ISSN 1831-9424

European Commission

JRC SCIENCE FOR POLICY REPORT

Glossary of human-centric artificial intelligence

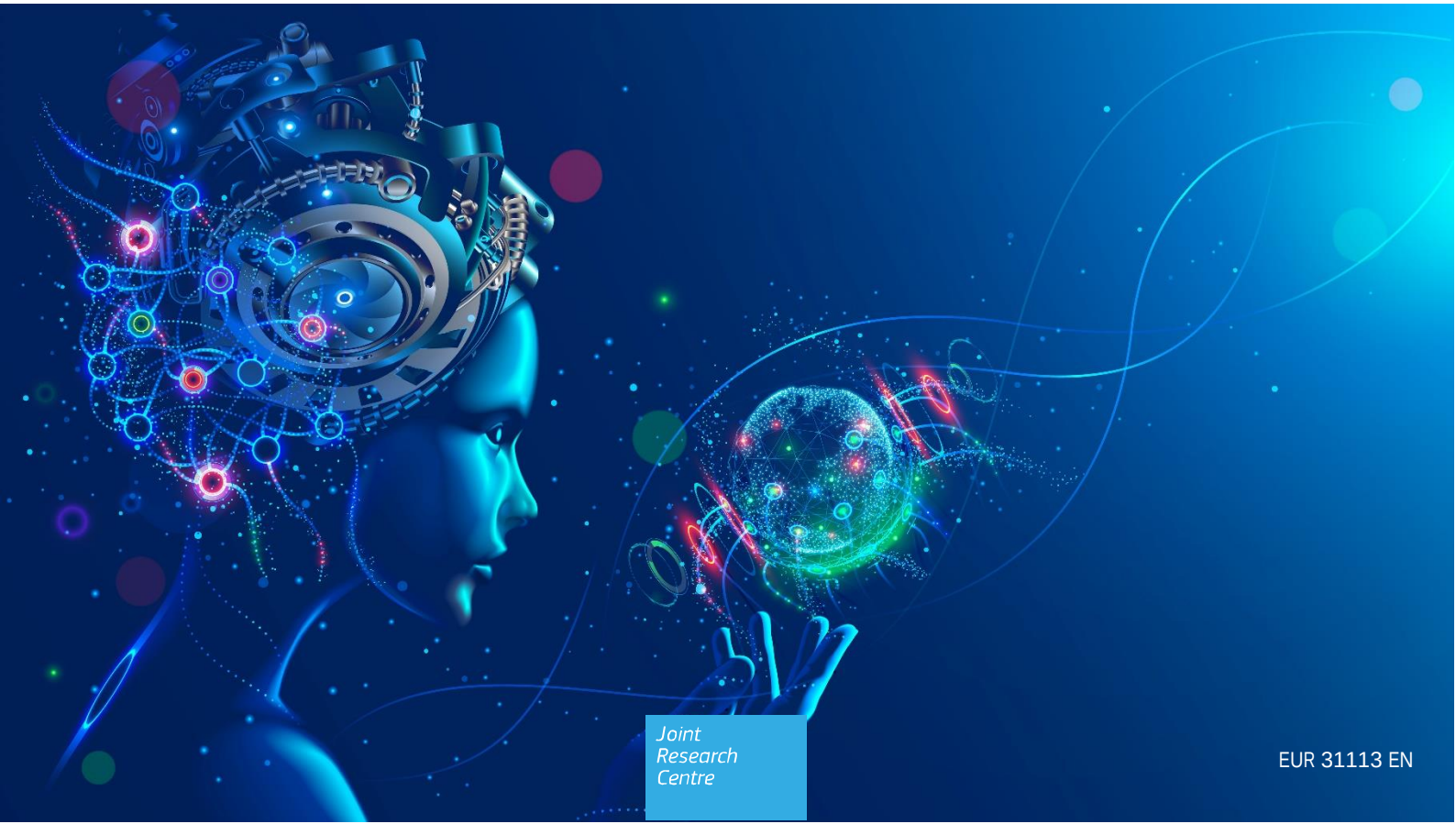
Estévez Almenzar, M.

Fernández Llorca, D.

Gómez, E.

Martínez Plumed, F.

2022



Joint Research Centre

EUR 31113 EN

Contact information
Name: Emilia Gómez
Email: emilia.gomez-gutierrez@ec.europa.eu

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC129614

EUR 31113 EN

PDF ISBN 978-92-76-53432-7 ISSN 1831-9424 doi:10.2760/860665 KJ-NA-31113-EN-N

Luxembourg: Publications Office of the European Union, 2022

How to cite this report: Estevez Almenzar, M., Fernández Llorca, D., Gómez, E., Martínez Plumed, F., *Glossary of human-centred artificial intelligence*, Publications Office of the European Union, Luxembourg, 2022, doi:10.2760/860665, JRC129614.

Contents

Abstract

Over the last few years, Artificial Intelligence (AI) has become a very active research topic, moving from a purely technical field to an interdisciplinary research domain and a very active topic in terms of policy developments. The European approach for AI focuses on two main areas: excellence and trust, enabling the development and uptake of AI while ensuring people's safety and fundamental rights. However, research and policy documentations do not always use the same vocabulary, often generating misunderstandings among researchers, policy makers, and the general public. Based on existing literature in the intersection between research, industry and policy, and given the expertise and know-now developed at the European Commission's Joint Research Centre, we present here a glossary of terms on AI, with a focus on a human-centric approach, covering concepts related to trustworthy artificial intelligence such as transparency, accountability or fairness. We have collected 230 different terms from more than 10 different general sources including standards, policy documents and legal texts, as well as multiple scientific references. Each term is accompanied by one or several definitions linked to references and complemented with our own definitions when no relevant source was found. We humbly hope that the work presented here can contribute to establishing the necessary common ground for the interdisciplinary and policy-centred debate on artificial intelligence.

Acknowledgements

We thank all members of the HUMAINT team and colleagues working on AI on different European Commission DGs, mainly CNECT, JUST and EAC, for engaging in interdisciplinary, policy-oriented discussions on AI which motivated and shaped this work. In particular, we would like to thank and acknowledge the work of our colleagues in the E.3 unit of the JRC, who have participated in the discussions of some of the terms of this report, and the reviewers that have participated in Pubsy:

Cachia, Romina (DG JRC)

Hamon, Ronan (DG JRC)

Junklewitz, Henrik (DG JRC)

Sánchez, Ignacio (DG JRC)

Authors

Estévez Almenzar, Marina

Fernández Llorca, David

Gómez, Emilia

Martínez Plumed, Fernando

Executive summary

Policy context

Over the last few years, Artificial Intelligence (AI) has become a very active field of research, which has evolved from a purely technical field to a research domain spanning different disciplines such as cognitive science, economics, or law. In addition, AI has become a very active topic in terms of policy developments, with governments and institutions defining investment strategies, educational programs or ethical guidelines. The European approach for AI focuses on establishing an ecosystem of excellence and trust in AI in Europe, enabling the development and uptake of AI while ensuring people´s safety and fundamental rights. While policy developments in AI should be in line with scientific and technical understanding, there are sometimes differences in vocabulary, which sometimes generate misunderstandings between different research communities or scientists and policy makers.

Key conclusions

Based on existing literature in the intersection between academia, industry and policy, and given the expertise and know-how developed at the European Commission´s Joint Research Centre, we present here a compact but comprehensive glossary of terms on AI, with a focus on a human-centric approach, intended to be used as a relevant reference for interdisciplinary and policy-centred discussions on the topic.

Main outcomes

We have collected and adapted 230 different terms from more than 10 relevant sources including standards, policy documents and legal texts, as well as multiple scientific references. These include concepts related to trustworthy and human-centred AI such as transparency, fairness or accountability.

Related and future JRC work

The glossary builds upon the work of the HUMAINT research programme, Joint Research Centre, which studies the impact of AI in human behaviour and provides scientific and technological support to AI policies at the European Commission. This includes contributions to the recently proposed European regulatory framework for AI (AI Act) and AI liability initiative.

Quick guide

The document is structured as follows. The document first includes a summary of the motivation, goals and structure of the glossary. It then provides the core contribution of the report, which is the list of terms, accompanied by one or several definitions linked to references, and complemented with own definitions when no relevant source was found. The glossary is then complemented by a short discussion on findings, limitations and steps for future work on the topic.

Note that definitions in some sources (such as ISO/IEC) contain annotations such as labels (e.g., "<engineering system>") to refer to the perspective adopted to define the term, or in some cases numerical references (e.g. "In a machine learning context (3.2.29)...") that refer to some terms that are included in the original source. See the original sources for more details on specific annotations.

1 Introduction

Over the last few years, Artificial Intelligence (AI) has become a very active research topic, moving from a purely technical field to an interdisciplinary research domain.

AI has been an area of major policy developments, and we have witnessed how countries governments and institutions around the world have been proposing investment plans, national strategies or ethical guidelines over the last few years. The European approach for AI focuses on two main areas: excellence and trust, enabling the development and uptake of AI while ensuring people´s safety and fundamental rights[1]. On the one hand, fostering excellence intends to strengthen Europe´s potential to compete globally by having member states joining forces on AI policy and investments, in the context of a Coordinated Plan on AI[2]. On the other hand, building trustworthy AI intends to create a safe and innovation-friendly environment, and the European Commission has proposed three main legal initiatives on this respect: a European regulatory framework for AI (AI Act[3]), rules to address liability aspects of new technologies, including AI, and a revision of sectorial safety legislation (e.g., Machinery regulation).

However, research and policy documentations do not always use the same vocabulary, often generating misunderstandings among researchers, policy makers, and the general public. In this respect, the work of Samoili et al. (2021) provide an analysis of 65 different definitions and taxonomies related to the term "Artificial Intelligence", which is defined in varied ways in different research, industry and policy-oriented documents.

The present document presents a list of definitions of terms and concepts relevant to the study of AI from a human-centric perspective. The glossary builds upon the work of the HUMAINT research programme (Gomez et al., 2021) that studies the impact of AI in human behaviour and yields in the science for policy interface. The goal of this document is then to provide, based on existing knowledge, a representative but still compact repository of terms, with the goal of serving AI researchers, practitioners and policy makers build a common understanding of AI concepts that are and will be of great importance in current and future policies on the area.

In Section 2 we provide a summary of the methodology designed to build this glossary. This will be followed by the list of terms, in alphabetical order, linked to sources and related references, which form the core contribution of this document. We conclude with a short conclusion of the study and directions for future work.

[1] European Commission, A European approach to artificial intelligence https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

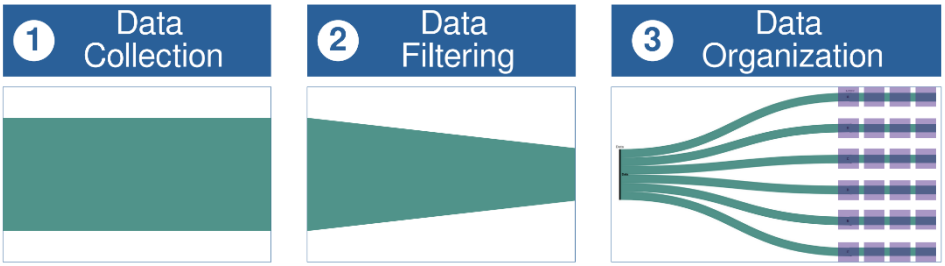
[2] European Commission, Coordinated Plan on Artificial Intelligence 2021 Review https://digital-strategy.ec.europa.eu/en/policies/plan-ai

[3] COM/2021/206 "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS", https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

2 Methodology

The presented glossary of terms related to human-centric AI has been created by means of a methodology based on three main pillars (as depicted in Fig. 1): (1) data collection, (2) data filtering, and (3) data analysis. Each pillar is composed of several steps. Each of these steps has been developed in parallel based on the expert analysis of the authors of this report. All participants are members of the HUMAINT team (Gomez et al., 2021) and have knowledge and experience in different aspects of human centric and trustworthy digital and IT systems, AI systems and recommender systems.

Figure 1. Three main tasks of the methodology to develop the glossary of terms: collection, filtering and organization.



Source: JRC, 2022.

The following is a summary of the main tasks addressed in each of the blocks.

2.1 Data Collection

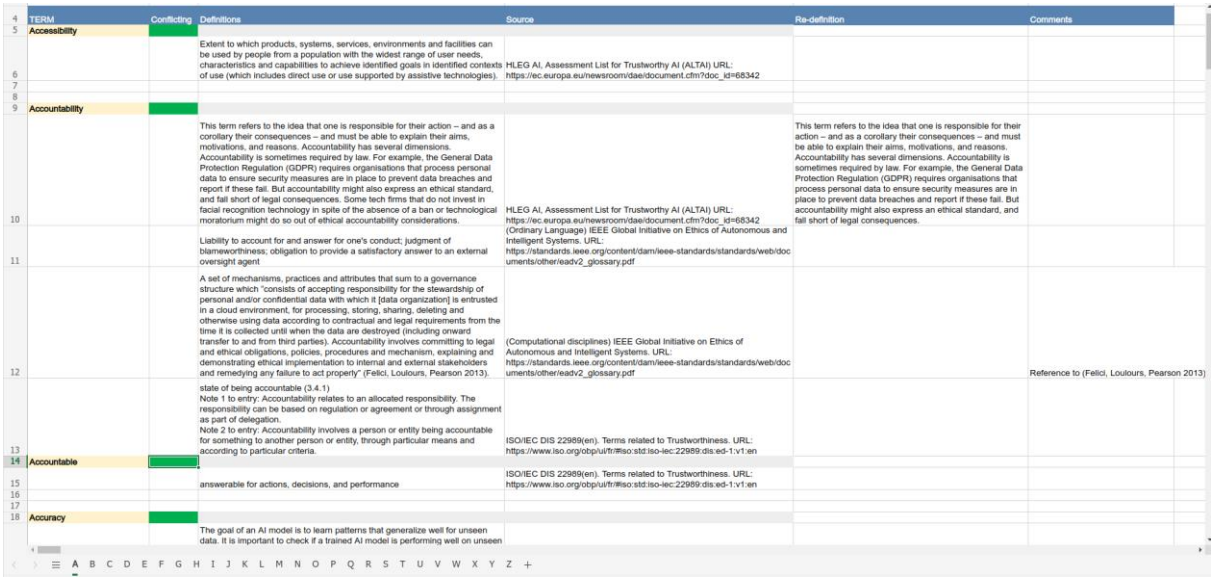
The data collection process was based on three main phases. First, the identification of similar glossaries of terms related to AI systems, both generic and specialised, and with different approaches depending on the application domain. In this first phase, quantity has been prioritised over quality, and broad and diverse application domains have been sought in order to try to collect as much information as possible.

The main glossaries and sources that have been identified and used are the following:

- Two glossaries from the High-Level Expert Group on Artificial Intelligence (AI HLEG) set up by the European Commission. The first one, with 16 terms, was included in the "Ethics Guidelines for Trustworthy Artificial Intelligence" (AI HLEG, 2019). The second one, with more than 40 terms, was presented in the "The Assessment List for Trustworthy AI (ALTAI)" (AI HLEG, 2020). In both cases, the glossaries are only intended to assist in the understanding of the terms used in both documents.
- The glossary of selected technical terms from the report on "Ethics of Connected and Automated Vehicles" developed by the Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659, 2020), which included 15 terms relevant to the content of the report.
- Different glossaries from the OECD such as the "OECD Glossary of Statistical Terms" (OECD, 2017), which contains a comprehensive set of definitions regarding statistical terms (highly relevant for AI systems) and the more recent "OECD Framework for the Classification of AI systems" (OECD, 2022). Another source from the OECD was the "Glossary of terms, in Promoting the Digital Transformation of African Portuguese-Speaking Countries and Timor-Leste" (OECD, 2018), which includes 36 terms related to digital technologies.
- The "ISO/IEC DIS 22989(en) Information technology - Artificial intelligence - Artificial intelligence concepts and terminology" (ISO, 2018), which includes terms related to AI, machine learning, neural networks, trustworthiness, and natural language processing.
- The "Glossary for Discussion of Ethics of Autonomous and Intelligent Systems" by the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (IEEE, 2017). This document includes multiple definitions from different perspectives, including ordinary language, computational disciplines, engineering, government, policy and social sciences, and ethics and philosophy.
- We also included definitions from different regulatory texts, such as the regulation on a "single market for digital services (Digital Services Act)" (COM(2020) 825), the regulation on "harmonised rules on fair access to and use of data (Data Act)" (COM(2022) 68) and the regulation "laying down harmonised rules on artificial intelligence (AI Act)" (COM(2021) 206).

