, Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems.
- Institute for the Future, (2018). "Ethical OS toolkit checklist, a guide to anticipating the future impact of today's technology", https://ethicalos.org/wp-content/uploads/2018/08/Ethical-OS-Toolkit.pdf .
- Mylrea, M., Robinson, N., (2023). AI Trust Framework and Maturity Model: Improving Security, Ethics and Trust in AI, Cybersecurity and Innovation Technology Journal, Vol.1, No.1, 2023, pp. 1-15 DOI. 10.52889/citj.v1i1.198 1.
- NIST, the National Institute of Standards and Technologies, (2023). "Artificial Intelligence Risk Management Framework" (AI RMF 1.0), doi.org/10.6028/NIST.AI.100-1.

- Owen, R., von Schomberg, R., Macnaghten, P., (2021). "An unfinished journey? Reflections on a decade of responsible research and innovation", Pages 217-233, https://doi.org/10.1080/23299460.2021.1948789.
- OECD, (2019). "Scoping the OECD AI Principles, Deliberations of the Expert Group on Artificial Intelligence (AIGO)", OECD Digital Economy Papers No. 291.
- OECD (2019). "Artificial Intelligence & Responsible Business Conduct".
- OECD (2023). "A portrait of AI adopters across countries: Firm characteristics, assets' complementarities and productivity". OECD Science, Technology and Industry working papers.
- Pornel, J., Saldaña, G., (2013). "Four Common Misuses of the Likert Scale", Philippine Journal of Social Sciences and Humanities Volume 18 No. 2 (2013) pages 12-19.
- Shuster, T. et al, (2021), "Maturity models for the assessment of artificial intelligence in small and medium-sized enterprises", PLAIS EuroSymposium on digital transformation, Part of the lecture notes in business information processing book series (LNBIP, volume 429).
- Stahl B., (2011). "Ethical issues of emerging ICT applications, Towards responsible research and innovation in the information and communication technologies and security technologies fields", European Commission, chapter 1, p. 22.
- Stahl, B. et al., (2017). "Responsible research and innovation (RRI) maturity model", Special issue responsible research and innovation (RRI) in industry.
- The Responsible AI Institute, (2022), The Responsible AI Certification - White Paper.
- Tanujaya, B. et al. (2023). "Likert Scale in Social Sciences Research: Problems and Difficulties", FWU Journal of Social Sciences 16(4):89-101, DOI:10.51709/19951272/Winter2022/7.
- Tashakkori, Abbas, and Charles Teddie. (2009). "Integrating Qualitative and Quantitative Approaches to Research". In The SAGE Handbook of Applied Social

Research Methods, edited by Leonard Bickman and Debra J. Rog, 2nd editio, 283–317. Los Angeles: Sage.

- UNESCO, (2023). "UNESCO Recommendation on the Ethics of Artificial Intelligence".
- von Schomberg, R., (2011). "Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields", SSRN Electronic Journal, DOI:10.2139/ssrn.2436399.

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *[Journal name]* and was not involved in the editorial review or the decision to publish this article.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: