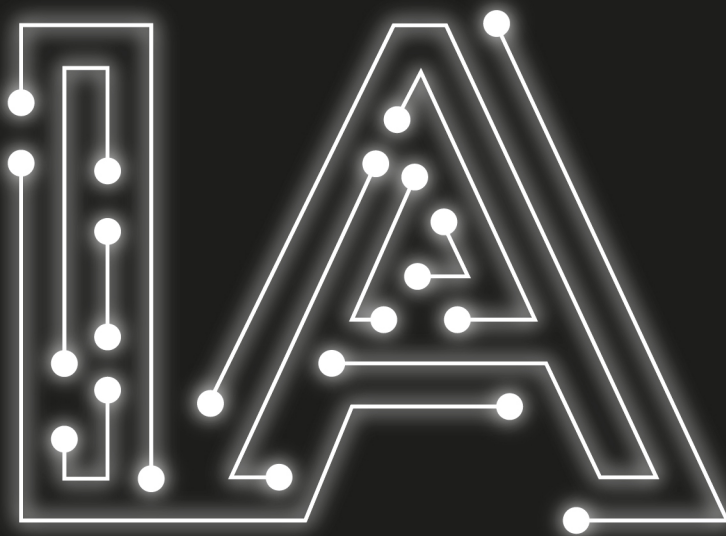



ONS.ANI

Inteligência Artificial CT223





Achieving Responsible
AI: The Portuguese
Committee on ISO/IEC
JTC 1/SC 42 Artificial
Intelligence.

Summary

Introduction	4
Building Ethical AI: A Guide to	6
Trustworthy AI Standard Development	6
The Characteristics of Trustworthy AI: A Comprehensive Framework	10
Technical Requirements for Safe and Reliable AI: A Look at Technical-related Characteristics.....	10
Stakeholder-related Characteristics in AI Standards: Ensuring Fairness and Inclusivity.....	11
Governance of Artificial Intelligence: The Importance of Accountability, Transparency, and Compliance.....	11
International Standards for Artificial Intelligence: An Overview of ISO/IEC JTC 1/SC 42 AI.....	12
Ensuring Responsible AI in Portugal: The Work of the Portuguese Committee on ISO/IEC JTC 1/SC 42 AI	14
References	17

Introduction

Artificial Intelligence (AI) is rapidly becoming an integral part of our daily lives, with applications spanning across various domains such as healthcare, finance, transportation, and many others. However, as the use of AI becomes more widespread, concerns about its ethical, fair and responsible use are also growing. In light of these concerns, international standards for AI are becoming increasingly important to ensure that the technology is developed and used in a way that is in the best interest of society. One of the organizations working on developing these standards is the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) through the ISO/IEC JTC 1/SC 42 Artificial Intelligence project.

The Portuguese Committee on ISO/IEC JTC 1/SC 42 Artificial Intelligence is a national body that is responsible for representing Portugal in the international standardization project on AI. The Committee is composed of experts from various fields such as computer science, engineering, ethics, law, and governance, and are representatives from industry, academia, government and civil society. The Committee works closely with the Portuguese National Innovation Agency (ANI) to ensure that the standards being developed align with the national

strategy for AI and take into account the interests of the country.

The goal of the Portuguese National Committee is to ensure that the interests and views of Portugal are taken into account in the development of international standards for AI. The Committee is actively involved in the work of ISO/IEC JTC 1/SC 42, participating in the working groups and subgroups of the project and contributing to the development of the standards. The Committee also organizes and participates in national and international events to disseminate the work of the project and to gather feedback and input from stakeholders.

One of the main contributions of the Portuguese Committee is the participation in the working group on Trustworthy AI, which aims to develop a set of guidelines for the development of trustworthy AI systems. The Portuguese Committee has been actively involved in the development of the guidelines, contributing to the identification of key issues and the development of the draft standards.

Another important area of work for the Portuguese Committee is the development of standards for the governance of AI systems. The Committee has been actively involved in the working group on governance, contributing to the development of standards for accountability, transparency, and compliance with laws

and regulations. The Committee also works to ensure that the standards being developed are in line with the national legal framework for AI, including data protection and privacy laws. This is essential for ensuring that AI systems are used in a responsible and ethical way, and that they are in compliance with the legal frameworks in place.

