



NIST Trustworthy and Responsible AI
NIST AI 300-1 ipd

Guidance and Templates for
Public-Facing AI Documentation

An AI Standards “Zero Draft”

Initial Public Draft

Razvan Amironesei
Jesse Dunietz

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.AI.300-1.ipd>

NIST Trustworthy and Responsible AI
NIST AI 300-1 ipd

**Guidance and Templates for
Public-Facing AI Documentation**

An AI Standards “Zero Draft”

Initial Public Draft

Razvan Amironesei
Jesse Dunietz

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.AI.300-1.ipd>

July 2026



U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Arvind Raman, NIST Director and Under Secretary of Commerce for Standards and Technology

1

2 **Guidance and Templates for Public-Facing AI**
3 **Documentation**

4

5 **An AI Standards “Zero Draft”**

6 Contents

7	1 Scope	1
8	2 Terms and definitions	1
9	3 Dimensions of potential outcomes of public-facing documentation processes	2
10	3.1 General	2
11	3.2 Process-driven outcome dimensions	4
12	3.2.1 Organizational resource demands	4
13	3.2.2 Internal collaboration	4
14	3.2.3 Internal decision-making	4
15	3.2.4 Release speed	4
16	3.3 Artifact-driven outcome dimensions	4
17	3.3.1 Reuse and integration	5
18	3.3.2 Malicious use or attacks	5
19	3.4 Jointly driven outcome dimensions	5
20	3.4.1 Accountability	6
21	3.4.2 Privacy preservation	6
22	3.4.3 Proprietary information protection	6
23	3.4.4 Legal obligations	6
24	3.4.5 Trust	6
25	4 Guidance on public-facing documentation processes and artifacts	7
26	4.1 General	7
27	4.2 Guidance on documentation artifacts	7
28	4.2.1 Artifact comprehensibility	7
29	4.2.2 Artifact informativeness	8
30	4.2.3 Artifact correctness	8
31	4.2.4 Artifact judiciousness	8
32	4.2.5 Artifact freshness	8
33	4.2.6 Artifact interoperability	8
34	4.2.7 Artifact findability and availability	9
35	4.2.8 Artifact maintainability	9
36	4.3 Guidance on documentation processes	9
37	4.3.1 General	9
38	4.3.2 Define documentation objectives	10
39	4.3.3 Provide organizational support	10
40	4.3.4 Keep processes manageable	10
41	4.3.5 Provide guidelines	10
42	4.3.6 Document continuously	10
43	4.3.7 Distribute documentation work	11
44	4.3.8 Incorporate audience input	11

45	4.4 Trade-offs between qualities of documentation artifacts and processes	11
46	4.4.1 Informativeness vs. judiciousness	12
47	4.4.2 Informativeness vs. comprehensibility, maintainability, and process	
48	manageability	12
49	4.4.3 Informativeness vs. freshness	12
50	4.4.4 Continuous documenting vs. keeping processes manageable	12
51	5 Documentation templates	12
52	5.1 General	12
53	5.2 Dataset documentation template	14
54	5.3 Model documentation template	15
55	6 Profiles	19
56	A Default profiles	20
57	A.1 General	20
58	A.2 Default AI dataset profile	21
59	A.3 Default AI Model Profile	25

Foreword: Background on the Zero Drafts Project

In March 2025, NIST announced its AI Standards Zero Drafts project. In this pilot project, NIST is collecting input on topics with a science-backed body of work and using it to develop “zero drafts”—preliminary, stakeholder-driven drafts of standards that are as thorough as possible. These drafts will then be submitted via the private sector-led standardization process as proposals for voluntary consensus standards. The project aims to broaden participation in and accelerate the creation of standards that meet the AI community’s needs and unleash AI innovation.

Based on community input, NIST selected two topics for the Zero Drafts pilot, one of which is documentation of AI datasets and models for transparency between AI stakeholders. This document is the initial public draft for the forthcoming zero draft on AI dataset and AI model documentation. NIST will publish a final and complete version of this document based on feedback received.

The resulting document will be submitted to INCITS/AI, the private sector-led committee that represents the United States in ISO/IEC JTC 1/SC 42. SC 42 is the subcommittee that focuses on cross-sectoral AI standards development within ISO and IEC (two prominent international standards developing organizations that collaborate closely through joint committees). Assuming that INCITS/AI proposes the document as a new project for SC 42 and that SC 42 accepts the proposal, the future of the document will be subject to the usual consensus processes of standards development. NIST does not expect to maintain the document further, and while the agency expects to contribute heavily through the ISO standardization process, NIST will be just one voice among many in SC 42 and INCITS/AI.

This document is intended for voluntary use, as would be any voluntary consensus standard that derives from it. In such a document, any usage of “shall” or stated “requirement” does not reflect any regulatory intent, nor does it indicate a rule or directed action from the U.S. federal government. Rather, use of “shall” or “requirement” indicates only what constitutes conformity: if a voluntary adopter of the document claims a process or artifact is conformant, that claim is valid only if the process or artifact upholds the requirement.

NIST has received strong indications from many stakeholders, most notably from industry, that standardizing AI documentation practices would lower the time, effort, and expertise required for both creating and consuming documentation, which would facilitate building on prior research and development efforts and ultimately accelerate innovation and adoption. The few requirements in this document aim to help to create the desired usefulness and consistency and integrate AI documentation processes into organizational workflows.

To help streamline the movement of this zero draft along its intended trajectory into the ISO/IEC ecosystem, the language in the draft is crafted to match ISO/IEC’s conventions. In particular:

- **“Shall”** indicates a requirement—a provision that is mandatory in order for something to be considered conformant to the document. **“Required”** and **“requirement[s]”** are also used in this sense; they do not indicate any regulatory or rulemaking implications of the requirements.
- **“Should”** indicates guidance, i.e., a recommendation—a provision that is advised but not mandatory. **“Guidance”** is also used in this sense; it does not indicate any regulatory or rulemaking implications of the recommendations.
- **“May”** indicates a permission or latitude—a provision that is allowed under the standard. No obligation or recommendation is implied.
- **“Can”** indicates a possibility or a capability—a statement of fact about what is technically feasible. It is not normative.