In the second phase of the data collection process, all available terms in the glossaries were recorded in alphabetical order in a shared file, with detailed reference information on the source of the definitions. In most cases, several definitions are available for each term. To maximize the number of available definitions, no filtering was performed at this stage. An example of the shared file used to collect all the definitions is depicted in Figure 2.

Figure 2. Example of the shared file used to collect all the definitions. Second column refers to the colour code for conflicting cases.



	TERM	Conflicting	Definitions	Source	Re-definition	Comments
5	Accessibility					
6			Extent to which products, systems, services, environments and facilities can be used by people from a population with the widest range of user needs, characteristics and capabilities to achieve identified goals in identified contexts of use (which includes direct use or use supported by assistive technologies).	HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342		
7						
8						
9	Accountability					
10			This term refers to the idea that one is responsible for their action -- and as a corollary their consequences -- and must be able to explain their aims, motivations, and reasons. Accountability has several dimensions. Accountability is sometimes required by law. For example, the General Data Protection Regulation (GDPR) requires organisations that process personal data to ensure security measures are in place to prevent data breaches and report if these fail. But accountability might also express an ethical standard, and fall short of legal consequences. Some tech firms that do not invest in facial recognition technology in spite of the absence of a ban or technological moratorium might do so out of ethical accountability considerations.	HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342	This term refers to the idea that one is responsible for their action -- and as a corollary their consequences -- and must be able to explain their aims, motivations, and reasons. Accountability has several dimensions. Accountability is sometimes required by law. For example, the General Data Protection Regulation (GDPR) requires organisations that process personal data to ensure security measures are in place to prevent data breaches and report if these fail. But accountability might also express an ethical standard, and fall short of legal consequences.	
11			Liability to account for and answer for one's conduct; judgment of blameworthiness; obligation to provide a satisfactory answer to an external oversight agent	(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf		
12			A set of mechanisms, practices and attributes that sum to a governance structure which "consists of accepting responsibility for the stewardship of personal and/or confidential data with which it [data organization] is entrusted in a cloud environment, for processing, storing, sharing, deleting and otherwise using data according to contractual and legal requirements from the time it is collected until when the data are destroyed (including onward transfer to and from third parties). Accountability involves committing to legal and ethical obligations, policies, procedures and mechanism, explaining and demonstrating ethical implementation to internal and external stakeholders and remedying any failure to act properly" (Felici, Loulours, Pearson 2013).	(Computational disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf		Reference to (Felici, Loulours, Pearson 2013)
13			state of being accountable (3.4.1) Note 1 to entry: Accountability relates to an allocated responsibility. The responsibility can be based on regulation or agreement or through assignment as part of delegation. Note 2 to entry: Accountability involves a person or entity being accountable for something to another person or entity, through particular means and according to particular criteria.	ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en		
14	Accountable					
15			answerable for actions, decisions, and performance	ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en		
16						
17						
18	Accuracy					
			The goal of an AI model is to learn patterns that generalize well for unseen data. It is important to check if a trained AI model is performing well on unseen			

Source: JRC, 2022.

Finally, in the third phase, terms relevant to the field of human-centric AI that were not found in any of the glossaries analysed were identified and specifically searched for in a wide range of sources, from scientific articles, digital encyclopaedias (e.g., Wikipedia) and even specialised websites or blogs (e.g., Towards Data Science). Depending on the source, data collected in this last phase were flagged as potentially conflicting for further review in the filtering phase.

2.2 Data Filtering

Once all the terms and definitions have been collected, they are filtered according to two main criteria. First, very similar definitions for the same term that can be considered repetitive are identified and one of them is selected, giving priority to those glossaries that have a greater number of terms. Second, those definitions that are not entirely appropriate from a human-centric perspective, or that are not entirely relevant from a policy perspective (e.g., highly technical) are either eliminated when this criterion is clear from experts' opinion, or are flagged as conflicting when experts have some doubts about applicability. For example, some definitions provided by the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, such as those from the perspective of the ordinary language, or in some cases from the computational or engineering disciplines, are finally removed since the definitions are either very specific to a particular discipline, or very generic and therefore of limited applicability.

Finally, a final review of all the terms marked as conflicting is carried out to establish two levels. On the one hand, those terms that only require some minor adaptation or modification of one of the available definitions are marked in orange (minor corrections needed). On the other hand, those terms that require a completely new definition are marked in red (major correction needed). The rest of the terms are marked as green (all definitions are accepted as they are) as shown in Figure 2. These labels will be used in the final phase.

2.3 Data Organization

In this last process of the proposed methodology, the final organization of the glossary is addressed, including two main steps. First, all the terms marked as conflicting are revisited and the following actions are carried out:

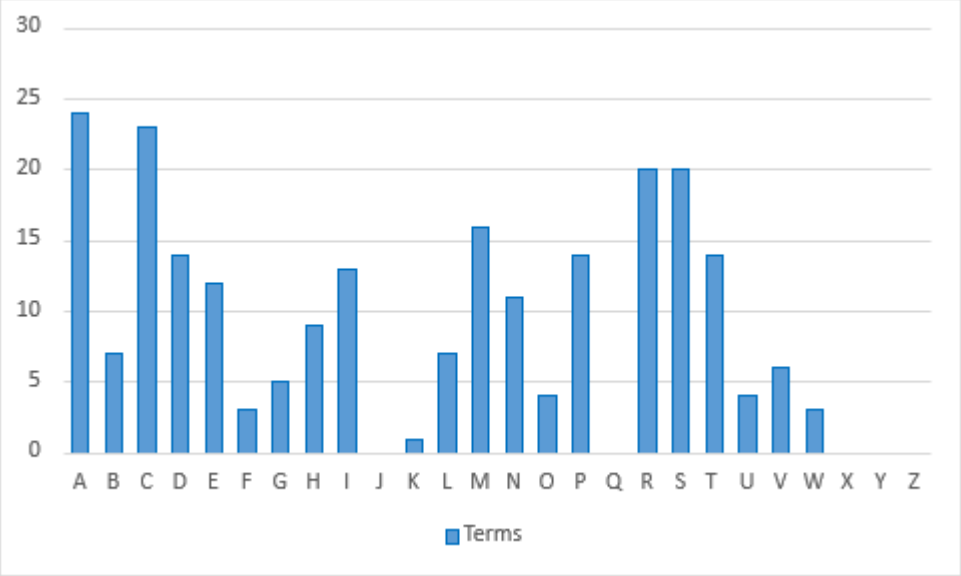
- For terms that have been marked by experts as needing minor corrections (marked in orange), modifications are applied to the most appropriate definition to adapt it to our requirements. In the glossary these cases are marked as "own elaboration, based on" or "own elaboration, adapted from".
- For those terms whose definitions were not considered satisfactory by the experts, i.e., those requiring major corrections (marked in red), definitions of our own, agreed upon by all the authors, have been developed. In the glossary these cases are marked as "Own elaboration".

It is important to note that AI is a highly multidisciplinary and evolving field. Therefore, it is very common to find different definitions for the same term. Sometimes the definitions differ only slightly. At other times, they differ considerably, as they are based on different perspectives. In order not to lose this multidisciplinary component, and in order not to discard different but relevant definitions, we have chosen to keep multiple definitions for some terms.

Finally, the last step of the methodology consists of proposing a specific order in the definitions for those terms with more than one definition available. We have ordered the terms in a consensual manner and based on expert judgement according to their importance in relation to the human-centric perspective and policy-oriented objectives.

In Figure 3 we depict the final distribution of terms obtained. We have collected a total of 230 terms in this first glossary of human-centric and policy-relevant terms.

Figure 3. Histogram with the distribution of terms of the glossary.



Source: JRC, 2022.

3 Glossary

3.1 A

Accessibility

Extent to which products, systems, services, environments, and facilities can be used by people from a population with the widest range of user needs, characteristics and capabilities to achieve identified goals in identified contexts of use (which includes direct use or use supported by assistive technologies).

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI)* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Accountability

This term refers to the idea that one is responsible for their action – and as a corollary their consequences – and must be able to explain their aims, motivations, and reasons. Accountability has several dimensions. Accountability is sometimes required by law. For example, the General Data Protection Regulation (GDPR) requires organisations that process personal data to ensure security measures are in place to prevent data breaches and report if these fail. But accountability might also express an ethical standard, and fall short of legal consequences.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI)* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Liability to account for and answer for one's conduct; judgment of blameworthiness; obligation to provide a satisfactory answer to an external oversight agent.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

A set of mechanisms, practices and attributes that sum to a governance structure which "consists of accepting responsibility for the stewardship of personal and/or confidential data with which it [data organization] is entrusted in a cloud environment, for processing, storing, sharing, deleting and otherwise using data according to contractual and legal requirements from the time it is collected until when the data are destroyed (including onward transfer to and from third parties). Accountability involves committing to legal and ethical obligations, policies, procedures and mechanism, explaining and demonstrating ethical implementation to internal and external stakeholders and remedying any failure to act properly" (Felici, Loulours, Pearson 2013).

Source: *(Computational disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

State of being accountable.

Note 1 to entry: Accountability relates to an allocated responsibility. The responsibility can be based on regulation or agreement or through assignment as part of delegation.

Note 2 to entry: Accountability involves a person or entity being accountable for something to another person or entity, through particular means and according to particular criteria.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Accountable

Answerable for actions, decisions, and performance.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Accuracy

The concept of accuracy of AI systems generally refers to the capability of the AI system to perform the task for which it has been designed. This includes, inter alia, making correct predictions about future outcomes, correctly classifying inputs, regrouping profiles into relevant categories, or generating faithful examples from abstract descriptions. Depending on the nature of the problem, various statistical metrics are used for evaluating the performance of AI systems, such as statistical accuracy, precision, recall, F1-score, mean square error, mean absolute error, to name but a few. These metrics are used in the training, validation and testing stages to give an indication of the behaviour of the model on the corresponding sets.

Source: *Own elaboration.*

The goal of an AI model is to learn patterns that generalize well for unseen data. It is important to check if a trained AI model is performing well on unseen examples that have not been used for training the model. To do this, the model is used to predict the answer on the test dataset and then the predicted target is compared to the actual answer. The concept of accuracy is used to evaluate the predictive capability of the AI model. Informally, accuracy is the fraction of predictions the model got right. A number of metrics are used in Machine Learning (ML) to measure the predictive accuracy of a model. The choice of the accuracy metric to be used depends on the ML task.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI). URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Adaptive

An adaptive AI is a system that changes its behaviour while in use. Adaptation may entail a change in the weights of the model or a change in the internal structure of the model itself. The new behaviour of the adapted system may produce different results than the previous system for the same inputs.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

An adaptive algorithm is an algorithm that changes its behaviour at the time it is run, based on a priori defined reward mechanism or criterion.

Source: *Glossary of artificial intelligence. Wikipedia. URL: https://en.wikipedia.org/wiki/Glossary_of_artificial_intelligence*

Adversarial Attack

A malicious attempt which tries to perturb the input of a machine learning model (e.g. adding some noise imperceptible by humans) to cause the model to draw incorrect conclusions (e.g. a misclassification, or an error in the confidence of the classification).

Source: *Own elaboration, based on Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572*

Action targeting a learning system to cause malfunction.

Source: *NIST. URL: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf*

Adversarial Example

An input to a Machine Learning model that is purposely designed to cause a model to make a mistake in its predictions despite resembling a valid input to a human.

Source: *Own elaboration, based on (Goodfellow et al., 2015)*

Machine Learning input sample formed by applying a small but intentionally worst-case perturbation to a clean example, such that the perturbed input causes a learned model to output an incorrect answer.

Source: *NIST. URL: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf*

Adversary

The agent who conducts or intends to conduct detrimental activities, perhaps by creating an adversarial example.

Source: *NIST. URL: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf*

Agent

An agent can be a physical or virtual entity that can act, perceive its environment (in a partial way) and communicate with others, is autonomous and has skills to achieve its goals and tendencies. It is in a multi-agent system (MAS) that contains an environment, objects and agents (the agents being the only ones to act), relations between all the entities, a set of operations that can be performed by the entities and the changes of the universe in time and due to these actions.

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Automated entity that perceives its environment and takes actions to achieve its goals.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

An intelligent being who acts by will, from intention, whether for its own ends or those of other agents.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Algorithm

An algorithm consists of a set of instructions or steps used to solve a problem (e.g., it does not include the data). The algorithm can be abstract and implemented in different programming languages and software libraries.

Source: *Own elaboration.*

A formula or set of rules (or procedure, processes, or instructions, or steps) for solving a problem or for performing a task. In Artificial Intelligence, the algorithm tells the machine how to find answers to a question or solutions to a problem. In Machine Learning, systems use many different types of algorithms. Common examples include decision trees, clustering algorithms, classification algorithms, or regression algorithms.

Source: *AI: A Glossary of Terms, Artificial Intelligence in Medical Imaging. URL: https://link.springer.com/content/pdf/bbm%3A978-3-319-94878-2%2F1.pdf*

Mechanisms for decision-making based on a set of digital rules and using input/output sources, encompassing Artificial intelligence (AI) algorithms, developed with the intention of mimicking human intelligence. Algorithms are usually embedded in hardware and software and can be based on other systems besides AI.

Source: *European Commission, Directorate-General for Research and Innovation, Ethics of connected and automated vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility, Publications Office, 2020. URL: https://data.europa.eu/doi/10.2777/966923*

Anonymisation

Anonymisation consists in preventing any identification of individuals from personal data.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI). URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Anticipatory Ethics

Analysis of the standards for good or bad actions taken when designing, developing, deploying, or decommissioning emerging technologies

Source: *(Ethics and Philosophy) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Application Programing Interface (API)

The calls, subroutines, or software interrupts that comprise a documented interface so that an application program can use the services and functions of another application, operating system, network operating system, driver, or other lower-level software program.

Source: *Shnier, M., Dictionary of PC Hardware and Data Communications Terms, O'Reilly Media, Inc., Sebastopol, CA, 1996.*

Set of well-defined methods, functions, protocols, routines or commands which application software uses with facilities of programming languages to invoke services

Note 1 to entry: An API is available for different types of software, including Web-based systems/ecosystems.

Source: *ISO/TS 23029:2020(en). URL: https://www.iso.org/obp/ui/#iso:std:iso:ts:23029:ed-1:v1:en*

A system access point or library function that has a well-defined syntax and is accessible from application programs or user code to provide well-defined functionality.

Source: *NIST, URL: https://csrc.nist.gov/glossary/term/application_programming_interface*

Artificial Intelligence

An AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives. It uses machine and/or human-based data and inputs to (i) perceive real and/or virtual environments; (ii) abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and (iii) use model inference to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy.

Source: *OECD AI Principles Overview. URL: https://oecd.ai/en/ai-principles*

Artificial intelligence (or machine intelligence) refers to systems that display intelligent behaviour by analysing their environment and taking actions - with some degree of autonomy - to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g., voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g., advanced robots, autonomous cars, drones, or Internet of Things applications). The term AI was first coined by John McCarthy in 1956.

Source: *AI: A Glossary of Terms, Artificial Intelligence in Medical Imaging. URL: https://link.springer.com/content/pdf/bbm%3A978-3-319-94878-2%2F1.pdf*

Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI). URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

AI systems are software (and possibly also hardware) systems that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.

Source: *European Commission, Directorate-General for Research and Innovation, Ethics of connected and automated vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility, Publications Office, 2020. URL: https://data.europa.eu/doi/10.2777/966923*

The capacity of computers or other machines to exhibit or simulate intelligent behaviour.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

<engineered system> Set of methods or automated entities that together build, optimize and apply a model so that the system can, for a given set of predefined tasks, compute predictions, recommendations, or decisions.

Note 1 to entry: AI systems are designed to operate with varying levels of automation.

Note 2 to entry: Predictions can refer to various kinds of data analysis or production (including translating text, creating synthetic images or diagnosing a previous power failure). It does not imply anteriority.

<discipline> Study of theories, mechanisms, developments and applications related to artificial intelligence <engineered system>

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Artificial Intelligence Practitioners

By AI practitioners we denote all individuals or organisations that develop (including research, design or provide data for) deploy (including implement) or use AI systems, excluding those that use AI systems in the capacity of end-user or consumer.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL:* https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

Artificial Intelligence System

A system based on artificial intelligence (see Artificial Intelligence).

Own elaboration.

Software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

Source: *EU Artificial Intelligence Act URL :* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Unity of concerns or techniques related to development of Artificial Intelligence that leads to design or development of Autonomous Agent Systems.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Engineered system featuring AI <engineered system>

Note 1 to entry: AI systems can be designed to generate outputs such as predictions, recommendations and classifications for a given set of human defined objectives.

Note 2 to entry: AI systems can be designed to operate with varying levels of automation.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Artificial Intelligence System Lifecycle

AI system lifecycle phases involve: i) 'design, data and models'; which is a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building; ii) 'verification and validation'; iii) 'deployment'; and iv) 'operation and monitoring'. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.

Source: *OECD, URL:* https://oecd.ai/en/ai-principles

Assistive Technology

Software or hardware that is added to or incorporated within an ICT (Information and Communication Technology) system to increase accessibility.

Often it is specifically designed to assist people with disabilities in carrying out daily activities. Assistive technology includes wheelchairs, reading machines, devices for grasping, etc. In the area of Web Accessibility, common software based assistive

technologies include screen readers, screen magnifiers, speech synthesizers, and voice input software that operate in conjunction with graphical desktop browsers (among other user agents). Hardware assistive technologies include alternative keyboards and pointing devices.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI). URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Devices and other solutions that assist people with deficits in physical, mental, or emotional functioning. Assistive technology devices are items frequently used by people with functional deficits as alternative ways of performing actions, tasks, and activities. Assistive technology also includes ways of controlling these devices. Software may control ordinary hardware systems in ways that facilitate their use by persons with functional deficits, like text-to-speech conversion software that runs on ordinary computers.

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Software and hardware purposively combined to augment or replace human sensory or cognitive tasks.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Augmented Reality

Virtual content layered over the real environment.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

A system that supplements the real world with virtual (computer generated) objects that appear to coexist in the same space as the real world. An AR system [will] have the following properties: combines real and virtual objects in a real environment; runs interactively, and in real time; and registers (aligns) real and virtual objects with each other" (Azuma et al 2001, 34).

Source: *(Engineering) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Augmented reality is the material/virtual nexus mediated through technology, information, and code, and enacted in specific and individualised space/time configurations (Graham, Zook, and Boulton 2012, 466).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Automatic/Automation/Automated

Pertaining to a process or system that, under specified conditions, functions without human intervention.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Automatic Summarization

Task of shortening a portion of natural language while retaining important semantic information.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Autonomy/Autonomous

Autonomy: Capability to perform with an absent or low degree of external influence.

Autonomous: Agent provided with autonomy.

Source: *Own elaboration.*

An autonomous AI system is an AI system that performs behaviours or tasks with a high degree of autonomy, that is, without external influence.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Characteristic of a system that is capable of modifying its operating domain or goal without external intervention, control or oversight.

Note 1 to entry: In jurisprudence, autonomy refers to the capacity for self-governance. In this sense, also, "autonomous" is a misnomer as applied to automated AI systems, because even the most advanced AI systems are not self-governing. Rather, AI systems operate based on algorithms and otherwise obey the commands of operators. For these reasons, this document does not use the popular term autonomous to describe automation.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

The ability of a person or artifact to govern itself including formation of intentions, goals, motivations, plans of action, and execution of those plans, with or without the assistance of other persons or systems.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Agents that are autonomous have control both over their internal state and over their own behaviour and autonomy means that the problem solvers have their own persistent thread of control (i.e., they are active) and that they decide for themselves which actions they should perform at what time".

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Auditability

Auditability refers to the ability of an AI system to undergo the assessment of the system's algorithms, data and design processes. This does not necessarily imply that information about business models and Intellectual Property related to the AI system must always be openly available. Ensuring traceability and logging mechanisms from the early design phase of the AI system can help enabling the system's auditability.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Availability

Property of being accessible and usable on demand by an authorised entity.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

3.2 B

Backpropagation

In Machine Learning, backpropagation is a widely used algorithm for training feedforward neural networks. Generalisations of backpropagation exist for other artificial neural networks (ANNs), and for functions generally. In fitting a neural network, backpropagation computes the gradient of the loss function with respect to the weights of the network for a single input–output example.

Source: *Wikipedia. URL:* https://en.wikipedia.org/wiki/Backpropagation

Neural network training method that uses the error at the output layer to adjust and optimise the weights for the connections from the successive previous layers.

Source: *ISO/IEC 23053:2022(en) Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML). URL:* https://www.iso.org/obp/ui/#iso:std:iso-iec:23053:ed-1:v1:en

Balanced Dataset

A balanced dataset refers to a dataset whose distribution of labels is approximately equal. Imbalanced datasets mean that the number of observations differs for the classes in a classification dataset.

Source: *Own elaboration.*

Bayesian Network

Probabilistic model that uses Bayesian inference for probability computations using a directed acyclic graph.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Bias

Bias (in AI) is an anomaly in the output of AI systems, due to the prejudices and/or erroneous assumptions made during the system development process or prejudices in the training data, so the results from the AI system cannot be generalised widely.

Source: *Own elaboration.*

Bias is a systematic deviation from a true state. From a statistical perspective an estimator is biased when there is a systematic error that causes it to not converge to the true value that it is trying to estimate. In humans, bias can manifest itself in deviating perception, thinking, remembering or judgment which can lead to decisions and outcomes differing for people based on their membership to a protected group. There are different forms of bias, such as the subjective bias of individuals, data and algorithm bias, developer bias and institutionalised biases that are ingrained in the underlying societal context of the decision.

Source: *Tolan, Songül. "Fair and unbiased algorithmic decision making: Current state and future challenges." arXiv preprint arXiv:1901.04730 (2019).*

Inclination of prejudice towards or against a person, object, or position. Bias can arise in many ways in AI systems. For example, in data-drive AI systems, such as those produced through machine learning, bias in data collection and training can result in an AI system demonstrating bias. In logic-based AI, such as rule-based systems, bias can arise due to how a knowledge engineer might view the rules that apply in a particular setting. Bias can also arise due to online learning and adaptation through interaction. It can also arise through personalisation whereby users are presented with recommendations or information feeds that are tailored to the user's tastes. It does not necessarily relate to human bias or human-driven data collection. It can arise, for example, through the limited contexts in which a system is used, in which case there is no opportunity to generalise it to other contexts. Bias can be good or bad, intentional or unintentional. In certain cases, bias can result in discriminatory and/or unfair outcomes, indicated in (HLEG AI, 2019) as unfair bias.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Systematic difference in treatment of certain objects, people, or groups in comparison to others.

Note 1 to entry: Treatment is any kind of action, including perception, observation, representation, prediction, or decision.

Note 2 to entry: Source info will be updated after ISO/IEC 24027 is published.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Big Data

An all-encompassing term for any collection of data sets so large or complex that they are difficult to store, manage and process with conventional, non-scalable technology.

Source: *Own elaboration. Adapted from the "Glossary of Big Data Terms" from SAGE. URL: https://campus.sagepub.com/blog/glossary-of-big-data-terms*

Extensive datasets - primarily in the data characteristics of volume, variety, velocity, and/or variability - that require a scalable technology for efficient storage, manipulation, management, and analysis.

Note 1 to entry: Big data is commonly used in many different ways, for example as the name of the scalable technology used to handle big data extensive datasets.

Source: *ISO/IEC 20546:2019(en) Information technology - Big data - Overview and vocabulary. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:20546:ed-1:v1:en*

Biometric data

Personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural persona, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data.

Source: *EU Artificial Intelligence Act URL : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206*

Black Box

In science, computing, and engineering, a black box is a system which can be viewed in terms of its inputs and outputs (or transfer characteristics), without any knowledge of its internal workings. Its implementation is "opaque" (black).

In neural networking or heuristic algorithms (computer terms generally used to describe 'learning' computers or 'AI simulations'), a black box is used to describe the constantly changing section of the program environment which cannot easily be tested by the programmers. This is also called a “white box” in the context that the program code can be seen, but the code is so complex that it is functionally equivalent to a black box.

Source: *Wikipedia, URL :* *https://en.wikipedia.org/wiki/Black_box*

3.3 C

Catastrophic forgetting / interference

Catastrophic forgetting (or catastrophic interference) is a problem in machine learning where a model forgets an existing learned pattern when learning a new one.

The model uses the same parameters to recognize both patterns and learning the second pattern overwrites the parameters’ configuration from having learned the first pattern.

Source: *Machine Learning Glossary (by James Mishra). URL:* *https://machinelearning.wtf/terms/catastrophic-forgetting/*

Under certain conditions, the process of learning a new set of patterns suddenly and completely erased a network’s knowledge of what it had already learned.

Source: *Robert M. French, "Catastrophic forgetting in connectionist networks", Trends in Cognitive Sciences, Volume 3, Issue 4, 1999, Pages 128-135, ISSN 1364-6613.*

Certainty / Uncertainty

Certainty: Quality of being reliably true. Refers to dealing with entities that are entirely deterministic and certain.

Uncertainty: Lack of certainty. Refers to situations involving imperfect or incomplete information. There are many sources of uncertainty in an AI system, including variance and noise in the specific data values, the (incomplete) sample of data collected from the domain, and in the imperfect nature of any models developed from such data.

Source: *Own elaboration*.

Chatbot

A computer program designed to simulate conversation with a human user, usually over the internet; especially one used to provide information or assistance to the user as part of an automated service.

Source: *Oxford English Dictionary. URL:*

https://www.oed.com/view/Entry/88357851?redirectedFrom=chatbot#eid

Cloud Computing

A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Source: *NIST, URL: https://csrc.nist.gov/glossary/term/cloud_computing*

Paradigm for enabling network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand.

Note 1 to entry: Examples of resources include servers, operating systems, networks, software, applications, and storage equipment.

Source: *ISO/IEC 20546:2019(en) Information technology - Big data - Overview and vocabulary. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:20546:ed-1:v1:en*

Cognition

The mental action or process of acquiring knowledge and understanding through thought, experience, and the senses.

Source: *Oxford University Press.*

A functional ontology for cognitive function includes 3 primary functions: "phonology (phonetic encoding and articulation), semantics (perceptual knowledge and functional knowledge), and orthography (visual synthesis of feature extraction and colour processing)" (Price and Friston 2005, 270).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Conscious knowledge.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Cognitive Computing

Programming designed to mimic human cognition.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Cognitive computing is an emerging paradigm of intelligent computing methodologies and systems based on cognitive informatics that implements computational intelligence by autonomous inferences and perceptions mimicking the mechanisms of the brain (Wang et al 2010, p. 1).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Category of AI systems to enable people and machines to interact more naturally.

Note 1 to entry: Cognitive computing tasks are associated with machine learning, speech processing, natural language processing, computer vision and human-machine interfaces.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Computation

Computation is the integration of numerical simulation, mathematical modelling, algorithm development and other forms of quantitative analysis to solve problems that theorization, experimentation, and/or observation cannot.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Computation is construed in 6 ways: "1. Formal symbol manipulation, 2. Effective computability, 3. Execution of an algorithm, 4. Digital state machines, 5. Information processing, 6. Physical symbol systems (Smith 2002, 3).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Computer Vision

Capability of an agent to acquire, process, and interpret visual data.

Note 1 to entry: Computer vision involves the use of sensors to create a digital image of a visual scene.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos, also seeking to understand and automate tasks that the human visual system can do.

Source: *Dana H. Ballard; Christopher M. Brown (1982). Computer Vision. Prentice Hall. ISBN 978-0-13-165316-0.*

Consciousness

The state or ability to be aware of self and environment.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Connectionism / Connectionist paradigm / Connectionist model / Connectionist approach

Form of cognitive modelling that uses a network of interconnected units which generally are simple computational units.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Consent

Agreement.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

The attachment of an agent's will to a proposal, action, or outcome, such that the agent accepts (some share of the) responsibility for the consequences and/or legitimizes an action or state of affairs which, in the absence of consent, would lack legitimacy or legality" (Reeve 2016).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Consensus

General agreement among a group.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

A consensus government is one in which multiple, independent perspectives are taken into account during decision making, rather than domination of decision-making by a winning party.

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Consistency

Also called monotonicity, refers to the condition of being unchanging or unvarying in tone.

Source: *Collins dictionary.*

Consumer

Means any natural person who is acting for purposes which are outside his or her trade, business or profession.

Source: *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. URL: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services*

Continuous Learning / Continual Learning / Lifelong Learning

The ability to continually learn over time by accommodating new knowledge while retaining previously learned experiences is referred to as continual or lifelong learning. Learning continually is crucial for agents and robots operating in changing environments and required to acquire, fine-tune, adapt, and transfer increasingly complex representations of knowledge. Such a continuous learning task has represented a long-standing challenge for machine learning and neural networks and, consequently, for the development of artificial intelligence (AI) systems. The main issue of computational models regarding lifelong learning is that they are prone to catastrophic forgetting or catastrophic interference, i.e., training a model with new information interferes with previously learned knowledge.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Incremental training of an AI system that takes place on an ongoing basis during the operation phase of the AI system life cycle.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

An adaptive algorithm capable of learning from a continuous stream of information, with such information becoming progressively available over time and where the number of tasks to be learned (e.g. membership classes in a classification task) are not predefined. Critically, the accommodation of new information should occur without catastrophic forgetting or interference.

Source: *Parisi et al. Continual Lifelong Learning with Neural Networks: a review, 2019.*

Control

The action or fact of holding in check or restraining; restraint.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Purposeful action on or in a process to meet specified objectives.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Control (adaptive)

An adaptive controller is a controller that can modify its behaviour in response to changes in the dynamics of the process and the disturbances. It can be considered as a special type of nonlinear feedback control in which the stages of the process can be separated into two categories, which can change at different rates (Bhatt and Shah 2002).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Control system

A control system manages, commands, directs, or regulates the behaviour of other devices or systems using control loops. It can range from a single home heating controller using a thermostat controlling a domestic boiler to large industrial control systems which are used for controlling processes or machines.

Source: *Wikipedia. URL: https://en.wikipedia.org/wiki/Control_system*

Controllability / Controllable

Property of an AI system that a human or other external agent can intervene in the system's functioning.

Note 1 to entry: Such a system is heteronomous.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Confidence score

A confidence score, or confidence interval, is a metric that quantifies the uncertainty or probability of an event, and represents the likelihood that the output of an AI system is correct and will satisfy a user's defined objective of an AI system's estimation.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Confidence scores, or confidence intervals, are a way of quantifying the uncertainty of such an estimate. A low confidence score associated with the output of an AI system means that the system is not too sure that the specific output is correct. Much of AI involves estimating some quantity, such as the probability that the output is a correct answer to the given input.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Convolutional Neural Network (CNN) / Deep Convolutional Neural Network (DCNN)

Feed forward neural network using convolution in at least one of its layers.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Convolution

Convolution is a mathematical operation on two functions (f and g) that produces a third function (f*g) that expresses how the shape of one is modified by the other. [...] It is defined as the integral of the product of the two functions after one is reversed and shifted. The integral is evaluated for all values of shift, producing the convolution function.

