

The Language of Trustworthy AI: An In-Depth Glossary of Terms (Mar. 2023)

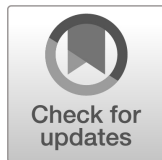
Daniel Atherton, et al.

National Institute of Standards and Technology

Copyright 2023. All rights reserved.

Reprinted with permission.





NIST Artificial Intelligence AI 100-3

The Language of Trustworthy AI: An In-Depth Glossary of Terms

Daniel Atherton
Reva Schwartz
Peter C. Fontana
Patrick Hall

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.AI.100-3>

NIST Artificial Intelligence AI 100-3

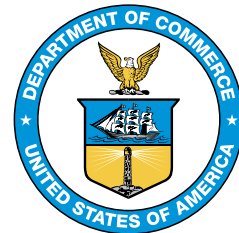
The Language of Trustworthy AI: An In-Depth Glossary of Terms

Daniel Atherton
Patrick Hall
BNH.AI

Reva Schwartz
Peter C. Fontana
*Information Technology Laboratory
National Institute of Standards and Technology*

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.AI.100-3>

March 2023



U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

NIST Technical Series Policies

[Copyright, Use, and Licensing Statements](#)

[NIST Technical Series Publication Identifier Syntax](#)

Publication History

Approved by the NIST Editorial Review Board on 2023-03-29

How to cite this NIST Technical Series Publication:

Daniel Atherton, Reva Schwartz, Peter C. Fontana, Patrick Hall (2023) The Language of Trustworthy AI: An In-Depth Glossary of Terms. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Artificial Intelligence AI 100-3. <https://doi.org/10.6028/NIST.AI.100-3>.

NIST Author ORCID iDs

Reva Schwartz: 0000-0002-9012-6306

Peter C. Fontana: 0000-0002-1101-2010

Contact Information

aiframework@nist.gov

Submit Comments

To provide input on the glossary please send comments to aiframework@nist.gov

Abstract

This document is a guide and record of the development for the NIST (National Institute of Standards and Technology) [glossary](#) of terms for trustworthy and responsible artificial intelligence (AI) and machine learning (ML). The glossary effort seeks to promote a shared understanding and improved communication among individuals and organizations seeking to operationalize trustworthy and responsible AI through approaches such as the NIST AI Risk Management Framework (AI RMF).

Keywords

definitions; responsible AI; trustworthy AI.

Table of Contents

1. Glossary Aims	1
2. Glossary Development and Methodology	2
2.1. Background	2
2.2. Term generation and selection	3
2.3. Definition identification and selection	3
3. Glossary Visualization	4

Acknowledgments

The authors would like to thank Dr. Gregory Falco and his students at Johns Hopkins University Whiting School of Engineering who assisted with identification of terms and definitions in the early stages of the glossary.

1. Glossary Aims

This document is a guide and record of the development for the NIST (National Institute of Standards and Technology) [glossary](#)¹ of terms for trustworthy and responsible artificial intelligence (AI) and machine learning (ML). The glossary effort seeks to promote a shared understanding and improved communication among individuals and organizations seeking to operationalize trustworthy and responsible AI through approaches such as the NIST AI Risk Management Framework (AI RMF).

Like the AI RMF, the glossary is non-sector specific and use-case agnostic, designed to be flexible for all organizations and sectors of society to use. The goal of this common vocabulary is not to declare one specific meaning for identified terms, but to provide interested parties with a broader awareness of the multiple meanings of commonly used terms within the interdisciplinary field of trustworthy and responsible AI.

NIST's collaborative engagement with private industry, civil society groups, academia, and public sector organizations during the development of the AI RMF was an influence in the development of the glossary. As the federal AI standards coordinator, NIST has worked to identify critical terms development activities, strategies, and gaps, and provide definitions for an expanding lexicon.

The glossary can:

- promote a shared understanding and improved communication among individuals and organizations seeking to operationalize trustworthy and responsible AI
- foster and encourage cross-collaboration among different disciplines.
- serve as a first-stop resource for those new to the field or looking to strengthen the field of artificial intelligence.
- be used in conjunction with the NIST AI RMF and related resources, or as a stand-alone document.

A comprehensive glossary for responsible artificial intelligence is a critical step towards promoting effective communication and a shared understanding of concepts. A deeper appreciation and broader perspective towards relevant terms and their multiple definitions is particularly useful for fostering cross-collaboration within the interdisciplinary trustworthy and responsible AI community. Broader awareness of lexical meaning can help expand understanding beyond a given silo within the field, and enhance collaboration. The process of building this glossary has involved — and will continue to include — collaboration with experts from various sectors and disciplines to ensure it remains a relevant and valuable resource. Core principles in designing the glossary include:

- **Alignment with existing international and industry standards, while integrating**

¹Glossary URL is https://airc.nist.gov/AI-RMF_Knowledge_Base/Glossary.

authoritative and accessible sources. While the glossary seeks to align with the definitions of terms used in international standards and industry-specific glossaries, (such as those produced by IEEE, ANSI, and ISO/IEC) it also assembles an array of sources that offer clear definitions for terms that are regularly in use but often not explicitly defined.

- **Inclusion of terms related to emerging AI technologies.** As technologies evolve, new terms arise. The glossary will continue to include new terms and updated definitions as needed.
- **Inclusion of definitions from a wide variety of domains and areas.** Information sources were consulted from the following domains and disciplines: machine learning, computational sciences, statistics, psychology, behavioral sciences, social sciences, and the law. This variety of information source enhances understanding and cross-collaboration across disciplines, which is particularly important in the multi-disciplinary field of trustworthy and responsible AI.
- **Collaborative, based on consultation with subject matter experts.** NIST has collaborated with subject matter experts from various organizations and sectors to ensure that the glossary is comprehensive and reflects the current understanding of AI terms. Rigorous discussion of terms and their nuances has been a core part of glossary development.
- **Regular review and feedback processes.** The glossary will be subject to review and feedback from the broader AI community to ensure accuracy and completeness. Comments and suggestions for new terms, revised definitions or additional definitions can be submitted to NIST at aiframework@nist.gov.
- **Dissemination and use.** NIST will promote the use and dissemination of the glossary to a broad range of stakeholders, including researchers, developers, and policymakers, to facilitate a shared understanding of trustworthy and responsible AI terminology.

2. Glossary Development and Methodology

2.1. Background

Work on the glossary began in early 2022 with the creation of an underlying database architecture to maximize simplicity and replicability. For each definition, a corresponding citation tag would correspond to a full citation in the next tab. Each row for each citation contains all the requisite cells to construct a citation; a simple scripting algorithm can generate citations for each tag according to any style guide. After the first approximately 150 terms were defined, which was in early May 2022, the project expanded in scope by bringing in preliminary feedback from additional stakeholders.

By fall 2022, the glossary had swelled to approximately 500 terms requiring definition. Based on preliminary feedback, each term provided the space for up to five different definitions, as well as a column each for legal-specific definitions, synonymous terms, related terms, and a space for clarifying notes.

2.2. Term generation and selection

Glossary development involved careful consideration of relevant and widely used terms professionals in the field of trustworthy and responsible AI. Certain terms also have connotations that may hold problematic valences and require appropriate synonyms. The term selection process required a sensitive understanding of the technical jargon and concepts that are commonly used by experts in the industry, while allowing for terms with cross-domain meanings.

2.3. Definition identification and selection

Once terms were selected, the goal was to provide a broad coverage of definitions from a variety of sources and disciplines. Definition selection focused on the following sources: NIST publications; standards published by IEEE, ISO, and ANSI; and peer-reviewed journals, conference proceedings, and textbooks across many domains and areas, including computer science, statistics, psychology, sociology, and the humanities. These sources were selected to ensure that the sources were well-vetted, and in many cases well-cited, and provided reliable definitions. Given the dynamic nature of many terms, we also drew from other sources such as arXiv preprints, journalism, and technology-based websites' glossaries. In some instances, we included definitions from dictionaries, both general and technical, to provide additional context and clarity. In the case of terms that have been in use for an extended period, we faced the challenge of finding earlier sources in computer science and statistics that described these terms. In a large number of cases, a variety of organizational glossary publications were consulted, as well as technical and general dictionaries. For a small number of terms, definitions were selected from easily accessible websites. This process also enabled a simultaneous search for more sources.

The primary process for identifying and selecting terms was to:

- use existing glossaries from well-known sources including ISO Standards and government reports.
- look for definitions from scholarly papers, online technical books, and software system documentation.
- if no definitions were found (or to supplement what was found), use other online definitions from the web as well as physical textbooks.

All definitions were checked for technical accuracy. The definition search team reflects different disciplinary backgrounds (humanities, linguistics, data science, computer science)

and sought content from a variety of technical sources and domains: behavioral, social, computational, statistical, etc. In some cases additional definitions were added when terms exhibited polysemy (were used in multiple ways and with multiple meanings). When written definitions did not exist for a specific term, sometimes the definition for similar terms was used instead.

The glossary includes 511 terms with at least one definition, and 218 terms with at least two. The bibliography contains 379 citations.

3. Glossary Visualization

Forthcoming.