NOTE TO REVIEWERS

NIST invites input on any aspect of this document. Input can be shared by email to ai-standards+doczd@nist.gov. NIST strongly encourages comments using the optional [commenting template](#), although input can also be provided via marked-up documents, bulleted lists of comments and concerns, reaction letters, or any other form that stakeholders find most convenient. NIST welcomes specific suggestions for replacement text, where applicable.

Submissions, including attachments and other supporting materials, will become part of the public record and may be subject to public disclosure. Significant contributions may be acknowledged in NIST's final product. Organizations are also welcome to host listening sessions in which they convene stakeholders to share verbal feedback with the agency.

NIST will consider input received by **September 16, 2026**. Input received later may still be considered for NIST inputs to subsequent standardization processes, or for other NIST contributions on AI documentation.

If AI assistants are used to help with feedback, please disclose this use, and ensure that real-world insights are accurately and concisely captured. This helps to ensure that the output prioritizes actionable advice and references that go beyond what could easily be elicited from AI assistants.

107 **Introduction**

108 (To be completed with the final draft)

109 **1 Scope**

110 This document provides guidance on objectives for and approaches to documenting AI datasets and models with
111 the intent of releasing the resulting documentation artifacts publicly. The document provides recommendations
112 for documentation processes, as well as requirements for documentation artifacts in the form of templates against
113 which conformity of documentation artifacts can be assessed. Furthermore, it specifies a mechanism for extending
114 the templates and related requirements via profiles.

115 This document is applicable to any organization, regardless of size, type, and nature, that provides or uses
116 products or services that utilize AI systems. It is applicable to any type of AI model, including generative AI and
117 large language models but also predictive machine learning models and others. It is applicable to public-facing
118 documentation of datasets and models whether or not the datasets and models themselves are publicly available.

119 *Note: This document addresses only documentation of AI datasets and models. It does not address documenting entire*
120 *AI systems, which may consist of large, complex arrangements of components that rely on or interact with a given*
121 *dataset or model. NIST assesses that documentation practices at this level are less mature, so they have been left to*
122 *future work.*

123 **2 Terms and definitions**

124 **AI dataset**

125 *dataset* intended for use in training or evaluating an *AI model* or *AI system*

126 **Documentation process**

127 set of organizational activities, procedures, and decisions undertaken with the objective of producing and
128 maintaining a *documentation artifact*

129 **Documentation artifact**

130 *Alternate term: documentation <entity>*

131 record describing a product or service, or some portion thereof

132 Note 1 to entry: Documentation artifacts can take many forms, including physical or digital publications such as
133 PDFs, JSON files, or web-based media.

134 Note 2 to entry: The product or service can be or include a digital object such as a dataset or model.

135 **Documentation <activity>**

136 activity that takes place as part of a *documentation process*

137 **Documentarian**

138 member of an organization's staff who coordinates documentation processes and release of documentation
139 artifacts

140 Note 1 to entry: A documentarian can hold other roles simultaneously.

141 **Field**

142 entry in a *documentation artifact* characterizing a defined aspect of the documented product or service

143 Note 1 to entry: An alternative term to describe entries in a documentation artifact would be "data elements,"
144 following standards such as ISO/IEC 11179-31:2023, where "data element" denotes a unit of data within a dataset.
145 However, using this term in this document would have created confusion between a unit of information in a
146 documentation artifact and a unit of data within a documented dataset.

147 **Subfield**

148 *field* that is a structured component of another *field*

149 **Parent field**

150 *field* of which a *subfield* is a component

151 **Root field**

152 *field* that has no *parent field*

153

NOTE FOR REVIEWERS

NIST is soliciting community feedback on the most appropriate terminology for what we have termed a “documentation artifact.” We welcome input on whether “artifact” is the clearest term for the record describing a product or service. If this term seems potentially inadequate or misleading, please suggest a more precise and suitable alternative.

154

155 **Definitions to be referenced from ISO/IEC 22989 (or amendments to it):**

156 **AI system**

157 engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given
158 set of human-defined objectives

159 Note 1 to entry: The engineered system can use various techniques and approaches related to *artificial intelligence*
160 to develop a *model* to represent data, *knowledge*, processes, etc. which can be used to conduct *tasks*.

161 **AI model**

162 *model* used in an *AI system*

163 Note 1 to entry: In this document, an AI model is considered to include only the model architecture and
164 parameters. Related system components, such as post-processing modules, are not included.

NOTE FOR REVIEWERS

The current scoping of “models” does not encompass components with which a model is typically shipped or served (e.g., guardrail classifiers that operate on model output). Expanding the scope beyond the model object—i.e., beyond the architecture and parameters—would touch on the system level, raising the question of how to document AI systems in their full architectural complexity. That question has been excluded from this document’s scope given the dearth of literature for standardization to build on. NIST has not identified a principled basis for scoping in certain system components, such as those sometimes referred to as part of a model, while scoping out others. If reviewers believe model documentation should encompass additional components, NIST welcomes input on how the scope for model documentation can be expanded without then needing to encompass entire systems.

165

166 **Dataset**

167 collection of data with a shared format

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

169 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*.

3 Dimensions of potential outcomes of public-facing documentation processes

3.1 General

The decisions of an AI dataset or model¹ provider about what and how to document can lead to wide variations in outcomes. As the provider makes these decisions, it is typically guided by organizational objectives, which can be thought of as preferences over possible outcome scenarios. These scenarios can be broken down into key variables, or outcome dimensions, that matter to the provider, such as the degree of consumer trust or the extent to which business confidential information remains confidential. This clause describes outcome dimensions that are both strongly affected by documentation processes and commonly referenced when assessing the degree of success, benefit, or harm of a given outcome scenario.

Despite this focus on positive and negative outcomes, the outcome dimensions are not presented below as a list of benefits and drawbacks. Rather, each dimension is conceived of as a variable that an organization would often foresee being realized differently under different possible outcome scenarios. For example, once a model and its documentation are released, it is possible that consumer trust will be very high, very low, or anywhere in between. From the perspective of the model provider designing and executing documentation processes, all of these are projected outcomes along the consumer trust dimension with the potential to eventually be realized.²

This conceptual structure provides three advantages:

1. It recognizes that the same outcome on a given dimension can be judged as positive, neutral, or negative depending on contextual factors such as use case, provider organization objectives, and which interested party is making the judgment. For example, documentation can help others to build on a model, but whether an organization considers this outcome to be highly desirable can depend on its objectives.
2. It accounts for the fact that even for a given context, outcomes are typically not binary; rather, they fall on a sliding scale of benefit or cost along each dimension. For example, thorough documentation might provide maximal benefits with respect to trust, but superficial documentation might still provide greater benefits than no documentation. Whether the trust outcome from superficial documentation is considered beneficial, adequate, or disappointing thus depends on the point of comparison.
3. It clarifies in a purely descriptive fashion the empirical basis (Clause 3) for normative guidance and judgments (Clause 4), supporting decision-making about when and how to apply the guidance.

This document considers three types of dimensions: (1) Process-driven impacts from the documentation process itself (primarily experienced by the originating organization); (2) artifact-driven impacts from making documentation artifacts available for others to use; and (3) jointly driven impacts, i.e., impacts affected by both the process followed and the artifacts produced.³

¹ This document assumes that documentation is provided for a model alone, rather than for a system that includes additional output post-processing (e.g., classifiers for large language model guardrails) or other architectural components that are sometimes referred to as part of an AI model. If these components are included in the scope of documentation, then much of the guidance and templates in this document will still apply, but they would likely require some adaptation or clarification.

² The concept of outcome dimensions can be compared to random variables in statistics. Each dimension can be thought of as a random variable representing some aspect of the state of the world as it will stand after the dataset or model and corresponding documentation are released. A specific outcome along a dimension is analogous to the corresponding random variable taking a particular value (an “event” in statistical terminology). For example, high trust in a company or product is equivalent to the random variable for trust taking a high value.

³ Processes and artifacts cannot be entirely divorced from each other; the processes are designed to yield the artifacts, and the artifacts are the result of the processes. Still, they can be considered in isolation: a process-driven impact would occur even if the artifact were discarded before release of the dataset or model, and an artifact-driven impact would occur even if

While each outcome dimension is presented independently, dimensions can interact with each other. For example, information about evaluation results that is useful for suitability assessment can also be useful to malicious actors. Clause 4.4 discusses trade-offs between the qualities of documentation artifacts and processes that affect various dimensions.

All outcome dimensions discussed below can be strongly affected by public-facing documentation, i.e., documentation artifacts available to anyone who seeks them. However, these dimensions are not necessarily unique to public-facing documentation as opposed to documentation that is provided to more limited sets of recipients (e.g., business partners or government bodies).

AI documentation processes often lead to changes in the degree of transparency that stakeholders have into the dataset or model and its creators, particularly into their characteristics and history. These shifts in transparency in turn can lead to changes in outcomes along the dimensions discussed below. For example, describing a model's training procedures in documentation can clarify how the originating organization framed the model's task and sought out data. This information can help a potential user determine whether the model is fit for their purposes or compliant with their policies. Transparency thus functions as an intermediate effect that modulates other outcomes of interest. Recognizing these interactions can drive decision-making about how to design documentation processes to properly calibrate the resulting transparency.

3.2 Process-driven outcome dimensions

This section describes a noncomprehensive list of dimensions that can be impacted by the documentation process itself. These impacts are primarily experienced by the originating organization.

3.2.1 Organizational resource demands

Documentation processes consume some portion of an organization's time, finances, and other resources, usually an amount proportional to the extensiveness of the documentation processes. Conversely, a more extensive process can correlate with lower resource consumption at later points, because more extensive documentation can facilitate later documentation and decision-making efforts (and vice versa).

3.2.2 Internal collaboration

Documentation processes influence how effectively teams collaborate within themselves and among each other: the more thorough an organization's documentation processes are and the more integrated they are across teams, the easier intra-organizational collaboration is likely to be (and vice versa). In particular, more thorough documentation is likely to support more seamless collaboration among staff across technical and governance roles, including executive leadership, policy, legal and compliance, marketing, product management, risk management, and other business-critical functions.

3.2.3 Internal decision-making

Robust documentation requires articulating information from many stages of a dataset's or model's lifecycle, including data sourcing decisions, model design choices, and evaluation procedures. In the process of documenting, an organization can compile information that might otherwise remain fragmented across internal stakeholders. The information needed by public audiences often differs from what is needed by internal stakeholders, but the information compiled during processes for public documentation can still be useful for internal teams as they develop, maintain, update, and retire the datasets or models. For example, documentation can provide information about personally identifiable information or other privileged information included in a dataset, which can be critical to procedures for removal, archival, destruction, stakeholder notification, and

accurate documentation somehow appeared with no internal process on the part of the dataset or model provider. For a jointly driven impact, these conditional independence relationships do not hold.

version transition planning. Accordingly, an organization's ability to make decisions deliberately and with reliable information depends partly on what documentation processes are implemented.

3.2.4 Release speed

Documenting a dataset or model can require additional steps in the development and release process. Eliminating or shortening these steps can lead to faster release cycles, whereas including or extending them can slow down release cycles. Conversely, increased documentation efforts can also sometimes accelerate release cycles by making it easier for internal stakeholders to obtain the information they need (and vice versa for decreased efforts).

This outcome dimension is particularly significant for more agile release cycles that involve publishing dataset or model updates at close intervals.

3.3 Artifact-driven outcome dimensions

Many impacts of documentation activity arise when the information in a documentation artifact is received and interpreted by parties external to the originating organization, or at least external to the teams developing the dataset or model.

The interested parties impacted by a documentation artifact need not be direct recipients of the documentation. For example, release of private information affects those whose information is released even if they are completely unaware of the relevant dataset or model and never see the documentation.

The following is a noncomprehensive list of artifact-driven outcome dimensions.

3.3.1 Reuse and integration

Decisions about documentation artifacts can affect the ability of external parties to efficiently locate models or datasets and assess their suitability for a given purpose so that they can ultimately build on them.

3.3.1.1 Discoverability and searchability

Decisions about documentation artifacts can affect the ability of external parties to discover and search for models or datasets. Artifacts with conventionalized content and structure can help external parties search repositories for datasets or models more easily than less standardized artifacts. While implementers will often need additional detailed information, the presence of easily searchable fields can enable them to identify candidate datasets and models for further investigation.

3.3.1.2 Suitability assessments

Documentation artifacts affect recipients' ability to assess quality and suitability for their purposes, including via the extent and manner of documentation of testing, evaluation, verification, and validation (TEVV) results. For example, documentation artifacts can, to a greater or lesser extent, do any or all of the following:

- Communicate characteristics that may be important for assessing suitability (e.g., dataset sourcing or model performance).
- Describe the contexts of use for which the dataset or model was designed.
- Explain technical requirements for use and considerations for implementation.
- Communicate the objectives, findings, and limitations of TEVV activities and metrics to facilitate assessment, including by indicating uses for which the dataset or model is or is not suitable.
- Facilitate reproducibility of experiments or TEVV activities.
- Improve evaluation practices themselves (e.g., the process of documenting how evaluation activities cover the intended use cases can reveal additional scenarios or test cases to assess). Such improvement in turn can lead to communicating more characteristics that are more helpful for assessing suitability.

- Communicate post-deployment performance data and real-world impact information when available.

3.3.1.3 Implementation and further development

Once an external party has determined that a dataset or model is suitable for its use, the information (or lack thereof) in a documentation artifact can continue to affect its ability to effectively build on the dataset or model. For example, the more clearly and thoroughly the documentation describes known benefits and risks (including risk mitigation measures already taken), the more able the external party will likely be to leverage the dataset or model to its maximum potential and to appropriately choose risk mitigation measures.

3.3.2 Malicious use or attacks

Documentation can provide malicious users or attackers with information that helps them misuse or compromise the dataset or model. The thoroughness, level of detail, and types of information included in a documentation artifact can affect how likely this impact is to occur.

3.4 Jointly driven outcome dimensions

The outcome dimensions below depend on both the documentation process and the published documentation artifact.

3.4.1 Accountability

Documentation can provide or withhold information about the processes and decisions of the organization that created the dataset or model. How much information is provided can affect the extent to which those processes and decisions can be examined and evaluated by interested parties, especially those outside the organization. Processes for development, decision-making, and post-deployment monitoring that are more thoroughly documented are typically more traceable and attributable (and vice versa).

This correlation is particularly noteworthy regarding decisions about risk management. For example, if an organization more thoroughly documents decision-making processes and conclusions around risks to accept, avoid, transfer, or mitigate, then these processes and outcomes are typically more open to deliberation and examination.

The degree of accountability is determined largely by the scope, level of detail, and accuracy of the published information, i.e., by the artifact's contents. Beyond the artifact, however, the documentation process itself affects accountability within the organization: documentation processes can, to a greater or lesser extent, enable or incentivize documentarians, subject matter experts, and others within the organization to examine other decisions and processes, including adherence to organizational policies and safeguards.

3.4.2 Privacy preservation

Documentation artifacts can include sensitive information about individuals whose information is present in a dataset or reflected in a model. The most straightforward impacts of documentation on privacy preservation thus occur via the documentation artifact: the extent to which artifacts include or exclude such information determines how well they protect these individuals' privacy.

Privacy impacts can also occur via the documentation process alone. For example, as information about a sensitive dataset is compiled internally during the documentation process, an organization can recognize that the dataset should be more fully de-identified to prevent later re-identification, enabling the organization to improve its risk posture. Conversely, compiling the information can enable individuals in the organization to re-identify people in the data without proper authorization. Either case represents an impact on privacy even if the documentation artifact is never published.

328 **3.4.3 Proprietary information protection**

329 Documentation processes can protect or reveal proprietary information. For example, there can be information
330 about a dataset that would enable competitors or attackers to draw conclusions about products or services in
331 development. Documentation processes and artifacts can reveal more or less of this information.

332 Similarly to privacy preservation, impacts on proprietary information protection can occur either via what is
333 included in the documentation artifact or via what information becomes visible to actors within the organization
334 during the documentation process. For example, documentation workflows can replicate sensitive data to
335 environments lacking sufficient access controls, risking unauthorized disclosure.

336 **3.4.4 Legal obligations**

337 Documentation processes can contribute, to varying degrees, to successful or unsuccessful fulfillment of
338 obligations under relevant laws, regulations, or contracts (e.g., transparency or reporting requirements or
339 obligations regarding data privacy or security). Such obligations often entail both process requirements and
340 requirements on the provision of information, i.e., documentation artifacts.

341 **3.4.5 Trust**

342 Datasets and models are often developed so that external parties can use or build on them. Those parties'
343 willingness to do so depends partly on the extent to which they feel equipped to understand, in a manner that
344 suits their needs, how well considered the processes and decisions behind the product are and how the product
345 can affect various interested parties. The visibility that a documentation artifact offers can leave external parties
346 with varying degrees of understanding (or assurance that the desired understanding is attainable). External
347 parties sometimes also interpret the degree of visibility as a signal of the provider organization's openness and
348 thus, potentially, of its trustworthiness.

349 Building trust can require documenting many process- or decision-oriented aspects beyond the datasets' or
350 models' technical characteristics. Examples include procedures for decision-making, data collection and
351 recruitment, and post-deployment experimentation.

352 Documentation modulates trust largely via the content of the artifact: trust is correlated with many of the artifact
353 qualities noted in Clause 4.3. However, the process of documenting can also affect trust independent of the artifact.
354 Trust is affected partly by the extent to which external parties believe decisions about design, data collection,
355 post-deployment processes, and the like were made via a deliberative, risk-aware process. Even if these parties
356 never have the opportunity to view a documentation artifact, their level of trust can increase if they know that
357 there was a documentation process in which trust-relevant questions were considered. Similarly, organizational
358 staff often behave differently when they expect that their decisions will be made visible through documentation.
359 Accordingly, external parties will likely trust a dataset or model more if they believe that it was developed with
360 processes that staff expected would lead to external release of documentation, even if the artifact is never
361 released.

362 **4 Guidance on public-facing documentation processes and artifacts**

363 **4.1 General**

364 This clause provides recommendations for designing documentation processes and artifacts to be fit for purpose.
365 In defining its objectives (which in turn define fitness for purpose), an organization should consider what
366 outcomes it and other interested parties deem preferable on the dimensions discussed in Clause 3. Explicit
367 connections are made to these outcome dimensions where they are not necessarily obvious.

368 Organizations should calibrate the recommendations in this clause based on their contextual objectives, including
369 the audiences the documentation artifacts are meant to serve.

370 In particular, not all documentation processes are geared toward public documentation. Organizations should
371 expect that internal documentation will often be more thorough than what is ultimately released to the public.

372 The more thorough internal processes will often simultaneously support public, business-to-business, business-
373 to-regulator, or other documentation needs. The recommendations in this clause concern public-facing
374 documentation; they are not necessarily applicable to documentation processes and artifacts aimed at other
375 audiences.

376 **4.2 Guidance on documentation artifacts**

377 Documentation artifacts should be tailored to the needs, uses, domains, interested parties, and technical outcomes
378 that the artifacts are meant to serve given organizational objectives. In particular, public-facing documentation
379 will typically be less comprehensive than documentation for limited external audiences (e.g., customers, partners,
380 or regulators) or internal audiences (e.g., developers, implementers, or compliance personnel). A difference in
381 comprehensiveness is particularly likely with respect to sensitive information, such as trade secrets or personal
382 information.

383 The following is a noncomprehensive list of specific qualities that should be sought in public-facing
384 documentation artifacts.

385 **4.2.1 Artifact correctness**

386 Documentation artifacts should contain factual, technically precise, evidence-based information about the
387 datasets or models. Inaccurate artifacts will fail to achieve most desired outcomes while wasting resources, and
388 can even expose organizations to legal, reputational, or other risks.

389 The rigor and precision of the information should be tailored to the needs and constraints of the intended
390 documentation audience, which may imply variation across different documentation artifacts. See also Clause
391 4.4.2.

392 **4.2.2 Artifact comprehensibility**

393 Documentation artifacts should be clear and easy to understand for their intended audiences. Particularly for
394 public audiences, content should be as plain and concise as feasible and organized clearly in a structure that
395 matches reader expectations. Adequate comprehensibility is necessary in order to achieve positive outcomes on
396 the dimensions of accountability, trust, suitability assessment, and reuse and integration.

397 NOTE: What information is easy to understand varies significantly between intended audiences even
398 within the general public. In particular, highly technical information can be incomprehensible to some
399 audiences even as it provides other audiences with precisely the details they seek. See Clause 4.4.2.

400 **4.2.3 Artifact informativeness**

401 Documentation artifacts should be as extensive and detailed as needed for the purpose, describing as much as the
402 intended audience needs to understand about the AI dataset or model, the development process, and decision-
403 making behind them. To that end, information from all parts of the dataset or model lifecycle should be provided
404 as needed, including about post-release updates and the processes behind them, if applicable.

405 When considering the audience’s information needs, documentation authors should focus on what information
406 the audience will likely want to act on or use (e.g., for assessing suitability). For example, contact points or other
407 redress mechanisms can be important information for some audiences, and stakeholders sometimes need to
408 know the reasons for an update in order to apply it appropriately. Providing the breadth and depth needed by the
409 audience is often a precondition for achieving desired outcomes on the dimensions of trust, accountability, and
410 reuse and integration.

411 **4.2.4 Artifact judiciousness**

412 Documentation artifacts should include only information that is relevant and useful for public audiences and
413 appropriate for public disclosure. In particular, artifacts should omit private, proprietary, or legally impermissible
414 information.

Including unnecessary information can complicate an audience's search for the information it needs, particularly if the artifact is not organized in a manner that makes it easy to skip irrelevant information. This can negatively impact suitability assessment, reuse and integration, accountability, and trust.

Including inappropriate information can also compromise privacy or proprietary information, violate legal obligations, facilitate malicious use or attacks, or reduce trust.

4.2.5 Artifact freshness

For documentation artifacts presented as canonical or current, the information conveyed should be up to date with the most recent version of a dataset or model. Older versions of documentation might remain useful, and organizations should consider whether and how to store, make available, mark, and otherwise manage these documents.

Organizations should ensure that documentation is available for the most recent version of a dataset or model.

Some documentation artifacts are updated after the dataset's or model's release to include information that was not available at the time of release (e.g., post-deployment monitoring results). In such cases, readers generally need clarity about what period of post-release information is included in the artifact. The artifact should therefore clearly indicate when it was last updated, at least regarding the post-release portions.