In conclusion, the Portuguese Committee on ISO/IEC JTC 1/SC 42 Artificial Intelligence plays a crucial role in representing Portugal in the international standardization project on AI. The work of the Portuguese Committee on ISO/IEC JTC 1/SC 42 Artificial Intelligence is essential for ensuring that the standards being developed for AI are inclusive, technically sound and aligned with the national and international context, thus promoting the responsible and ethical use of AI in Portugal.

Building Ethical AI: A Guide to Trustworthy AI Standard Development

Trustworthy AI Standard Development is a critical aspect of ensuring that the technology is used ethically, fairly, and in a way that is in the best interest of society^{1,2}. The development of trustworthy AI systems can help to mitigate the potential risks of unintended consequences and negative impacts

associated with AI and can ensure that AI is developed and used in a way that is beneficial for society²⁻⁴.

The development of trustworthy AI requires the consideration of several technical, stakeholder, and governance-related characteristics. Technical characteristics include the ability to make accurate predictions and decisions, the ability to explain its reasoning, and the ability to operate in a safe and secure manner⁵. Stakeholder characteristics include the ability to respect user privacy and data protection, the ability to promote fairness and non-discrimination, and the ability to respect human values and ethics⁶. Governance characteristics include the ability to be accountable for its actions, the ability to be transparent in its decision-making processes, and the ability to be governed by laws and regulations^{7,8}.

The development of trustworthy AI standards is a multidisciplinary effort, involving experts from various fields such as computer science, engineering, ethics, law, and governance. The development process typically involves several stages, including the identification of key issues, the development of draft standards, public consultation, and finalization of the standard^{9,10}.

One of the key issues in the development of trustworthy AI standards is ensuring that the standards are aligned with the latest research and developments in the field. This requires close collaboration between researchers and standardization experts, to ensure that the standards are

based on the latest scientific knowledge and are technically sound^{6,11-13}. Additionally, the standards should be developed in a way that is inclusive and takes into account the views and needs of all stakeholders, including industry, governments, and civil society^{6,11-13}.

Another important aspect of trustworthy AI standard development is ensuring that the standards are flexible and adaptable to the rapidly changing landscape of AI. The standards should be designed in a way that can be updated and revised as new research and developments emerge, to ensure that they remain relevant and effective¹⁴. Additionally, the standards should be designed in a way that is easy to implement and comply with, to ensure that they are widely adopted and used¹⁴⁻¹⁷.

Governance characteristics of Trustworthy AI are also important for standard development. These include the ability to be accountable for its actions, the ability to be transparent in its decision-making processes, and the ability to be governed by laws and regulations¹⁸. Accountability means that AI systems should be able to explain their decision-making processes and be answerable for any negative impacts they may cause. Transparency means that the decision-making processes of AI systems should be explainable and understandable to humans^{15,18}. Laws and regulations ensure that AI systems are used in a way that is compliant with relevant laws and regulations, such as those related to data privacy, data protection, and human rights, as

highlighted in "Achieving Trustworthy AI: A Framework for Responsible Development and Deployment" by IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2018).

International Organization for Standardization (ISO) has been working on the standardization of AI since 2018, through its Technical Committee ISO/IEC JTC 1/SC 42 Artificial intelligence. This committee is focused on the development of international standards for the governance and use of AI. The committee aims to create standards that promote the development of trustworthy AI systems and to provide a framework for the governance of AI. The committee's work is ongoing, and it is expected that the first standards will be published in the near future.

Moreover, many other organizations around the world are also working on developing AI standards such as IEEE, ITU, and ISO/IEC JTC 1/SC 42 Artificial intelligence. These organizations are working in collaboration with governments, industry, and civil society to develop standards that promote the responsible development and use of AI, as highlighted in "Trustworthy AI: a framework for responsible development" by European Commission (2021).

The Characteristics of Trustworthy AI: A Comprehensive Framework

Characterization of trustworthy AI refers to the set of attributes or characteristics that define a trustworthy AI system¹⁹⁻²¹. These characteristics are used to evaluate the safety, reliability, and trustworthiness of an AI system, and to ensure that it is used for the benefit of society²². The characterization of trustworthy AI is a critical aspect of the development of AI standards, as it provides a framework for the evaluation and governance of AI systems²¹.

The characterization of trustworthy AI is a critical aspect of the development of AI standards. Together, these characteristics provide a comprehensive framework for the evaluation and governance of AI systems, which is crucial for ensuring that AI is developed and used for the benefit of society²¹.