Source: *Wikipedia URL :* *https://en.wikipedia.org/wiki/Convolution*

Mathematical operation involving a sliding dot product or cross-correlation of the input data.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Culture

Culture is that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society (Tylor 1871).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Culture is a well-organised unity divided into two fundamental aspects—a body of artifacts and a system of customs (Malinowski 1931, 623). Culture is an historically transmitted pattern of meanings embodied in symbols (Geertz 1973, 89).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

3.4 D

Data

Symbols representing information that can be manipulated.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Data means "things given" in Latin— although we tend to use it as a mass noun in English, as if it denotes a substance—and ultimately, almost all useful data is given to us either by nature, as a reward for careful observation of physical processes, or by other people, usually inadvertently (consider logs of Web hits or retail transactions, both common sources of big data). As a result, in the real world, data is not just a big set of random numbers; it tends to exhibit predictable characteristics. For one thing, as a rule, the largest cardinalities of most datasets— specifically, the number of distinct entities about which observations are made—are small compared with the total number of observations (Jacobs 2009, 39).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

A value or set of values representing a specific concept or concepts. Data become "information" when analysed and possibly combined with other data in order to extract meaning and to provide context. The meaning of data can vary depending on its context.

A dataset is an organised collection of data. The most basic representation of a dataset is data elements presented in tabular form. Each column represents a particular variable. Each row corresponds to a given value of that column's variable. A dataset may also present information in a variety of non-tabular formats, such as an extended mark-up language (XML) file, a geospatial data file, or an image file (Data.gov, no date).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Data augmentation

Process of creating new data samples by manipulating the original data.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Data annotation

Process of attaching a set of descriptive information to data without any change to that data.

Note 1 to entry: The descriptive information can take the form of metadata, labels and anchors.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Database

Collection of data organised according to a conceptual structure describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas.

Source: *ISO/IEC 20546:2019(en) Information technology - Big data - Overview and vocabulary. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:20546:ed-1:v1:en*

A structured set of data held in computer storage and typically accessed or manipulated by means of specialised software.

Source: *Oxford English Dictionary. URL:*

https://www.oed.com/view/Entry/47411?redirectedFrom=database#eid

Data mining

Computational process that extracts patterns by analysing quantitative data from different perspectives and dimensions, categorizing it, and summarizing potential relationships and impacts.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Data quality checking

Process in which data is examined for completeness, bias and other factors which affect its usefulness for an AI system.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Data sampling

Process to select a subset of data samples intended to present patterns and trends similar to that of the larger dataset being analysed.

Note 1 to entry: Ideally, the subset of data samples will be representative of the larger dataset.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Datasets

Some AI systems rely on data sets to infer the logical mechanisms at play in the production of outcomes. Data sets are made of examples adapted to the task (e.g., pairs of inputs and labels for classification tasks), and are often divided into three parts, used in the three typical stages in the development of AI systems.

Training set: used for learning, that is, fitting the learnable parameters of a model (e.g., the weights of a neural network), for example using optimization techniques.

Validation set: used for validating the model, that is, in the context of AI development, providing an unbiased evaluation of the model after training and tuning the non-learnable parameters of the model and the learning process. The validation stage aims to prevent overfitting (the model begins to "memorize" training data rather than "learn" to generalize). The validation dataset can be a separate dataset or part of the training dataset, either as a fixed or variable split.

Test set: used for testing the final model, that is, providing an independent evaluation of the model after training and validation. The test dataset must be independent from the training and validation datasets, that is, data in the test dataset should not be used in training or validation. Testing here is meant to be seen as an internal stage in the development of an AI system to ensure good performance and may not substitute the testing phase with regard to other obligations.

Source: *Own elaboration*.

Collection of data with a shared format and goal-relevant content

EXAMPLE 1: Micro-blogging posts from June 2020 associated with hashtags #rugby and #football.

EXAMPLE 2: Macro photographs of flowers in 256x256 pixels.

Note 1 to entry: Datasets can be used for validating or testing an AI model. In a machine learning context, datasets can also be used to train a machine learning algorithm.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Data poisoning

Data poisoning occurs when an adversarial actor attacks an AI system and is able to inject bad data into the AI model's training set, thus making the AI system learn something that it should not learn. Examples show that in some cases these data poisoning attacks on neural nets can be very effective, causing a significant drop in accuracy even with very little data poisoning. Other kinds of poisoning attacks do not aim to change the behaviour of the AI system, but rather they insert a backdoor, which is data that the model's designer is not aware of, but that the attacker can leverage to get the AI system to do what they want.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Decision tree

Model for which inference is encoded as paths from the root to a leaf node in a tree structure.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Deep Learning

<artificial intelligence> Approach to creating rich hierarchical representations through the training of neural networks with many hidden layers.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Deep learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Deep learning architectures have been applied to fields including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, climate science, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.

Source: *Bengio, Y.; Courville, A.; Vincent, P. (2013). "Representation Learning: A Review and New Perspectives". IEEE Transactions on Pattern Analysis and Machine Intelligence. 35 (8): 1798–1828. arXiv:1206.5538. doi:10.1109/tpami.2013.50. PMID 23787338. S2CID 393948.*

Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". Neural Networks. 61: 85–117. arXiv:1404.7828. doi:10.1016/j.neunet.2014.09.003. PMID 25462637. S2CID 11715509.

Bengio, Yoshua; LeCun, Yann; Hinton, Geoffrey (2015). "Deep Learning". Nature. 521 (7553): 436–444. Bibcode:2015Natur.521..436L. doi:10.1038/nature14539. PMID 26017442. S2CID 3074096.

Development

A process of maturation of a plan or product from idea to fruition.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Dialog Management

Task of choosing the appropriate next move in a dialogue based on user input, the dialogue history and other contextual knowledge, to meet a desired goal.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Discrimination

Differentiation for the purpose of separating persons to determine entitlements, rights, or eligibility.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Discrimination algorithms are those that allow computer vision technologies, such as LiDAR, to differentiate types of objects or states of matter (see Hu et al 2009). Algorithms which reproduce social preferences that are discriminatory may be considered to be discriminatory algorithms.

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

The US Equal Employment Opportunity Commission describes types of discrimination. By: age, disability, genetic information, national origin, pregnancy, race/colour, religion, or sex.

Race discrimination involves treating someone (an applicant or employee) unfavourably because he/she is of a certain race or because of personal characteristics associated with race (such as hair texture, skin colour, or certain facial features). Colour discrimination involves treating someone unfavourably because of skin colour complexion" (EEOC no date).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

3.5 E

Emotion Recognition

Task of computationally identifying and categorizing emotions expressed in a piece of text, speech or image.

Note 1 to entry: Examples of emotions include happiness, sadness, anger and delight.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Encryption

Encryption is the procedure whereby clear text information is disguised by using a hash key. Encrypted results are unintelligible data for persons who do not have the encryption key.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

End-user

An end-user is the person that ultimately uses or is intended to ultimately use the AI system. This could either be a consumer or a professional within a public or private organisation. The end-user stands in contrast to users who support or maintain the product, such as system administrators, database administrators, information technology experts, software professionals and computer technicians.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI). URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

End-users (or intended users) of an AI system or application are the individuals or groups that use the system for a specific purpose.

Source: *OECD Framework for the Classification of AI systems. URL: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1645798047&id=id&accname=guest&checksum=1F40FEDBAC6979FE2EF85A8DF4B481CE*

Equality

Sameness in relevant respects (e.g., quantity, value).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

In the abstract, it means that people who are similarly situated in morally relevant respects should be treated similarly. Possible interpretations include equality before the law, equality of political power, equality of opportunity for social and economic advancement, equality of resources, equality of welfare, equality of freedom, and equality of respect (Nagel 2005).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Equity

By definition, equity is concerned with justice. On a societal level, equity is concerned with the just distribution of resources in society. Because a wide range of theories of just distribution exist, equity considerations are multifaceted and create a normative conceptual space in which theories can be considered, argued, and applied.

Source: *Lewis, E. O. C., MacKenzie, D., & Kaminsky, J. (2021). Exploring equity: How equity norms have been applied implicitly and explicitly in transportation research and practice.*

Ethics

Ethics is an academic discipline which is a subfield of philosophy. In general terms, it deals with questions like "What is a good action?", "What is the value of a human life?", "What is justice?", or "What is the good life?". In academic ethics, there are four major fields of research: (i) Meta-ethics, mostly concerning the meaning and reference of normative sentence, and the question how their truth values can be determined (if they have any); (ii) normative ethics, the practical means of determining a moral course of action by examining the standards for right and wrong action and assigning a value to specific actions; (iii) descriptive ethics, which aims at an empirical investigation of people's moral behaviour and beliefs; and (iv) applied ethics, concerning what we are obligated (or permitted) to do in a specific (often historically new) situation or a particular domain of (often historically unprecedented) possibilities for action. Applied ethics deals with real-life situations, where decisions have to be made under time pressure, and often limited rationality. AI Ethics is generally viewed as an example of applied ethics and focuses on the normative issues raised by the design, development, implementation, and use of AI. Within ethical discussions, the terms "moral" and "ethical" are often used. The term "moral" refers to the concrete, factual patterns of behaviour, the customs, and conventions that can be found in specific cultures, groups, or individuals at a certain time. The term "ethical" refers to an evaluative assessment of such concrete actions and behaviours from a systematic, academic perspective.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL:* https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

Of or relating to moral principles, especially as forming a system, or the branch of knowledge or study dealing with these. (OED)

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pd

---

Computer ethics is the analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology. I use the phrase "computer technology" because I take the subject matter of the field broadly to include computers and associated technology. For instance, I include concerns about software as well as hardware and concerns about networks connecting computers as well as computers themselves. A typical problem in computer ethics arises because there is a policy vacuum about how computer technology should be used. Computers provide us with new capabilities and these in turn give us new choices for action. Often, either no policies for conduct in these situations exist or existing policies seem inadequate. A central task of computer ethics is to determine what we should do in such cases, i.e., to formulate policies to guide our actions. Of course, some ethical situations confront us as individuals and some as a society. Computer ethics includes consideration of both personal and social policies for the ethical use of computer technology (Moor 1985, 266).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Ethical AI

Term used to indicate the development, deployment and use of AI that ensures compliance with ethical norms, including fundamental rights as special moral entitlements, ethical principles, and related core values. It is the second of the three core elements necessary for achieving Trustworthy AI.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL:* https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

Evasion (Model Evasion)

Evasion is one of the most common attacks on Machine Learning models (ML) performed during production. It refers to designing an input, which seems normal for a human but is wrongly classified by ML models. A typical example is to change some pixels in a picture before uploading, so that the image recognition system fails to classify the result.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Expert System

An interactive computer program that asks the same questions a human expert would ask, and from the information given to it by the user, (attempts to) provide the same answer the expert would provide (Quinn, 1990).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

An expert system consists of three main pairs: 1. Knowledge base. The actual information in the expert system. 2. Inference engine. The name given to the software that makes the expert system work. The software works with input data supplied by the user to search the knowledge base in order to reach a conclusion. 3. User interface. Screens and or menus through which the expert system communicates with users (Duval and Main 1994, 44).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

AI system that encapsulates knowledge provided by a human expert in a specific domain to infer solutions to problems.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Explainability

Feature of an AI system that is intelligible to non-experts. An AI system is intelligible if its functionality and operations can be explained non technically to a person not skilled in the art.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Property of an AI system to express important factors influencing the AI system results in a way that humans can understand.

Note 1 to entry: It is intended to answer the question “Why?” without actually attempting to argue that the course of action that was taken was necessarily optimal.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

It refers to methods and techniques in AI such that the results of the solution can be understood by humans. It contrasts with the concept of the "black box" in machine learning where even its designers cannot explain why the AI arrived at a particular decision. An AI system is intelligible if its functionality and operations can be explained non technically to a person not skilled in the art.

Source: *Edwards, Lilian; Veale, Michael (2017). «Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For». Duke Law and Technology Review 16: 18.*

Explicit programming

Specific implementation according to a set of step-by-step instructions from input to output.

Source: *Own elaboration.*

Exploding gradient

Phenomenon of backpropagation training (3.2.21) in neural networks where large error gradients accumulate and result in very large updates to the weights, making the model (3.1.26) unstable.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

3.6 F

Fairness

Fairness refers to a variety of ideas known as equity, impartiality, egalitarianism, non-discrimination and justice. Fairness embodies an ideal of equal treatment between individuals or between groups of individuals. This is what is generally referred to as 'substantive' fairness. But fairness also encompasses a procedural perspective, that is the ability to seek and obtain relief when individual rights and freedoms are violated.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Fault tolerance

Fault tolerance is the property that enables a system to continue operating properly in the event of the failure of (or one or more faults within) some of its components. If its operating quality decreases at all, the decrease is proportional to the severity of the failure, as compared to a naively designed system, in which even a small failure can cause total breakdown. Fault tolerance is particularly sought after in high-availability or safety critical systems. Redundancy or duplication is the provision of additional functional capabilities that would be unnecessary in a fault-free environment. This can consist of backup components that automatically 'kick in' if one component fails.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Feed Forward Neural Network (FFNN)

Neural network where information is fed from the input layer to the output layer in one direction only.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

3.7 G

General AI (AGI)

Also referred to as strong AI, AGI is the hypothetical ability of an intelligent system that can successfully understand, learn and perform any intellectual task that a human being can.

Source: *Own elaboration.*

<system> AI that addresses a broad range of tasks with a satisfactory level of performance

Note 1 to entry: Compared to narrow AI.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

General Purpose AI (GPAI) / Foundational Models

Also referred to as "foundational models". Large AI models trained on a vast quantity of data (generally unlabeled data and using self-supervision learning at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks.

Source: *Own elaboration adapted from the Center for Research on Foundation Models (CRFM), "On the Opportunities and Risks of Foundation Models". URL:* https://crfm.stanford.edu/report.html

General purpose AI systems are AI systems that have a wide range of possible uses, both intended and unintended by the developers. They can be applied to many different tasks in various fields, often without substantial modification and fine-tuning. Current general purpose AI systems are characterised by their scale (a lot of memory, data and powerful hardware) as well as their reliance on transfer learning (applying knowledge from one task to another).

These systems are sometimes referred to as "foundation models" and are characterised by their widespread use as pre-trained models for other, more specialised AI systems. For example, a single general purpose AI system for language processing can be used as the foundation for several hundred applied models (e.g. chatbots, ad generation, decision assistants, spambots, translation, etc.), some of which can then be further fine-tuned into a number of applications tailored to the customer.

Source: *Future of Life Institute (FLI). "General Purpose AI and the AI Act", May 2022. URL:* https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf

Genetic Algorithm

Algorithm which simulates natural selection by creating and evolving a population of individuals (solutions) for optimization problems.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Governance

The process of collective decision-making and policy implementation, used distinctly from government to reflect broader concern with norms and processes relating to the delivery of public goods (Brown, 2018).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Governance consists of the traditions and institutions by which authority in a country is exercised. This includes the process by which governments are selected, monitored and replaced; the capacity of the government to effectively formulate and implement sound policies; and the respect of citizens and the state for the institutions that govern economic and social interactions among them. (World Bank 2017).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Ground Truth

Value of the target variable for a particular item of labelled input data.

Note 1 to entry: Ground truth is not always the same as absolute truth.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning.* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

3.8 H

Heteronomy / Heteronomous

Characteristic of a system operating under the constraint of external intervention, control or oversight.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

High displacement potential

Displacement potential attributed to AI systems that perform tasks that use clearly defined processes and outputs (e.g. tasks performed by clinical lab technicians, optometrists, chemical engineers, actuaries, credit analysts, accountants, operations research analysts, concierges, mechanical drafters, brokerage clerks and quality control inspectors). This does not imply a high likelihood of being replaced by AI. That would require a more complex assessment of the technical feasibility and context of the task to be performed.