4.2.6 Artifact interoperability

Documentation artifacts should follow established conventions for content, structure, and format, matching artifacts for other AI datasets and models. Following the templates in Clause 5 can help to achieve this aim.

Organizations should provide both machine-readable and human-readable artifacts, which may or may not be the same artifact. Human-readable artifacts support interested parties' decision-making, particularly parties without technical AI backgrounds. Machine-readable artifacts support technical accessibility, reusability, and analysis of the products (e.g., by facilitating searches). Each type of artifact should be tailored to the functions it supports.

Where the information in a machine-readable artifact cannot be fully structured, the artifact can still associate free text or other unstructured content with broad machine-readable fields or categories (e.g., intended use, evaluation results).

Interoperable artifacts enhance the findability of dataset and model documentation (e.g., via indexed metadata) and make the products easier to compare, ultimately enhancing trust, reuse and integration, suitability assessments, and accountability. Interoperable artifacts also support comprehensibility: those immersed in the discipline will find it easier and faster to make sense of an artifact in a familiar format. Interoperability can support continuous documentation processes (see Clause 4.2.6) by enabling automation of portions of the documentation process.

4.2.7 Artifact findability and availability

Documentation artifacts should be made available with minimal barriers to access. Specifically:

- Artifacts should be published so that they are easy to locate. In particular, they should be published along with pointers (e.g., hyperlinks), descriptions or metadata, and possibly accompanying materials (e.g., blog posts), all in intuitive locations (e.g., via well-organized access portals).
- Artifacts can be distributed via a number of mechanisms (e.g., open download, registration required, or tiered access). Mechanisms that minimize the number of steps and approvals required should be preferred, to the extent consistent with organizational objectives.

4.2.8 Artifact maintainability

Documentation artifacts and their components should be easy to create, extend, update, and maintain across the AI lifecycle.

Maintainability includes modularity. Information should be separated into distinct components, each of which can easily be replaced with an expanded or condensed version if desired or omitted altogether. (The templates in Clause 5 can facilitate such modularity.) Indeed, in some artifacts, it can be desirable to omit certain components.

Maintainability facilitates other important artifact and process qualities, including correctness, freshness, and process manageability. Maintainable artifact design can both prevent accumulation of unmet documentation needs and reduce resource expenditure over the AI dataset or model lifecycle.

4.3 Guidance on documentation processes

4.3.1 General

Documentation processes should be designed to yield artifacts with the qualities described in Clause 4.2, as well as to directly drive preferred outcomes from Clause 3.

As organizations develop their documentation processes, they should consider the needs of stakeholders relevant to their organizational objectives, decide on trade-offs, and maintain and continually improve their practices. For example, depending on use case, regulatory requirements, industry vertical, and third-party interests, organizations may have to weigh the interests of different parties in areas such as comprehensibility, completeness, and freshness.

Specific recommended practices for designing and executing documentation processes are provided in the subclauses below.

4.3.2 Define documentation objectives

Organizations should define their objectives for their documentation artifacts and processes and design the remainder of their documentation processes and the resulting artifacts accordingly. Objectives will likely involve many outcomes from Clause 3.

Organizations should assess, continuously or periodically, the extent to which their documentation objectives are being met.

4.3.3 Provide organizational support

To the extent called for by organizational objectives, organizational leadership should support documentation by allocating necessary resources (e.g., time, budget, and headcount) and establishing institutional incentives for prioritizing documentation (e.g., management recognition, explicit mention in performance evaluation, and associated compensation).

Resources and incentives should be tailored to the objectives, including desired outcomes from Clause 3.

4.3.4 Keep processes manageable

Documentation processes should be manageable to execute, given available resources. Otherwise, the organization risks exhausting its resources, and can risk compromising the documentation altogether or even the dataset or model. The availability of resources for documentation work and the proportional burden that these processes impose will vary substantially based on an organization's size and capacity, and processes should be calibrated accordingly.

To mitigate the complexity of collaborating on documentation and integrating it with other parts of the dataset or model lifecycle, organizations should:

1. Consider periodically creating internal documentation artifacts, perhaps partial or unfinished ones, with records from specific milestones. This approach helps documenters focus on recording relevant information at a given stage of development of an AI dataset or model, thus providing a trail of artifacts that can be further consolidated by others into a final documentation artifact for public release.
2. Designate documentarians and encourage them to play an active role. Documentarians should facilitate collaboration between others, state concerns, raise questions, seek clarifications, and facilitate an

environment of deliberation about the dataset or model development process as a whole. Documentarians should not, however, be expected to do all the work of documenting; accurate and complete documentation depends on knowledge from many subject matter experts involved in creating the dataset or model.

4.3.5 Provide guidelines

Organizations should create and maintain internal documentation guidelines for documentarians and subject matter experts, particularly regarding how to execute documentation processes and produce documentation artifacts that satisfy the organization's and interested parties' needs. Guidelines may need to reflect the variety of forms of documentation: organizations might need various sets of guidelines for the various ways they use documentation, with only some guidelines applying to processes for public-facing documentation.

4.3.6 Document continuously

Documentation activities should be undertaken iteratively throughout the lifecycle of the dataset or model.

Organizations should seek to avoid delaying documentation until the dataset or model is largely complete. Rather, documentation should be created in earlier stages of the lifecycle (e.g., in the design and development stage, planned data sources and collection processes can be documented).

Documentation should then be revised and fleshed out in later stages based on more detailed or up-to-date information (e.g., final task input format), findings/results to date, and transformations to the dataset or model (e.g., model retraining, data pipeline changes). It is often advisable for subject matter experts to dynamically create and adapt summaries of their technical work as they are doing it, in coordination with documentarians. Continuous documentation can include documenting post-deployment impacts or experimental results.

Revision throughout the lifecycle is necessary to keep artifacts up to date, and helps to compile the many decisions and design choices to facilitate adequate informativeness and a risk-aware mindset.