Technical Requirements for Safe and Reliable AI: A Look at Technical Characteristics

Technical characteristics of trustworthy AI include accuracy, robustness, explainability, safety, and adaptability¹⁴. These characteristics ensure that the AI system can be trusted to make decisions that are in the best interest of society and that it is safe, reliable, and

transparent in its decision-making process¹⁴. They are crucial for responsible use of AI in various domains such as healthcare, finance, and transportation.

Stakeholder Characteristics in AI Standards: Ensuring Fairness and Inclusivity

Stakeholder characteristics of trustworthy AI are essential for ensuring that the system is inclusive and respectful of human rights. Characteristics such as respecting user privacy and data protection, promoting fairness and non-discrimination, respecting human values and ethics, and considering the social and economic impact of its decisions are crucial for ensuring that the system is used in a responsible and ethical manner^{12,14}.

Governance of Artificial Intelligence: The Importance of Accountability, Transparency, and Compliance

Governance characteristics of trustworthy AI refer to the attributes that ensure the system is accountable, transparent, and follows laws and regulations. These characteristics are essential for the responsible and ethical use of AI, including accountability for negative impacts, clear explanations of decision-making processes, compliance with laws and regulations, and being auditable for transparency and fairness^{14,17}.

International Standards for Artificial Intelligence: An Overview of ISO/IEC JTC 1/SC 42 AI

ISO/IEC JTC 1/SC 42 Artificial Intelligence is an international standardization project on AI that was launched in 2018. The project is being carried out by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). The goal is to develop international standards for AI that will help to ensure that the technology is used ethically, fairly, and in a way that is in the best interest of society.

The project is focused on several key areas, including the development of standards for trustworthy AI, the characterization of trustworthy AI, and the governance of AI. The development of standards for trustworthy AI involves the identification of key issues, the development of draft standards, public consultation, and the finalization of the standards. The characterization of trustworthy AI involves the development of a set of attributes or characteristics that define a trustworthy AI system, which is used to evaluate the safety, reliability, and trustworthiness of an AI system. The governance of AI involves the development of standards for the governance of AI

systems, including accountability, transparency, and compliance with laws and regulations.

The project is expected to take several years to complete, and the standards developed will be regularly updated and revised as new research and developments emerge in the field of AI. The final standards will be available for use by industry, governments, and other organizations worldwide.

In conclusion, ISO/IEC JTC 1/SC 42 Artificial Intelligence is an important international standardization project on AI that aims to ensure that the technology is used ethically, fairly, and in a way that is in the best interest of society. The project is focused on the development of standards for trustworthy AI, the characterization of trustworthy AI, and the governance of AI. The project is being carried out by a multidisciplinary team of experts and is expected to take several years to complete. The final standards will be available for use by industry, governments, and other organizations worldwide and will be regularly updated as new research and developments in the field of AI emerge.

Ensuring Responsible AI in Portugal: The Work of the National Committee on ISO/IEC JTC 1/SC 42 AI

The Portuguese National Committee on ISO/IEC JTC 1/SC 42 Artificial Intelligence is a national body that is responsible for representing Portugal in the international standardization project that is being carried out by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). The goal of the Portuguese Committee is to ensure that the interests and views of Portugal are taken into account in the development of international standards for AI.

The Portuguese Committee is composed of experts from various fields such as computer science, engineering, ethics, law, and governance. The Committee is represented by the Portuguese National Innovation Agency (ANI) to ensure that the standards being developed align with the national strategy for AI and take into account the interests of the country.

The Portuguese Committee is actively involved in the work of ISO/IEC JTC 1/SC 42, participating in the working groups and subgroups of the project and contributing to the development of the standards. The

committee also organizes and participates in national and international events to disseminate the work of the project and to gather feedback and input from stakeholders.

One of the main contributions of the Portuguese Committee is the participation in the working group on Trustworthy AI, which aims to develop a set of guidelines for the development of trustworthy AI systems. The Portuguese Committee has been actively involved in the development of the guidelines, contributing to the identification of key issues and the development of the draft standards.

Another important area of work for the Portuguese Committee is the development of standards for the governance of AI systems. The Committee has been actively involved in the working group on governance, contributing to the development of standards for accountability, transparency, and compliance with laws and regulations. The Committee also works to ensure that the standards being developed are in line with the national legal framework for AI, including data protection and privacy laws.