Source: *OECD Framework for the Classification of AI systems. URL:* https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1645798047&id=id&accname=guest&checksum=1F40FEDBAC6979FE2EF85A8DF4B481CE

Human-machine teaming

Integration of human interaction with machine intelligence capabilities.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Human oversight

Human oversight helps ensure that an AI system does not undermine human autonomy or causes other adverse effects.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

The capability for human intervention in every decision cycle of the system (human-in-the-loop), or during the design cycle of the system and monitoring the system's operation (human-on-the-loop), or the capability to oversee the overall activity of the system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation (human-in-command). Human oversight can include the decision not to use the system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the

ability to override a decision made by a system. Oversight mechanisms can be required in varying degrees to support other safety and control measures, depending on the system's application area and potential risk.

Source: *Based on HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

AI system that encapsulates knowledge provided by a human expert in a specific domain to infer solutions to problems.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Human-centric AI

The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. This also entails consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Human-in-the-loop

Human-in-the-loop (HITL) is one of the governance mechanisms addressed by human oversight. HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Human-on-the-loop

Human-on-the-loop (HOTL) is one of the governance mechanisms addressed by human oversight. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Human-in-command

Human-in-command is one of the governance mechanisms addressed by human oversight. It refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Hyperparameter

<machine learning> Characteristic of a machine learning algorithm that affects its learning process.

Note 1 to entry: Hyperparameters are selected prior to training and can be used in processes to help estimate model parameters.

Note 2 to entry: Examples of hyperparameters include number of network layers, width of each layer, type of activation function, optimization method, learning rate for neural networks; the choice of kernel function in a support vector machine; number of leaves or depth of a tree; the K for K-means clustering; the maximum number of iterations of the expectation maximization algorithm; the number of Gaussians in a Gaussian mixture.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

3.9 I

Implementation

The process of adopting, integrating, carrying out, executing, or practicing a plan, a method, a system or any design, idea, model, specification, standard or policy for doing something.

Source: *Own elaboration.*

Putting a plan or policy into action.

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Imputation

Procedure where missing data are replaced by estimated or modelled data.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Individually Identifiable Data

Information which can be linked to a single person.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Individually Identifiable Data is data that identifies the person that the data is about, or that can be used to identify that individual. This generally refers to data that contains either an identification number, or factors relating to physical, mental, economic, cultural, or social identity that could be used to link the data to an individual. Regulatory requirements for privacy generally apply (only) to individually identifiable data (Clifton 2009, 1471-1472).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

The term “personally identifiable information” refers to information which can be used to distinguish or trace an individual's identity, such as their name, Social Security Number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother’s maiden name, etc. (iDASH no date).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Inference

Reasoning by which conclusions are derived from known premises.

Note 1 to entry: In artificial intelligence, a premise is either a fact, a rule, a model, a feature, or raw data.

Note 2 to entry: The term "inference" refers both to the process and its result.

Note 3 to entry: inference: term and definition standardised by ISO/IEC [ISO/IEC 2382-28:1995].

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Information

Statements that carry meaning.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Information Retrieval

Task of retrieving relevant documents or parts of documents from a dataset, typically based on keywords or natural language queries.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Internet of Things (IoT)

Infrastructure of interconnected entities, people, systems and information resources together with services which process and react to information from the physical world and virtual world.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

IoT device

Entity of an IoT system that interacts and communicates with the physical world through sensing or actuating.

Note 1 to entry: An IoT device can be a sensor or an actuator.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Input data

Data provided to or directly acquired by an AI system on the basis of which the system produces an output.

Source: *EU Artificial Intelligence Act URL:* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Intelligence

The faculty of understanding; intellect. Also as a count noun: a mental manifestation of this faculty, a capacity to understand (OED) "Intelligence measures an agent's ability to achieve goals in a wide range of environments." S. Legg and M. Hutter (for a review of 70+definitions, See Legg and Hutter 2007).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Intelligent systems are expected to work, and work well, in many different environments. Their property of intelligence allows them to maximize the probability of success even if full knowledge of the situation is not available. Functioning of intelligent systems cannot be considered separately from the environment and the concrete situation including the goal (R. R. Gudwin).

Intelligence is the ability to process information properly in a complex environment. The criteria of properness are not predefined and hence not available beforehand. They are acquired as a result of the information processing (H. Nakashima).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Intelligent Agent

An autonomous entity capable of successfully adapting to its environment by effecting its own will.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Intelligent agents continuously perform three functions: perception of dynamic conditions in the environment; action to affect conditions in the environment; and reasoning to interpret perceptions, solve problems, draw inferences, and determine actions (Hayes-Roth, 1995).

Intelligent agents are software entities that carry out some set of operations on behalf of a user or another program with some degree of independence or autonomy, and in so doing, employ some knowledge or representation of the user's goals or desires" (IBM quoted in Franklin and Graesser 1996, 23). Intelligence is the ability to process information properly in a complex environment. The criteria of properness are not predefined and hence not available beforehand. They are acquired as a result of the information processing. H. Nakashima

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Interpretability

Interpretability refers to the concept of comprehensibility, explainability, or understandability. When an element of an AI system is interpretable, this means that it is possible at least for an external observer to understand it and find its meaning.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

(AI) Models are interpretable when humans can readily understand the reasoning behind predictions and decisions made by the model.

The more interpretable the models are, the easier it is for someone to comprehend and trust the model.

Models such as deep learning and gradient boosting are not interpretable and are referred to as black-box models because they are too complex for human understanding. It is impossible for a human to comprehend the entire model at once and understand the reasoning behind each decision.

Source: *Interpretable AI. URL:* *https://www.interpretable.ai/interpretability/what*

Inversion (Model Inversion)

Model inversion refers to a kind of attack to AI models, in which the access to a model is abused to infer information about the training data. So, model inversion turns the usual path from training data into a machine-learned model from a one-way one to a two-way one, permitting the training data to be estimated from the model with varying degrees of accuracy. Such attacks raise serious concerns given that training data usually contain privacy-sensitive information.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Impacted Stakeholders

Impacted stakeholders can be indirectly or directly affected by the deployment of an AI system or application but do not necessarily interact with the system. An AI system or application can impact several different stakeholder groups.

Source: *OECD Framework for the Classification of AI systems. URL:* *https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1645798047&id=id&accname=guest&checksum=1F40FEDBAC6979FE2EF85A8DF4B481CE*

3.10 J

3.11 K

Knowledge-based system (KBS)

A knowledge-based system (KBS) is a type of Artificial Intelligence (AI) that provides logical reasoning capabilities on knowledge to solve an application problem. It is typically made of a knowledge base that represents knowledge explicitly and an inference engine that generates outputs by reasoning on the knowledge base.

Source: *Own elaboration, adapted from multiple sources.*

3.12 L

Label

A marker, description, or tag, characterising and identifying raw data (images, text files, videos, etc.) and providing context so that an AI system can learn from it.

Source: *Own elaboration*.

Law (scientific)

In general, a scientific law is the description of an observed phenomenon. It doesn't explain why the phenomenon exists or what causes it. The explanation of a phenomenon is called a scientific theory (Bradford 2017).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

An axiomatic statement.

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Law Enforcement

Law enforcement means activities carried out by law enforcement authorities for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security.

Source: *EU Artificial Intelligence Act. URL:* *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206*

Legal Personhood

An individual who has legal status with a state, such as citizenship. “The function of legal personhood is to attribute value and rights to the individual” (Dyschkant 2015, 2107).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

While there is disagreement about how precisely to formulate a definition of legal personhood, the key element of legal personhood seems to be the ability to bear rights and duties. Black’s Law Dictionary defines a legal person as an entity given certain legal rights and duties of a human being; a being, real or imaginary, who for the purpose of legal reasoning is treated more or less as a human being (Dyschkant 2015, 2076).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Lifecycle

The lifecycle of an AI system includes several interdependent phases ranging from its design and development (including sub-phases such as requirement analysis, data collection, training, testing, integration), installation, deployment, operation, maintenance, and disposal. Given the complexity of AI (and in general information) systems, several models and methodologies have been defined to manage this complexity, especially during the design and development phases, such as waterfall, spiral, agile software development, rapid prototyping, and incremental.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

An AI system's life cycle encompasses its development (including research, design, data provision, and limited trials), deployment (including implementation) and use phase.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Evolution of a system, product, service, project or other human-made entity, from conception through retirement.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Long Short-Term Memory (LSTM)

Type of recurrent neural network that processes sequential data with a satisfactory performance for both long and short span dependencies.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Low displacement potential

Displacement potential attributed to AI systems perform tasks that require reasoning about novel situations (e.g. research), interpersonal skills (e.g. teachers and managers, some baristas) and physical occupations that require perception and manipulation of a plurality of irregular objects in uncontrolled environments with limited room for mobility (e.g. maids, cleaners, cafeteria attendants, hotel porters, roofers and painters, massage therapists, plasterers and stucco masons). This does not necessarily mean that the occupation will not see significant automation of key tasks.

Source: *OECD Framework for the Classification of AI systems. URL: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1645798047&id=id&accname=guest&checksum=1F40FEDBAC6979FE2EF85A8DF4B481CE*

3.13 M

Machine Learning

Machine Learning is a branch of artificial intelligence (AI) and computer science which focuses on development of systems that are able to learn and adapt without following explicit instructions imitating the way that humans learn, gradually improving its accuracy, by using algorithms and statistical models to analyse and draw inferences from patterns in data.

Source: *Own elaboration.*

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that focuses on the development of systems capable of learning from data to solve an application problem without being explicitly programmed. Learning refers to the computational process of optimizing model parameters from data, according to a given criteria. The model is a mathematical construct that generates an output based on input data.

Source: *Adapted from multiple terms from ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Machine Learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data.

Source: *Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892*

Process of optimizing model parameters (3.1.28) through computational techniques, such that the model's behaviour reflects the data or experience.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Machine Learning algorithm

Algorithm to establish parameters, according to a given criteria, of a machine learning model from data.

EXAMPLE: Consider solving a univariate linear function y = θ0 + θ1x where y is an output, or result, x is an input, θ0 is an intercept (the value of y where x=0) and θ1 is a weight. In machine learning, the process of determining the intercept and weights for a linear function is known as linear regression.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Machine Learning model

Mathematical construct that generates an inference, or prediction, based on input data.

Note 1 to entry: A machine learning model results from training based on a machine learning algorithm.

EXAMPLE: If a univariate linear function (y = θ0 + θ1x) has been trained using linear regression, the resulting model can be y = 3 + 7x.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Machine Translation (MT)

Automated translation of text or speech from one natural language to another using a computer system.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Maleficent

Acts intentionally taken to promote evil or confound good.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Malfeasant

Acts intentionally taken by persons or organizations in a position of power to promote evil or confound good.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Malfeasance is failure of officials to faithfully execute their duties, whether as enforcement of rightful law or policy, chiefly for their own gain in funds or leisure (Becker and Stigler 1974).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Metadata

Data that provides information about different aspects of other data, but not the content of the data. It is used to summarise basic information about the data that can facilitate working with the data. There are many distinct types of metadata, including descriptive, structural, administrative, reference, statistical, legal, etc.

Source: *Own elaboration, adapted from Wikipedia https://en.wikipedia.org/wiki/Metadata*

Data about data or data elements, possibly including their data descriptions, and data about data ownership, access paths, access rights and data volatility.

Source: *ISO/IEC 20546:2019(en) Information technology - Big data - Overview and vocabulary. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:20546:ed-1:v1:en*

Methodology

Methodology is defined as the research strategy that outlines the way one goes about undertaking a research project, whereas methods identify means or modes of data collection (Howell 2012, viii).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

OECD glossary of statistical terms defines methodology as a structured approach to solve a problem.

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Mind

A person's cognitive, rational, or intellectual powers; the intellect; esp. as distinguished from the emotions; a person of intellectual prowess; an intellectual combination of the neural architecture and effects of the transmissions of this architecture on the formation of emotions, mental representations, correspondences between sensation and mental representations of that which is sensed, computation of internal and external data, and decisions, plans and intentions made on the basis of the unity of all of these.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

According to a Classical Computational Theory of Mind, the mind is a computational system similar in important respects to a Turing machine, and core mental processes (e.g., reasoning, decision- making, and problem solving) are computations similar in important respects to computations executed by a Turing machine (Rescorla 2015).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Mitigation

Plan to lessen the impact of a harm.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Risk mitigation planning is the process of developing options and actions to enhance opportunities and reduce threats to project objectives. Risk mitigation implementation is the process of executing risk mitigation actions. Risk mitigation progress monitoring includes tracking identified risks, identifying new risks, and evaluating risk process effectiveness throughout the project (Project Management Institute 2008).

Source: *(Engineering) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Mixed Reality

A type of virtual reality system.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

The most straightforward way to view a Mixed Reality environment, therefore, is one in which real world and virtual world objects are presented together within a single display, that is, anywhere between the extremes of the virtuality continuum (Milgram and Kishino 1994).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Model

In AI, this keyword mostly refers to a machine learning model or statistical model, which can make predictions/decisions over data. So only a subset of algorithms are models.

Source: *Own elaboration.*

Physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, process or data

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Moral

Thought and discourse about moral questions; moral philosophy, ethics (OED); Pertaining to the meaning of good and evil and establishment of ethical standards to foster those Meanings.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

A moral Turing test (MTT) might similarly be proposed to bypass disagreements about ethical standards by restricting the standard Turing test to conversations about morality. If human "interrogators" cannot identify the machine at above chance accuracy, then the machine is, on this criterion, a moral agent" (Allen et al 2000, quoted in Arnold and Schuetz 2016, 104).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Moral Agent

An agent able to define and implement their meaning of good and evil.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

A suitably generic characterization might be that a moral agent is an individual who takes into consideration the interests of others rather than acting solely to advance his, her, or its self-interest (Allen et al 2000, 252).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Cua defines moral agents with respect to the principle of impartiality. As moral agents, the principle of autonomy appears to be the basis for applying the principle of impartiality, for in the notion of balance implicit in the moral point of view it is suggested that the interests of all individuals in dispute have an equal claim to respect in adjudication. Unless morality is to be viewed primarily as a product of external factors, every moral agent is entitled to administer its function so long as the principle of impartiality is applied and maintained (Cua 1967, 164-165).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Moral Autonomy

Cognitive capacity to self-define the meaning of good and evil, with or without the ability to fully act upon it.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

An artificial system's achievement to pass the moral Turing test.

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Moral Norms

Perceptions about the moral correctness or wrongness of actions that have been codified by a community into standards against which behaviours are judged, praised or punished; standards which pertain to the meaning of good and evil and are held as such by a community.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

3.14 N

Named Entity Recognition (NER)

Task of recognizing and labelling the denotational names of entities and their categories for sequences of words in a stream of text or speech.

Note 1 to entry: Entity refers to concrete or abstract thing of interest, including associations among things.

Note 2 to entry: "Named entity" refers to an entity with a denotational name where a specific or unique meaning exists.

Note 3 to entry: Denotational names include the specific names of persons, locations, organizations, and other proper names based on the domain or application.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Narrow AI

Term used to describe AI systems that are specified to handle a singular or limited task. Many currently existing AI systems that are likely operating as a narrow AI focused on a specific problem. For instance, digital assistants are all examples of narrow AI as they operate within a limited pre-defined range of functions.

Source: *Own elaboration.*

<system> AI that is focused on defined tasks to address a specific problem.

Note 1 to entry: Compared to general AI.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Natural Language

Language which is or was in active use in a community of people, and the rules of which are mainly deduced from the usage.

Note 1 to entry: Natural language is any human language, which can be expressed in text, speech, sign language etc.

Note 2 to entry: Natural language is any human language, such as English, Spanish, Arabic, Chinese, or Japanese, to be distinguished from programming and formal languages, such as Java, Fortran, C++, or First-Order Logic.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Natural Language Generation (NLG)

Task of converting data carrying semantics into natural language.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Natural Language Processing (NLP)

<system> Information processing based upon natural language understanding and natural language generation.