4.3.7 Distribute documentation work

Documentation work should be collaborative and distributed across multiple people and roles. Organizations should avoid assigning documentation responsibilities to a single person. Artifacts produced by distributed, collaborative processes will tend to be more informative and modular than those created by a single person, and the process will likely be more manageable and continuous. A distributed documentation process is also more likely to yield process-driven benefits such as facilitating internal collaboration and deliberative decision-making.

Accordingly, organizations should provide for multiple functions that will typically be served by multiple people. Various subject matter experts should provide asynchronous technical accounts of their own work developing specific dataset and model subcomponents (e.g., data collection, annotation, versioning, metrics). Documentarians should oversee and record progress of documentation processes, coordinating between other documenters.

4.3.8 Incorporate audience input

Documentation processes are meant mainly to produce artifacts that help the intended audiences. By incorporating audiences' input, provider organizations can better understand their needs. Audience input can help the organization to more quickly identify when existing artifacts are failing to exhibit desired qualities or to provide the intended benefits. For example, audience feedback can notify an organization if artifacts are proving hard to understand or include someone's personal information.

Accordingly, to the extent needed to achieve organizational objectives, documentation processes should allow for and incorporate feedback, user testing, and other input from the audiences, potentially up to and including co-design of the documentation artifacts. If speaking directly with intended audiences is difficult, organizations can rely on other sources of information about what interested parties need and how well existing documentation

artifacts are serving those needs. Examples of information-gathering mechanisms include surveys, relevant standards, and industry publications.

When input has been sought from audiences, both trust and accountability can be increased by clarifying in the documentation artifact or in associated documents what input was received, how it was or was not acted upon, and what benefits, risks, or constraints motivated those decisions. Organizations will need to weigh these outcomes against impacts on other outcome dimensions, particularly resource demands and judiciousness; see also Clause 4.4.

Many factors can shape the audience involvement called for by organizational objectives. Such factors include mission, resources, sensitivity of information, and redress or recourse needs.

4.4 Trade-offs between qualities of documentation artifacts and processes

While each of the qualities in Clauses 4.2 and 4.3 is valuable, maximizing one often comes at the expense of others, necessitating that organizations make trade-offs.

Organizations should make trade-offs between qualities of documentation artifacts and processes based on their context and organizational objectives.

While the specific trade-offs can vary, they often include those listed below.

4.4.1 Informativeness vs. judiciousness

There is often information that would be useful to some interested parties, but which other interested parties, possibly including the originating organization, would prefer or even be obliged to protect. For example, if a potential user of a model that processes personal information wants to assess the model's usability, it is possible that they would benefit from extensive details about the training dataset. At the same time, it is possible that some of those details would reveal sensitive information about the individuals whose data were included.

As organizations evaluate such trade-offs, they should both consider user consent regarding data use and sharing and respect any prior agreements regarding consent. Similarly, some information can be useful to both legitimate interested parties and malicious attackers, so its release can require careful judgment.

4.4.2 Informativeness vs. comprehensibility, maintainability, and process manageability

Documentation that is more thorough can provide more of the information that interested parties need. However, long or highly detailed documents can be burdensome or overwhelming, potentially making the documentation difficult to understand or use. Such documentation can also tax organizations' resources for creating and maintaining it. This concern is particularly relevant for public documentation, for which non-expert members of the public are frequently part of the intended audience.

Organizations can partly address the comprehensibility issue by providing different versions of the documentation that address different audiences, at the expense of further burdening the documentation's development and maintenance processes. It can also be helpful to adhere to a clear structure that enables some audiences to skip portions they do not need to understand. In some documentation formats, progressive disclosure mechanisms can further facilitate use by multiple audiences.

4.4.3 Informativeness vs. freshness

The more thorough a documentation artifact is, the more difficult it can be to keep it current, at least while keeping documentation processes manageable.

4.4.4 Continuous documenting vs. keeping processes manageable

Documenting throughout the lifecycle can yield better outcomes, but it can also burden an organization's resources more and make documentation workflows more complex.

5 Documentation templates

5.1 General

The templates in this Clause are designed to capture information that is frequently needed by interested parties across many kinds of organizations, roles, application domains, geographic regions, and datasets or model types. The templates are designed to help organizations to implement the guidance from Clause 4 and to achieve better outcomes on the dimensions discussed in Clause 3.

Documentation artifacts for AI datasets and models shall adhere to the structure and requirements of the templates in this Clause. Specifically:

- If a documentation artifact includes a field from this Clause, that field shall be populated with information matching the field description and guidance. Information not relevant to the field's stated description and guidance shall not be included.
- If a documentation artifact includes information matching the description and guidance of a field, that field shall be present in the artifact and shall contain said information (or a reference to it).
- All fields designated in the templates as required shall be present and populated in a documentation artifact.

Fields marked as "recommended" should be populated under most circumstances, but are not required to be.

A field may be populated with any mix of free text (potentially including formatted or multimedia content), structured values, references to other fields, references to external artifacts, or subfields. Subfields may likewise be populated with any mix of these types of data.

A documentation artifact may contain additional top-level fields beyond those listed in this Clause.

These templates are expected to guide documentation artifacts across all contexts, so they have been designed to maximize flexibility. The templates consist of root fields only, with most fields designated as optional. In contrast, more detailed subfields with more recommendations and requirements are offered in Annex A's default profile (see Clause 6). This profile can accommodate more prescriptiveness because organizations can always resort to a modified or alternative profile for contexts where the prescriptions are not well suited.

The guidance in Clause 4 should be followed when using these templates. Organizations should follow the guidance on documentation artifact qualities in Clause 4.2 when determining how to populate a field and how to prepare and disseminate the final artifact. Similarly, although the templates cannot capture organizational process elements, organizations should follow the process guidance in Clause 4.3 as they work with the templates. For example, as part of continuous documentation, it is possible that an organization would fill out templates, perhaps at different levels of detail, at different stages of the lifecycle.