In addition to its participation in the working groups and subgroups of the project, the Portuguese Committee also organizes and participates in national and international events to disseminate the work of the project and to gather feedback and input from

stakeholders. These events include workshops, seminars, and conferences, which provide an opportunity for experts and stakeholders to exchange ideas.

In conclusion, the Portuguese Committee on ISO/IEC JTC 1/SC 42 Artificial Intelligence plays an important role in representing Portugal in the international standardization project on AI. The work of the Portuguese Committee is essential for ensuring that the standards being developed for AI are inclusive, technically sound and aligned with the national and international context.

References

1. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. The ethics of algorithms: Mapping the debate. <http://dx.doi.org/10.1177/2053951716679679> 3, (2016).
2. Reinhardt, K. Trust and trustworthiness in AI ethics. *AI and Ethics* 2022 1, 1-10 (2022).
3. Bostrom, N. & Yudkowsky, E. The Ethics of Artificial Intelligence. *Artificial Intelligence Safety and Security* 57-69 (2018) doi:10.1201/9781351251389-4.
4. Taddeo, M. & Floridi, L. How AI can be a force for good. *Science* (1979) 361, 751-752 (2018).
5. Oneto, L. et al. Towards learning trustworthily, automatically, and with guarantees on graphs: An overview. *Neurocomputing* 493, 217-243 (2022).
6. Lewis, D. & Filip, D. An Ontology for Standardising Trustworthy AI.
7. AlShamsi, M., Salloum, S. A., Alshurideh, M. & Abdallah, S. Artificial intelligence and blockchain for transparency in governance. *Studies in Computational Intelligence* 912, 219-230 (2021).
8. Medaglia, R., Gil-Garcia, J. R. & Pardo, T. A. Artificial Intelligence in Government: Taking Stock and Moving Forward. <https://doi.org/10.1177/08944393211034087> (2021) doi:10.1177/08944393211034087.
9. Sharkov, G., Todorova, C. & Varbanov, & P. Strategies, Policies, and Standards in the EU Towards a Roadmap for Robust and Trustworthy AI Certification. 50, 11-22 (2021).
10. Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat Mach Intell* 1, 389-399 (2019).
11. Lewis, D., Pandit, H. J., Wall, P. J. & Filip, D. Standardization and the Governance of Artificial Intelligence.
12. Floridi, L. Establishing the rules for building trustworthy AI. *Nature Machine Intelligence* 2019 1:6 1, 261-262 (2019).
13. Lewis, D., Hogan, L., Filip, D. & Wall, P. J. Global Challenges in the Standardization of Ethics for Trustworthy AI. *Journal of ICT Standardization* 8, 123-150-123-150 (2020).
14. European Commission. Ethics Guidelines for Trustworthy AI. <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>.

15. Hickman, E. & Petrin, M. Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective. *European Business Organization Law Review* 22, 593–625 (2021).
16. Wickramasinghe, C. S., Marino, D. L., Grandio, J. & Manic, M. Trustworthy AI development guidelines for human system interaction. *International Conference on Human System Interaction, HSI 2020–June*, 130–136 (2020).
17. Smuha, N. A. The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. *Computer Law Review International* 20, 97–106 (2019).
18. European Commission. Artificial Intelligence Act. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF (2021).
19. Liang, W. et al. Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence* 2022 4:8 4, 669–677 (2022).
20. Reinhardt, K. Trust and trustworthiness in AI ethics. *AI and Ethics* (2022) doi:10.1007/S43681-022-00200-5.
21. Floridi, L. Establishing the rules for building trustworthy AI. *Nature Machine Intelligence* 2019 1:6 1, 261–262 (2019).
22. Zou, J. & Schiebinger, L. AI can be sexist and racist—it's time to make it fair. *Nature* 559, 324–326 (2018).

Achieving Responsible AI: The Portuguese Committee on ISO/IEC JTC 1/SC 42 Artificial Intelligence.

National Innovation Agency
ONS.ANI
Contact Helena Rosário Da Costa and Carolina Morais
E-mail helena.costa@ani.pt and carolina.morais@ani.pt

This document has been prepared for National Innovation Agency.
More information on the National Innovation Agency is able on the internet (<https://www.ani.pt/>).

© National Innovation Agency



ONS.ANI

Inteligência Artificial . CT223