<discipline> Discipline concerned with the way computers process natural language data.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Natural Language Understanding (NLU) / Natural Language Comprehension

Extraction of information, by a functional unit, from text or speech communicated to it in a natural language, and the production of a description for both the given text or speech, and what it represents.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Neural Network (NN) / Artificial Neural Network (ANN)

Network of two or more layers of neurons connected by weighted links with adjustable weights, which takes input data and produces an output.

Note 1 to entry: Whereas some neural networks are intended to simulate the functioning of biological neurons in the nervous system, most neural networks are used in artificial intelligence as realizations of the connectionist model.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Neuron

<artificial intelligence> Primitive processing element which takes one or more input values and produces an output value by combining the input values and applying an activation function on the result.

Note 1 to entry: Examples of nonlinear activation functions are a threshold function, a sigmoid function and a polynomial function.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

No displacement potential

Some AI systems execute tasks that could not be performed by humans with the same accuracy, specificity or scale (e.g. AI systems used in cybersecurity and threat detection).

Source: *OECD Framework for the Classification of AI systems. URL: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1645798047&id=id&accname=guest&checksum=1F40FEDBAC6979FE2EF85A8DF4B481CE*

Norms

That which is a model or a pattern; a type, a standard; A value used as a reference standard for purposes of comparison.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

In mathematics, norms are functions assigning a strictly positive length or size to each vector in a vector space (other than zero vectors).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

A collective evaluation of behaviour in terms of what it ought to be; a collective expectation as to what behaviour will be; and/or particular reactions to behaviour, including attempts to apply sanctions or otherwise induce a particular kind of conduct. (Gibbs 1965, 589)

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Normative System

A system based on what is established as the norm (OED); Organised parameters of action designed to promote good.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Normative systems include systems of law, abstract models of computer systems, and hybrid systems consisting of human and computer agents in interaction (Jones and Sergot 1993, 275).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

3.15 O

Online (machine) learning / Batch learning

A machine learning technique in which data becomes available in a sequential order and is used to update the model parameters at each training step (one sample at a time), as opposed to batch learning techniques which use the entire training data set or batches of the training data set to update the model parameters at each training step.

Source: *Own elaboration.*

Online Interface

Means any software, including a website or a part thereof, and applications, including mobile applications.

Source: *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. URL: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services*

Online Platform

Means a provider of a hosting service which, at the request of a recipient of the service, stores and disseminates to the public information, unless that activity is a minor and purely ancillary feature of another service and, for objective and technical reasons cannot be used without that other service, and the integration of the feature into the other service is not a means to circumvent the applicability of this Regulation.

Source: *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. URL: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services*

Ontology

The same ontological theory may commit to different conceptualizations, as well as the same conceptualization may underlie different ontological theories. The term "ontology" will be used ambiguously, either as synonym of "ontological theory" or as synonym of conceptualization".

Conceptualization: an intentional semantic structure which encodes the implicit rules constraining the structure of a piece of reality.

Formal Ontology: the systematic, formal, axiomatic development of the logic of all forms and modes of being.

Ontological commitment: a partial semantic account of the intended conceptualization of a logical theory.

Ontological engineering: the branch of knowledge engineering which exploits the principles of (formal) Ontology to build ontologies.

Ontological theory: a set of formulas intended to be always true according to a certain conceptualization.

Ontology: that branch of philosophy which deals with the nature and the organisation of reality.

Ontology: (sense 1) a logical theory which gives an explicit, partial account of a conceptualization; synonym of conceptualization (Guarino and Giaretta 1995).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

The study of what there is.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

3.16 P

Parameter / Model Parameter

<machine learning> Internal variable of a model that affects how it computes its outputs.

Note 1 to entry: Examples of parameters include the weights in a neural network, or the transition probabilities in a Markov model.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Patients

Agents who are acted upon by other agents.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Individuals who are treated by healthcare practitioners and whose data— Protected Health Information—is covered as Individually identifiable health information which means any information, including demographic information collected from an individual, that--(A) is created or received by a health care provider, health plan, employer, or health care clearinghouse; and (B) relates to the past, present, or future physical or mental health or condition of an individual, the provision of health care to an individual, or the past, present, or future payment for the provision of health care to an individual, and--(i) identifies the individual; or (ii) with respect to which there is a reasonable basis to believe that the information can be used to identify the individual (42 U.S.C. 1301.1171(6)).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Pen test

A penetration test, colloquially known as a pen test, pentest or ethical hacking, is an authorised simulated cyberattack on a computer system, performed to evaluate the security of the system. The test is performed to identify both weaknesses (also referred to as vulnerabilities), including the potential for unauthorised parties to gain access to the system's features and data, as well as strengths, enabling a full risk assessment to be completed.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Performance

Measurable result.

Note 1 to entry: Performance can relate either to quantitative or qualitative findings.

Note 2 to entry: Performance can relate to managing activities, processes, products (including services), systems or organizations.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Personal Data

Facts about an individual which may be used to identify them.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

"Personal data" means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (General Data Protection Regulation, Article 4.1).

"Sensitive Personal Data" are personal data, revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership; data concerning health or sex life and sexual orientation; genetic data or biometric data" (General Data Protection Regulation, Article 8.1).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Persuasion

The action or an act of persuading or attempting to persuade; the addressing of arguments or appeals to a person in order to induce cooperation, submission, or agreement; the presenting of persuasive reasoning or compelling arguments.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

The process by which agent action becomes social structure, ideas become norms, and the subjective becomes the intersubjective (Finnemore and Sikkink, 1998: 914).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

See Persuasive technology.

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Persuasive Technology

(Also known as "Captology") Software systems, which may or may not be integrated with specialised hardware, designed to change the behaviors or attitudes of end users in order to achieve a desirable end.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Captology focuses on the planned persuasive effects of computer technology. Computers function as a tool or instrument to increase capabilities in order to reduce barriers, increase self-efficacy, provide information for better decision making, change mental models; Computers function as a medium to provide experiences in order to provide first-hand learning, insight, visualization and resolve, and to promote understanding of cause-and-effect relationships. Computers function as social actors to create relationships in order to establish social norms, invoke social rules and dynamics, and provide social support or sanction (Fogg, Cuelar and Danielson 2009, 110; 116).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Planning

<artificial intelligence> Computational processes that compose a workflow out of a set of actions, aiming at reaching a specified goal.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

---

Field of Artificial Intelligence which explores the process of using autonomous techniques to solve planning and scheduling problems. A planning problem is one in which we have some initial starting state, which we wish to transform into a desired goal state through the application of a set of actions.

Source: *Planning. Wiki - The AI Planning & PDDL Wiki. URL:* https://planning.wiki/

Policy

A strategy used to pursue some goals. The policy dictates the actions to be taken as a function of the states of the system and the environment.

Source: *Own elaboration.*

Authoritative plans of action.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

A guide to action to change what would otherwise occur; a decision about amounts and allocations of resources; a statement of commitment to certain areas of concern; the distribution of the amount shows the priorities of decision makers. Public policy is policy at any level of government (Porta 2016).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Predictability

Property of an AI system that enables reliable assumptions by stakeholders about the output.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Prediction

<machine learning> Output of a machine learning model when provided with input data.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Principles

A fundamental source from which something proceeds; A primary element, force, or law which produces or determines particular results.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Principles such as the Church-Turing Principle, are statements that may be testable hypotheses or axioms used in computation (Deutsch 1985).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Privacy

The protection of select information through the use of mechanical or statistical masking mechanisms for the purpose of protecting individual or group dignity, desire for seclusion or concealment, property, secrets, or freedom of choice.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Freedom from surveillance (see Lyon and Zureik 1996).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

One aspect of privacy is the withholding or concealment of information (Posner 1977, 393).

Bostwick gives a typology of privacy as: "the privacy of repose, the privacy of sanctuary, and the privacy of intimate decision. Repose means peace, quiet, and calm for the individual protected. Sanctuary means prohibiting other persons from seeing, hearing, and knowing. The zone of intimate decision is an area within which the personal calculus used by an individual to make fundamental decisions must be allowed to operate without the injection of disruptive factors by the state. This privacy is less "freedom from" and more "freedom to" (Bostwick 1976).

The OECD Privacy Framework Privacy Principles include: collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability.

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Proprietary

Owned as property.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

A protocol confined to a particular proprietary set of software or hardware. This is in contrast to Internet protocols which are completely open (Ince 2013).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Proprietary capacity means the capacity or interest of a producer or handler that, either directly or through one or more intermediaries, is a property owner together with all the appurtenant rights of an owner including the right to vote the interest in that capacity as an individual, a shareholder, member of a cooperative, partner, trustee or in any other capacity with respect to any other business unit.

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

3.17 Q

3.18 R

Recommender System

A recommender system is a type of information retrieval (IR) system whose goal is to suggest items from a large collection that meets the preference of a user.

Source: *F. Ricci, L. Rokach, B. Shapira, P. Kantor, Recommender Systems Handbook, Springer, 2011. doi:10.1007/978-0-387-85820-3*

Fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service, including as a result of a search initiated by the recipient or otherwise determining the relative order or prominence of information displayed.

Source: *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. URL:* https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services

Recurrent Neural Network (RNN)

A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed or undirected graph along a temporal sequence. This allows it to exhibit temporal dynamic behaviour. Derived from feedforward neural networks, RNNs can use their internal state (memory) to process variable length sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition.

Source: *Dupond, Samuel (2019). "A thorough review on the current advance of neural network structures". Annual Reviews in Control. 14: 200–230.*

Abiodun, Oludare Isaac; Jantan, Aman; Omolara, Abiodun Esther; Dada, Kemi Victoria; Mohamed, Nachaat Abdelatif; Arshad, Humaira (2018-11-01). "State-of-the-art in artificial neural network applications: A survey". Heliyon. 4 (11): e00938. doi:10.1016/j.heliyon.2018.e00938. ISSN 2405-8440. PMC 6260436. PMID 30519653.

Neural network in which outputs from both the previous layer and the previous processing step are fed into the current layer.

Source: *ISO/IEC DIS 22989(en). Terms related to Neural Networks. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Red team

Red teaming is the practice whereby a red team or independent group challenges an organisation to improve its effectiveness by assuming an adversarial role or point of view. It is often used to help identify and address potential security vulnerabilities.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

Red teaming is the practice whereby a "red team" or independent group challenges an organisation to improve its effectiveness by assuming an adversarial role or point of view. It is particularly used to help identifying and addressing potential security vulnerabilities.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL:* https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

Reinforcement Learning

Machine Learning utilizing a reward function to optimize either a policy function or a value function by sequential interaction with an environment.

Note 1 to entry: Policy functions and value functions express a strategy that is learned by the environment.

Note 2 to entry: The environment can be any stateful model.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Reliability

Property of consistent intended behaviour and results.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Repeatability

Same team, same experimental setup. The measurement can be obtained with stated precision by the same team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials. For computational experiments, this means that a researcher can reliably repeat her own computation.

Source: *Association for Computing Machinery (ACM). URL:* https://www.acm.org/publications/policies/artifact-review-badging

Replicability

Different team, same experimental setup. The measurement can be obtained with stated precision by a different team using the same measurement procedure and the same measuring system, under the same operating conditions, in the same or a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using the author's own artifacts.

Source: *Association for Computing Machinery (ACM). URL:* https://www.acm.org/publications/policies/artifact-review-badging

Representative Learning

Learning representations of the data that make it easier to extract useful information when building classifiers or other predictors.

Source: *Bengio; A. Courville; P. Vincent (2013). "Representation Learning: A Review and New Perspectives". IEEE Transactions on Pattern Analysis and Machine Intelligence. 35 (8): 1798–1828. arXiv:1206.5538. doi:10.1109/tpami.2013.50. PMID 23787338.*

In Machine Learning, feature learning or representation learning is a set of techniques that allows a system to automatically discover the representations needed for feature detection or classification from raw data. This replaces manual feature engineering and allows a machine to both learn the features and use them to perform a specific task.

Source: *Wikipedia. URL: https://en.wikipedia.org/wiki/Feature_learning*

Reproducibility

Reproducibility refers to the closeness between the results of two actions, such as two scientific experiments, that are given the same input and use the methodology, as described in a corresponding scientific evidence (such as a scientific publication). A related concept is replication, which is the ability to independently achieve non-identical conclusions that are at least similar, when differences in sampling, research procedures and data analysis methods may exist. Reproducibility and replicability together are among the main tools of the scientific method.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Reproducibility describes whether an AI experiment exhibits the same behaviour when repeated under the same conditions.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Different team, different experimental setup. The measurement can be obtained with stated precision by a different team and a different measuring system, in a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using artifacts that they develop completely independently.

Source: *Association for Computing Machinery (ACM). URL: https://www.acm.org/publications/policies/artifact-review-badging*

Resilience

Ability of a system to recover operational condition quickly following an incident.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Research

Systematic inquiry into real phenomena.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any particular application or use in view (OECD Glossary of Statistical Terms 2017).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported under a program which is considered research for other purposes.

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Responsibility

Capability of fulfilling an obligation or duty; The quality of being reliable or trustworthy; the state or fact of being accountable for actions; liability for some action.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

A government that is responsive to public opinion, that pursues policies that are prudent and mutually consistent, and that is accountable to the representatives of the electors (Grant 2016).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Retraining

Updating a trained model by training with different training data.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Reward function

A function that maps each perceived state (or state-action pair) of the environment to a single number, a reward, indicating the intrinsic desirability of that state.

Source: *Source: "Reinforcement Learning: An Introduction", Richard S. Sutton and Andrew G. Barto. A Bradford Book. The MIT Press. Cambridge, Massachusetts. London, England. URL: http://www.incompleteideas.net/book/ebook/the-book.html*

Rights

That which is considered proper, correct, or consonant with justice, and related uses; The standard of permitted and forbidden action within a particular sphere.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Legal or moral recognition of choices or interests to which particular weight is attached. Very often, statements about rights draw on more than one of the four relations identified (Reeve 2016).:

1. A right is a liberty: a person has a liberty to X means that he has no obligation not to X.
2. A right is a right 'strictly speaking' or a claim right: a person has a right to X means others have a duty to him in respect of X.

3. A right is a power, that is, the capacity to change legal relations (and others are liable to have their position altered).
4. A right is an immunity, that is the absence of the liability to have the legal position altered

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Risk

Possible loss or harm.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Risk exposure is [equal to] the probability of an unsatisfactory outcome and the loss to the parties affected if the outcome is unsatisfactory (Boehm 1991, 33).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Effect of uncertainty on objectives.

Note 1 to entry: An effect is a deviation from the expected. It can be positive, negative or both, and can address, create or result in opportunities and threats.

Note 2 to entry: Objectives can have different aspects and categories, and can be applied at different levels.

Note 3 to entry: Risk is usually expressed in terms of risk sources, potential events, their consequences and their likelihood.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Robot

Automation system with actuators that performs intended tasks in the physical world, by means of sensing its environment and a software control system.

Note 1 to entry: A robot includes the control system and interface of a control system.

Note 2 to entry: The classification of robot into industrial robot or service robot is done according to its intended application.

Note 3 to entry: In order to properly perform its tasks, a robot makes use of different kinds of sensors to confirm its current state and perceive the elements composing the environment in which it operates.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Robotics

Science and practice of designing, manufacturing, and applying robots.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Robustness

Ability of a system to maintain its level of performance under any circumstances.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

---

Robustness refers to the ability of an algorithm or system to deal with execution errors, erroneous inputs, or unseen data.

Source: *Own elaboration.*

Robustness AI

Robustness of an AI system encompasses both its technical robustness (appropriate in a given context, such as the application domain or life cycle phase) and as well as its robustness from a social perspective (ensuring that the AI system duly takes into account the context and environment in which the system operates). This is crucial to ensure that, even with good intentions, no unintentional harm can occur. Robustness is the third of the three components necessary for achieving Trustworthy AI.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

3.19 S

Safety

Prevention of accidents.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

AI safety is described as mitigating accident risks from machine learning. "The problem of accidents in machine learning systems. We define accidents as unintended and harmful behaviour that may emerge from machine learning systems when we specify the wrong objective function, are not careful about the learning process, or commit other machine learning related implementation errors" (Amodei et al 2016, 1-2).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Freedom from unacceptable risk.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Sample

Atomic data element processed in quantities by a machine learning algorithm.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Self-learning AI system / Self-supervised learning

Self-learning (or self-supervised learning) AI systems recognize patterns in the training data in an autonomous way, without the need for supervision.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Learning from an internal knowledge base, or from new input data, without introduction of explicit external knowledge.

Source: *ISO/IEC 2382-31:1997(en). Information technology — Vocabulary — Part 31: Artificial intelligence — Machine learning. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-31:ed-1:v1:en*

Semantic Computing

Processing that aims to understand the intentions of users and the meanings of information and to express them in a machine processable form.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Semi-supervised Machine Learning

Machine learning that makes use of both labelled and unlabelled data during training.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Sentiment Analysis

Task of computationally identifying and categorizing opinions expressed in a piece of text, speech or image, to determine feeling and attitude from positive to neutral to negative.

Note 1 to entry: Examples of sentiments include approval, disapproval, positive toward, negative toward, agreement and disagreement.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Speech recognition / Speech-To-Text (STT)

Conversion, by a functional unit, of a speech signal to a representation of the content of the speech.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Speech Synthesis / Text-To-Speech (TTS)

Generation of artificial speech.

Source: *ISO/IEC DIS 22989(en). Terms related to Natural Language Processing. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Stakeholders

By stakeholders we denote all those that research, develop, design, deploy or use AI, as well as those that are (directly or indirectly) affected by AI – including but not limited to companies, organisations, researchers, public services, institutions, civil society organisations, governments, regulators, social partners, individuals, citizens, workers and consumers.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Any individual, group, or organization that can affect, be affected by, or perceive itself to be affected by a decision or activity.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Stakeholders encompass all organisations and individuals involved in or affected by AI systems, directly or indirectly.

Source: *OECD Framework for the Classification of AI systems. URL: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1645798047&id=id&accname=guest&checksum=1F40FEDBAC6979FE2EF85A8DF4B481CE*

Standards

Standards are norms designed by industry and/or Governments that set product or services' specifications. They are a key part of our society as they ensure quality and safety in both products and services in international trade. Businesses can be seen to benefit from standards as they can help cut costs by improved systems and procedures put in place. Standards are internationally agreed by experts, and they usually represent what the experts think is the best way of doing something. It could be about making a product, managing a process, delivering a service or supplying materials – standards cover a huge range of activities. Standards are released by international organizations, such as ISO (International Organisation for Standardisation), IEEE (The Institute of Electrical and Electronics Engineers) Standard Association, and NIST (National Institute of Standards and Technology).

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Soft Computing

Processing that is tolerant of and exploits imprecision, uncertainty, partial truth and approximation in its input data and provides usable results with tractability, robustness and low solution cost.

Note 1 to entry: soft computing is a fusion of research in evolutionary algorithms, genetic programming, swarm intelligence, neural science, neural net systems, fuzzy set theory, fuzzy systems, probabilistic reasoning, chaos theory, chaotic systems.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Social Norms

Formal and informal rules defined by a social group.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Sociotechnical System

A social system operating on a technical base.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Subject

A subject is a person, or a group of persons, affected by the AI system (such as the recipient of benefits where the decision to grant or reject benefits is underpinned by an AI system, or the general public for facial recognition).

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Subsymbolic AI

AI <engineered system> based on techniques and models using a numeric representation and implicit information encoding.

Note 1 to entry: Compared to symbolic AI.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Superintelligence

The capacity to apprehend what is beyond the normal range of human intelligence or understanding; spiritual or paranormal insight or awareness, spiritualism. (OED)

Source: *(Ordinary Languege) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

An intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills (Bostrom, 2006,11).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Support Vector Machine (SVM)

Machine learning algorithm that finds decision boundaries with maximal margins.

Note 1 to entry: Support vectors are data points that define the positioning of the decision boundaries (hyper-planes).

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Sustainability

The Brundtland Report defines sustainable development as "Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs".

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

A sustainable system is one which survives or persists (Costanza and Patten 1995, p. 193).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Symbolic AI

Symbolic artificial intelligence is the collective name for all AI research methods that rely on high-level "symbolic" representations of problems, mathematical logic, and search. It is also known as GOFAI ("Good Old-Fashioned Artificial Intelligence"). Symbolic AI used tools such as logic programming, production rules, semantic nets and frames, and it developed applications such as expert systems. It was the dominant paradigm of AI research from the mid-1950s to the late 1980s. Later, more recent sub-symbolic approaches to AI were introduced, based on neural networks, statistics, numerical optimization and other techniques. Symbolic AI is still applied in some smaller domains (such as knowledge representation), but most AI applications in the 21st century do not employ readable symbols as their primary objects.

Source: *Own elaboration.*

AI <engineered system> based on techniques and models using symbols and structures.

Note 1 to entry: Compared to. subsymbolic AI.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

System

Integration of individual units into a purposive whole.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

A state of a system may be defined as an undisturbed motion that is restricted by as many conditions or data as are theoretically possible without mutual interference or contradiction (Dirac 1981, 11).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Socio-technical systems [are] arrangements of multiple purposive actors and material artifacts interacting in ways that require analysing the total system and not just the constituent subsystems. (Rophol 1999, quoted in Bauer and Herder 2004).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

3.20 T

Task

Actions required to achieve a specific goal.

Note 1 to entry: These actions can be physical or cognitive.

Note 2 to entry: Examples of tasks include classification, regression, ranking, clustering and dimensionality reduction.

Source: *ISO/IEC DIS 22989(en). Terms related to Artificial Intelligence. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Technical Norms

Parameters of action which a professional community has determined confer some benefit based upon their uses.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Technology

The branch of knowledge dealing with the mechanical arts and applied sciences; the study of this; The application of such knowledge for practical purposes, esp. in industry, manufacturing, etc.; the sphere of activity concerned with this; the mechanical arts and applied sciences collectively (OED); Application of scientific, mathematical, design, or engineering practices to creation of artifacts (SM-J).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Technology is the application of science, engineering and industrial organization to create a human-build world (Rhodes 1999, p. 19).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Any equipment or interconnected system or subsystem of equipment that is used in the automatic acquisition, storage, manipulation, management, movement, control, display, switching, interchange, transmission, or reception of data or information by the executive agency. For purposes of the preceding sentence, equipment is used by an executive agency if the equipment is used by the executive agency directly or is used by a contractor under a contract with the executive agency which 1) requires the use of such equipment; or 2) requires the use, to a significant extent, of such equipment in the performance of a service or the furnishing of a product. The term information technology includes computers, ancillary equipment, software, firmware and similar procedures, services (including support services), and related resources (NIST 2013).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Test

Testing is defined as assessment of the fitness of a product to achieve its stated goals.

Source: *((Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Models of software testing emphasize different testing goals.

Demonstration phase models test to make sure that the software satisfies its specification, while destruction phase models test to detect implementation faults. Life Cycle Evaluation models test to detect requirements, design and implementation faults while Life Cycle Prevention models test to prevent requirements, design and implementation faults (Gelperin and Hetzel 1988, 688).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Test data / Evaluation data

Data used to assess the performance of a final machine learning model.

Note 1 to entry: Test data is disjoint from training data and validation data.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Data used for providing an independent evaluation of the trained and validated AI system in order to confirm the expected performance of that system before its placing on the market or putting into service.

Source: *EU Artificial Intelligence Act URL : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206*

Traceability

Ability to track the journey of a data input through all stages of sampling, labelling, processing and decision making.

Note 1 to entry: Test data is disjoint from training data and validation data.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Traceability of an AI system refers to the capability to keep track of the system's data, development and deployment processes, typically by means of documented recorded identification.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1*

Trader

Any natural person, or any legal person irrespective of whether privately or publicly owned, who is acting, including through any person acting in his or her name or on his or her behalf, for purposes relating to his or her trade, business, craft or profession.

Source: *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. URL: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services*

Trained model

Result of model training.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Training

Goal oriented teaching, particularly to develop a skill.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Training data is a portion of data used to fit a model.

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

A training program is the method through which the State agency carries out a plan of educational and training activities to improve the operation of its programs.

a. Initial in-service training means a period of intensive, task-oriented training to prepare new employees to assume job responsibilities.

b. Continuing training means an on-going program of training planned to enable employees to: (1) Reinforce their basic knowledge and develop the required skills for the performance of specific functions, and (2) acquire additional knowledge and skill to meet changes such as enactment of new legislation, development of new policies, or shifts in program emphasis.

c. Full-time training means training that requires employees to be relieved of all responsibility for performance of current work to participate in a training program.

d. Part-time training means training that allows employees to continue full time in their jobs or requires only partial reduction of work activities to participate in a training program outside of the State or local agency.

e. Long-term training means training for eight consecutive work weeks or longer.

f. Short-term training means training for less than eight consecutive work weeks.

(45 CFR 235.61 - Definition of terms).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Process to establish or to improve the parameters of a machine learning model, based on a Machine Learning algorithm, by using training data.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Training data

Subset of input data samples used to train a Machine Learning model.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en

Data used for training an AI system through fitting its learnable parameters, including the weights of a neural network.

Source: *EU Artificial Intelligence Act URL :* *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206*

Transparency

Easily seen through, recognised, understood, or detected (OED); Sufficient illumination to confer comprehension.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Transparency is a characteristic which describes a process whereby information is requested and then disclosed completely within the limits of public law, without distortion, and with respect to the computational and cognitive capacities of the information recipient in order to enable those recipients to interpret the information so that they are able to make rational, informed, decisions.

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

<organization> Property of an organization that appropriate activities and decisions are communicated to relevant stakeholders in a comprehensive, accessible and understandable manner.

Note 1 to entry: Inappropriate communication of activities and decisions can violate security, privacy, or confidentiality requirements.

<system> Property of a system that appropriate information about the system is communicated to relevant stakeholders.

Note 1 to entry: Appropriate information for system transparency can include aspects such as features, components, procedures, measures, design goals, design choices and assumptions.

Note 2 to entry: Inappropriate disclosure of some aspects of a system can violate security, privacy, or confidentiality requirements.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Triple Bottom Line

People, Planet, Profit.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

3BL (triple bottom line) advocates believe that social (and environmental) performance can be measured in fairly objective ways, and that firms should use these results in order to improve their social (and environmental) performance. Moreover, they should report these results as a matter of principle, and in using and reporting on these additional "bottom lines' firms can expect to do better by their financial bottom line in the long run (Norman and MacDonald 246).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Trust

Trust is viewed as: (1) a set of specific beliefs dealing with benevolence, competence, integrity, and predictability (trusting beliefs); (2) the willingness of one party to depend on another in a risky situation (trusting intention); or (3) the combination of these elements. While “Trust” is usually not a property ascribed to machines, this document aims to stress the importance

of being able to trust not only in the fact that AI systems are legally compliant, ethically adherent and robust, but also that such trust can be ascribed to all people and processes involved in the AI system's life cycle.

Source: *HLEG AI, Ethics Guidelines for Trustworthy AI. URL:* https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

Firm belief in the reliability, truth, or ability of someone or something; To believe or accept a statement, story, etc., without seeking verification or evidence for it.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Trust models are developed for multiagent communication: A reputation-based trust model collects, distributes, and aggregates feedback about participants' past behaviour. These models help agents decide whom to trust, encourage trustworthy behaviour, and discourage participation by agents who are dishonest. Reputation-based trust models are basically divided into two categories based on the way information is aggregated from an evaluator's perspective. They are "Direct/ Local experience model" and "Indirect/Global reputation model" where direct experience is derived from direct encounters or observations (first-hand experience) and indirect reputation is derived from inferences based on information gathered indirectly (second-hand evidence such as by word of mouth) (Das and Islam 2012).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

---

Legal definitions of trust include:

1. An equitable or beneficial right or title to land or other property, held for the beneficiary but another person, in whom resides the legal tile or ownership, recognised and enforced by courts of chancery.
2. An obligation arising out of a confidence reposed in the trustee or representative, who has the legal title to property conveyed to him, that he will faithfully apply the property according to the confidence reposed or, in other words, according to the wishes of the grantor of trust.
3. An equitable obligation, either express or implied, resting upon a person by reason of a confidence reposed in him, to apply or deal with the property for the benefit of some other person, or for the benefit of himself and another or others, according to such confidence (Black's Law Dictionary Online).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf

Trustworthy AI

Trustworthy AI has three components: (1) it should be lawful, ensuring compliance with all applicable laws and regulations (2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values and (3) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the AI system's life cycle.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342

3.21 U

Universal Design

Terms such as "Design for All", "Universal Design", "accessible design", "barrier free design", "inclusive design" and "transgenerational design" are often used interchangeably with the same meaning. These concepts have been developed by different stakeholders working to deliver high levels of accessibility. A parallel development of human centred design emerged within ergonomics focusing on usability. These related concepts are expressed in the human rights perspective of

the Design for All approach. The Design for All approach focuses on user involvement and experiences during the design and development process to achieve accessibility and usability. It should be applied from the earliest possible time, and throughout all stages in the life of products and services which are intended for mainstream use. A Design for All approach also focuses on user requirements and interoperability between products and services across the end-to-end chain of use to reach inclusive and non-stigmatizing solutions.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

Unsupervised (Machine) Learning

Machine learning that makes use of unlabelled data during training.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Use case

A use case is a specific situation in which a product or service could potentially be used. For example, self-driving cars or care robots are use cases for AI.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

User

A user is a person that uses, supports or maintains the product, such as system administrators, database administrators, information technology experts, software professionals and computer technicians.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

3.22 V

Values

Worth or quality as measured by a standard of equivalence; The relative worth, usefulness, or importance of a thing or (occasionally) a person; the estimation in which a thing is held according to its real or supposed desirability or utility (OED).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Value consists in the relation of harmony or fitness. It finds its point of contact with common sense in the popular expression "good for"... or "good of its kind" and the relationship is that of the particular to its universal. "Value" consists in the fulfilment of interest as such (Perry 1914).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Validation

A check for accuracy of relationships between claims and data supporting or refuting those claims.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Validation is the process of building an acceptable level of confidence that an inference about a simulated process is a correct or valid inference for the actual process (Van Horn quoted in Jagdev et al 1995, 333).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Validation means establishing by objective evidence that the particular requirements for a specific intended use can be consistently fulfilled. Process validation means establishing by objective evidence that a process consistently produces a result or product meeting its predetermined specifications. Design validation means establishing by objective evidence that device specifications conform with user needs and intended uses (CFR 21 Part 820.3 Definitions (z) (1,2)).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Validation data

Data used for providing an evaluation of the trained AI system and for tuning its non-learnable parameters and its learning process, among other things, in order to prevent overfitting; whereas the validation dataset can be a separate dataset or part of the training dataset, either fixed or variable split.

Source: *EU Artificial Intelligence Act URL :* *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206*

Data samples used to assess the performance of one or more candidate machine learning models.

Note 1 to entry: Validation data is disjoint from training data and test data.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL:* *https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

A validation data set is a sample of data held back from training an AI system that is used to give an estimate of model skill while tuning the model's hyperparameters (e.g., the number of hidden units in a neural network).

Source: *Own elaboration.*

Verification

A check for accuracy of a proposed solution.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Verification is "the process of confirming that the conceptual model has been correctly translated into an operational computer programme and that the calculations made with this programme utilize the correct input data" (Schlesinger et al 1974).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Verification means confirmation by examination and provision of objective evidence that specified requirements have been fulfilled (CFR 21 Part 820.3 Definitions (aa)).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Confirmation, through the provision of objective evidence, that specified requirements have been fulfilled.

Note 1 to entry: Verification only provides assurance that a product conforms to its specification.

Source: *ISO/IEC DIS 22989(en). Terms related to Trustworthiness. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

Virtual Reality

A "virtual reality" is defined as a real or simulated environment in which a perceiver experiences telepresence (Steuer 1992, 6).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Virtual Reality is an alternate world filled with computer-generated images that respond to human movements. These simulated environments are usually visited with the aid of an expensive data suit which features stereophonic video goggles and fiber-optic data gloves (Greenbaum, 1992; quoted in Steuer 1992, 5).

Source: *(Computational Disciplines) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Validation data

Data used for providing an evaluation of the trained AI system and for tuning its non-learnable parameters and its learning process, among other things, in order to prevent overfitting; whereas the validation dataset can be a separate dataset or part of the training dataset, either fixed or variable split.

Source: *EU Artificial Intelligence Act URL : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206*

Data samples used to assess the performance of one or more candidate machine learning models.

Note 1 to entry: Validation data is disjoint from training data and test data.

Source: *ISO/IEC DIS 22989(en). Terms related to Machine Learning. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso-iec:22989:dis:ed-1:v1:en*

A validation data set is a sample of data held back from training an AI system that is used to give an estimate of model skill while tuning model's hyperparameters (e.g., the number of hidden units in a neural network).

Source: *Own elaboration.*

Vulnerable (persons or groups)

No commonly accepted or widely agreed legal definition of vulnerable persons exists, due to their heterogeneity. What constitutes a vulnerable person or group is often context-specific. Temporary life events (such as childhood or illness), market factors (such as information asymmetry or market power), economic factors (such as poverty), factors linked to one's identity (such as gender, religion or culture) or other factors can play a role. The Charter of Fundamental Rights of the EU encompasses under Article 21 on non-discrimination the following grounds, which can be a reference point amongst others: namely sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age and sexual orientation. Other articles of law address the rights of specific groups, in addition to those listed above. Any such list is not exhaustive and may change over time. A vulnerable group is a group of persons who share one or several characteristics of vulnerability.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL:* *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

3.23 W

Weapon System

A weapon system consists of a weapon and the items associated with its employment (Schmitt 2013, 3).

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

An autonomous weapon system is a weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation (Department of Defense 2012, Directive 3000.09, quoted in Schmitt 2013, 5).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Wellbeing

With reference to a person or community: the state of being healthy, happy, or prosperous;

Physical, psychological, or moral welfare.

Source: *(Ordinary Language) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

The OECD recommends two areas of individual wellbeing dimensions that can be broken into eleven dimensions:

"Material Living Conditions include income and wealth, jobs and earnings, and housing. Quality of Life: health status, work and life balance, education and skills, social connections, civic engagement and governance, environmental quality, personal security, and subjective wellbeing". The OECD suggests that these wellbeing domains are sustained over time by natural capital, economic capital, human capital, and social capital (OECD 2011, 6).

Source: *(Government, Policy and Social Science) IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL:* *https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/eadv2_glossary.pdf*

Workflow of the model

The workflow of an AI model shows the phases needed to build the model and their interdependencies. Typical phases are: Data collection and preparation, Model development, Model training, Model accuracy evaluation, Hyperparameters' tuning, Model usage, Model maintenance, Model versioning. These stages are usually iterative: one may need to re-evaluate and go back to a previous step at any point in the process.

Source: *HLEG AI, Assessment List for Trustworthy AI (ALTAI) URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342*

3.24 X

3.25 Y

3.26 Z

---

4 Conclusions

This document presents a glossary of terms related to artificial intelligence, from a human-centric perspective and with a focus on a wide audience covering both researchers and policy makers. The glossary was built by collecting and contrasting existing related glossaries. A total of 230 terms were finally selected and incorporated in the glossary. We hope that this glossary will serve as a compact and relevant resource for discussions around AI, involving people from different backgrounds and interests, from scientific, technical and policy perspectives.. As future work, we see the need to engage in deep discussions for some of the terms, given the different sensitivities and concepts behind. Also, to adopt a dynamic approach for the glossary, so that it can be dynamically updated with new definitions and terms.

References

A

AI HLEG, “Ethics Guidelines for Trustworthy AI”, High Level Expert Group on Artificial Intelligence, B-1049 Brussels, 2019.

AI HLEG, “The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment”, High Level Expert Group on Artificial Intelligence, B-1049 Brussels, 2020.

Allen, Colin, Gary Varner, and Jason Zinser. "Prolegomena to any future artificial moral agent." Journal of Experimental & Theoretical Artificial Intelligence 12, no. 3 (2000): 251-261.

Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. "Concrete problems in AI safety." arXiv preprint arXiv:1606.06565 (2016).

Arnold, Thomas, and Matthias Scheutz. "Against the moral Turing test: accountable design and the moral reasoning of autonomous systems." Ethics and Information Technology 18, no. 2 (2016): 103-115.

Azuma, Ronald T. "A survey of augmented reality." Presence: Teleoperators and virtual environments 6, no. 4 (1997): 355-385.

B

Bauer, Johannes M., and Paulien M. Herder. "Designing socio-technical systems." Philosophy of technology and engineering sciences. North-Holland, 2009. 601-630.

Boehm, Barry W. 1991. "Software Risk Management: Principles and Practices." IEEE Software 8 (1): 32-41. doi: http://dx.doi.org/10.1109/52.62930

Bostrom, Nick. “How long before superintelligence?” Linguistic and Philosophical Investigations, 2006, Vol. 5, No. 1, pp. 11-30.

Bostwick, Gary L. "A taxonomy of privacy: Repose, sanctuary, and intimate decision." Cal. L. Rev. 64 (1976): 1447.

Bradford, Alina. “What is a law in science?” https://www.livescience.com/21457-what-is-a-law-in-science-definition-of-scientific-law.html

Brown, Garrett W., McLean, Iain and McMillan, Alistair. “A Concise Oxford Dictionary of Politics and International Relations (4 ed.)”, Oxford University Press, 2018.

C

Clifton, Chris. "Individually Identifiable Data." In Encyclopedia of Database Systems, pp. 1471-1472. Springer US, 2009.

Costanza, R., & Patten, B. C. (1995). Commentary: Defining and predicting sustainability. Ecological Economics, 15(3), 193–196.

Cua, A. S. “Toward an Ethics of Moral Agents.” Philosophy and Phenomenological Research, vol. 28, no. 2, 1967, pp. 163–174.

D

Das, Anupam, and Mohammad Mahfuzul Islam. "SecuredTrust: a dynamic trust computation model for secured communication in multi-agent systems." IEEE Transactions on Dependable and Secure Computing 9, no. 2 (2012): 261-274.

Data.gov. “Glossary of Terms”. No date. https://www.data.gov/glossary

Deutsch, David. "Quantum theory, the Church-Turing principle and the universal quantum computer." In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 400, no. 1818, pp. 97-117. The Royal Society, 1985.

Dirac, Paul Adrien Maurice. The principles of quantum mechanics. No. 27. Oxford university press, 1981.

Dyschkant, Alexis. "Legal personhood: How we are getting it wrong." U. Ill. L. Rev. (2015): 2075.

Duval, Beverly K and Linda Main. “Expert Systems: What is an Expert System?”. Library Software Review. (1994).

E

EEOC. “Race/ Color Discrimination” no date. https://www.eeoc.gov/laws/types/race_color.cfm

E03659, European Commission, “Ethics of connected and automated vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility”, Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659), 2020.

F

Felici, Massimo, Theofrastos Koulouris, and Siani Pearson. "Accountability for data governance in cloud ecosystems." In Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on, vol. 2, pp. 327-332. IEEE, 2013.

Finnemore, Martha, and Kathryn Sikkink. "International norm dynamics and political change." International organization 52.4 (1998): 887-917.

Fogg, B. J., Gregory Cuellar, and David Danielson. "Motivating, influencing, and persuading users: An introduction to captology." Human Computer Interaction Fundamentals (2009): 109-122.

Franklin, Stan, and Art Graesser. "Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents." In International Workshop on Agent Theories, Architectures, and Languages, pp. 21-35. Springer, Berlin, Heidelberg, 1996.

G

Geertz, C. (1973). Interpretation of Cultures. New York: Basic Books.

Gelperin, David, and Bill Hetzel. "The growth of software testing." Communications of the ACM 31, no. 6 (1988): 687-695.

General Data Protection Regulation. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46." Official Journal of the European Union (OJ) 59 (2016): 1-88.

Gibbs, Jack P. "Norms: The problem of definition and classification." American Journal of Sociology 70, no. 5 (1965): 586-594.

Gomez, E., Charisi, V., Tolan, S., Miron, M., Martinez Plumed, F. and Escobar Planas, M., HUMAINT: Understanding the impact of Artificial Intelligence on human behaviour, Amran, G. editor(s), Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-28212-9 (online), 978-92-76-28211-2 (print), doi:10.2760/23970 (online), 10.2760/043359 (print), JRC122667.

Graham, Mark, Matthew Zook, and Andrew Boulton. "Augmented reality in urban places: contested content and the duplicity of code." Transactions of the Institute of British Geographers 38, no. 3 (2013): 464-479.

Grant, Wyn. “Responsible Government”. Oxford Concise Dictionary of Politics. 2016. Online.

Greenbaum, P. (1992, March). The lawnmower man. Film and video, 9 (3), pp. 58-62.

Guarino, N. and Giaretta, P. 1995. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. Mars (ed.) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995. IOS Press, Amsterdam: 25-32.

H

H. Nakashima. AI as complex information processing. Minds and machines, 9:57–80, 1999.

Hayes-Roth, Barbara. An Architecture for Adaptive Intelligent Systems. Artificial Intelligence: Special Issue on Agents and Interactivity, (72): 329-365, 1995.

Howell, Kerry E. An introduction to the philosophy of methodology. Sage, 2012.

Hu, Yongxiang, David Winker, Mark Vaughan, Bing Lin, Ali Omar, Charles Trepte, David Flittner et al. "CALIPSO/CALIOP cloud phase discrimination algorithm." Journal of Atmospheric and Oceanic Technology 26, no. 11 (2009): 2293-2309.

I

IEEE (2017), "A Glossary for Discussion of Ethics of Autonomous and Intelligent Systems, Version 1 - Prepared for The IEEE Global Initiative for Ethically Aligned Design", URL: https://standards.ieee.org/wp-content/uploads/import/documents/other/eadv2_glossary.pdf

iDASH. "PHI and PII Definition and Data Elements". No date. https://idash.ucsd.edu/phi-and-pii-definition-and-data-elements.

Ince, Darrel. A Dictionary of the Internet, 3rd edition. 2013. Oxford Online.

ISO (2018) "ISO/IEC 22989 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology", URL: https://www.iso.org/standard/74296.html

J

Jacobs, Adam. "The pathologies of big data." Communications of the ACM 52, no. 8 (2009): 36-44.

Jagdev, H. S., Jim Browne, and Paddy Jordan. "Verification and validation issues in manufacturing models." Computers in industry 25.3 (1995): 331-353.

Jones, Andrew JI, and Marek Sergot. "On the characterisation of law and computer systems: The normative systems perspective." Deontic logic in computer science: normative system specification (1993): 275-307.

L

Legg, Shane, and Marcus Hutter. "A collection of definitions of intelligence." Frontiers in Artificial Intelligence and applications157 (2007): 17.

Lyon, David, and Elia Zureik, eds. Computers, surveillance, and privacy. U of Minnesota Press, 1996.

M

Malinowski, B. (1931). Culture. In E.R.A. Seligman (ed.), Encyclopedia of the Social Sciences, Vol. 4 (pp. 621–646). New York: Macmillan.

Milgram, Paul, and Fumio Kishino. "A taxonomy of mixed reality visual displays." IEICE TRANSACTIONS on Information and Systems 77, no. 12 (1994): 1321-1329.

MLADictionary, Blacks Law. "Blacks law dictionary." URL: https://dictionary.thelaw.com/truth [January 13th, 2021] (1990).

Moor, James H. "What is computer ethics?." Metaphilosophy 16, no. 4 (1985): 266-275.

N

Nagel, Thomas. "Equality". The Oxford Companion to Philosophy. 2nd edition. 2005. Online.

NIST Interagency/Internal Report (NISTIR) - 7298rev2. 2013

Norman, Wayne, and Chris MacDonald. "Getting to the bottom of "triple bottom line"." Business ethics quarterly 14.2 (2004): 243-262.

O

OECD (2011) Better Life Initiative. "Compendium of OECD Well-being Indicators."

OECD (2017) "Glossary of Statistical Terms". URL: https://stats.oecd.org/glossary/

OECD (2018) "Glossary of terms" in Promoting the Digital Transformation of African Portuguese-Speaking Countries and Timor-Leste. OECD Publishing, Paris, https://doi.org/10.1787/9789264307131-9-en.

OECD (2022) "OECD Framework for the Classification of AI systems", OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en.

P

Perry, Ralph Barton. "The Definition of Value." The Journal of Philosophy, Psychology and Scientific Methods, vol. 11, no. 6, 1914, pp. 141– 162.

Porta, Miquel. "A Dictionary of Epidemiology (6 ed.)", Oxford University Press. 2016.

Posner, Richard A. "The right of privacy." Ga. L. Rev. 12 (1977): 393.

Price, Cathy J., and Karl J. Friston. "Functional ontologies for cognition: The systematic definition of structure and function." Cognitive Neuropsychology 22, no. 3-4 (2005): 262-275.

Project Management Institute, A Guide to the Project Management Body of Knowledge, (PMBOK Guide), Fourth Edition, ANSI/PMI 99-001- 2008, pp. 273-312.

Q

Quinn, Kenneth. "Expert System Shells: What to Look For," Reference Services Review 18 (1), (Spring 1990): 83.

R

R. R. Gudwin. Evaluating intelligence: A computational semiotics perspective. In IEEE International conference on systems, man and cybernetics, pages 2080– 2085, Nashville, Tenessee, USA, 2000.

Reeve, Andrew. "Consent". The Concise Oxford Dictionary of Politics, 3rd edition. Online version.

Reeve, Andrew. "Rights". The Concise Oxford Dictionary of Politics, 3rd edition. Online version.

Rescorla, Michael. "Computational Theory of Mind". https://plato.stanford.edu/entries/computational-mind/

Rhodes, David G. "A practical approach to problem-based learning: simple technology makes PBL accessible." American Journal of Pharmaceutical Education 63.4 (1999): 410-414.

Ropohl, Günter. "Philosophy of socio-technical systems." Society for Philosophy and Technology Quarterly Electronic Journal 4.3 (1999): 186-194.

S

Samoili, S., Lopez Cobo, M., Delipetrev, B., Martinez-Plumed, F., Gomez Gutierrez, E. and De Prato, G., AI Watch. Defining Artificial Intelligence 2.0, EUR 30873 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-42648-6, doi:10.2760/019901, JRC126426.

Schmitt, Michael N., Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics (December 4, 2012). Harvard National Security Journal Feature (2013). Available at SSRN: https://ssrn.com/abstract=2184826 or http://dx.doi.org/10.2139/ssrn.2184826

Smith, Brian C. The foundations of computing, 2002. http://www.ageofsignificance.org/people/bcsmith/print/smith-foundtns.pdf (accessed October 18, 2017)

Steuer, Jonathan. "Defining virtual reality: Dimensions determining telepresence." Journal of communication 42.4 (1992): 73-93.

T

Tylor, Edward Burnett. Primitive culture: researches into the development of mythology, philosophy, religion, art, and custom. Vol. 2. J. Murray, 1871.

W

Wang, Yingxu, George Baciu, Yiyu Yao, Witold Kinsner, Keith Chan, Bo Zhang, Stuart Hameroff et al. "Perspectives on cognitive informatics and cognitive computing." International Journal of Cognitive Informatics and Natural Intelligence (IJCINI) 4, no. 1 (2010): 1-29.

Worldbank. "Worldwide Governance Indicators (WGI) Project". 2017. http://info.worldbank.org/governance/wgi/#home

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
joint-research-centre.ec.europa.eu

 @EU_ScienceHub

 EU Science Hub - Joint Research Centre

 EU Science, Research and Innovation

 EU Science Hub

 EU Science

Publications Office
of the European Union