For examples of the kinds of information appropriate for each field, see the subfields specified in the default profile in Annex A.

NOTE TO REVIEWERS

Stakeholder input has indicated a desire for concrete examples for the templates in Clause 5 and the default profile in Annex A. NIST welcomes input on types and forms of examples that would best help readers to understand what content is appropriate for each field and how to apply the guidance. Choices to make include:

- Whether to list example types of information that could be included in a field inline with the template. The main alternative, currently implemented, is to simply refer to the default profile as a source of examples.
- Whether to provide examples of filled-in fields for some representative dataset(s) and model(s).

- If yes, whether to provide these filled-in examples inline with the templates vs. providing one or more annexes with complete example documentation artifact(s).

NIST also welcomes concrete examples that could be provided in the text, as well as views on the extent to which it would be helpful to draw explicit links between the fields in Clause 5 and Annex A and the guidance in Clause 4.

620

621 **5.2 Dataset documentation template**

622 The template for documenting datasets can be found in Table 1.

623

Table 1: Dataset documentation template				
ID	Root Field Name	Description	Designation	Additional Guidance and Requirements
1	Identifying Descriptors	Information that uniquely identifies the dataset or characterizes its conditions of creation	Required	May contain any information that would help interested parties to discover, obtain, reference, or attribute the dataset or to manage different versions of it. Shall provide enough information to determine whether a given dataset is the one described by the documentation artifact.
2	Intended Use	Information characterizing the applications for which the dataset was designed	Optional	May provide information on primary applications for which the dataset is suitable, which may include applications the dataset is known to be unsuitable for. It is often helpful to align the information in this field with evidence described in Field 5.
3	Usage Rights and Restrictions	Information about terms, conditions, and rights governing the use, sharing, and modification of the dataset	Optional	May provide information about the legal or policy framework within which use, modification, and redistribution of the dataset is expected to occur. Such information can help interested parties to determine before use whether their intended application is permissible and under what conditions. This is distinct from Dataset Governance (Root Field 7), which covers the practices followed by the provider organization during the dataset’s development and release. Information in this field may include requirements or restrictions related to handling of personally identifiable information or other sensitive information included in the dataset.
4	Composition and Provenance	Information characterizing what	Recommended	May provide information about how, by whom, and when the data has been designed (if applicable), acquired, or processed, which may include information about efforts to ensure high-quality data. Composition and

Table 1: Dataset documentation template

ID	Root Field Name	Description	Designation	Additional Guidance and Requirements
		the dataset contains and how it was built		provenance information can help interested parties to assess what purposes the dataset's construction process makes it suitable for, including what a model trained or evaluated on the dataset is likely or unlikely to be able to do.
5	Evaluation	Evidence regarding the dataset's quality and fitness for use	Optional	May provide interested parties with evidence that helps them assess how well the dataset performs its intended function or potential risks regarding use of the data in model training or evaluation. This may include descriptions of the underlying evaluation activities undertaken during or after data acquisition and processing. If evaluation activities occurred as part of dataset construction, descriptions of such activities may be incorporated by referencing content from Root Field 4 (or vice versa).
6	Maintenance and Monitoring	Information about processes and responsibilities for dataset oversight and curation after release	Optional	May provide information about how, by whom, and when the dataset is updated, versioned, shared, maintained, and ultimately retired. Maintenance and monitoring information can help interested parties to understand the operational lifecycle of the dataset—whether it is actively maintained, how changes are tracked and communicated, and what will happen when it is retired—which is often important for assessing suitability and integrating the dataset into an organization's operations.
7	Dataset Governance	Information characterizing organizational practices applied over the dataset lifecycle	Recommended	May describe frameworks or organizational practices that guided the provider organization's activities over the course of the dataset's lifecycle. Also may include, reference, or describe documents or records resulting from the application of these frameworks or practices. This field can help an interested party to assess whether the dataset was produced in conformance with applicable obligations or policies and to integrate the dataset and the documentation artifact with the interested party's own organizational policies. The focus on the provider organization's activities differentiates this field from Usage Rights and Restrictions (Root Field 3), which addresses the terms under which another entity may work with the dataset.

625 **5.3 Model documentation template**

626 The template for documenting models can be found in Table 2.

Table 2: Model documentation template				
ID	Root Field Name	Description	Designation	Additional Guidance and Requirements
1	Identifying Descriptors	Information that uniquely identifies the model or characterizes its conditions of creation	Required	May contain any information that would help interested parties to discover, obtain, reference, or attribute the model or to manage its different versions. Shall provide enough information to determine whether a given model is the one described by the documentation artifact.
2	Intended Use	Information characterizing the applications for which the model was designed	Optional	May provide information on primary applications for which the model is suitable, which may include applications the dataset is known to be unsuitable for. This may also include information about the intended task or tasks, intended application domain or domains, functionality offered, and types of AI system that the model can be integrated into. It is often helpful to align the information in this field with evidence described in Field 6.
3	Usage Rights and Restrictions	Information about terms, conditions, and rights governing the use, sharing, and modification of the model	Optional	May provide information about the legal or policy framework within which use, modification, and redistribution of the dataset is expected to occur. Such information can help interested parties to determine before use whether their intended application is permissible and under what conditions. This is distinct from Model Governance (Root Field 8), which covers the practices followed by the provider organization during the model’s development and deployment.
4	Design	Information describing of the model’s architecture and technical assumptions	Optional	May describe how the model is structured and how it is meant to be used within a data processing pipeline, i.e., what inputs it expects and what outputs it produces. This field may include information about objectives or observations that drove design decisions. Model design information can help interested parties to assess what purposes the

Table 2: Model documentation template

ID	Root Field Name	Description	Designation	Additional Guidance and Requirements
				model's design makes it suitable for, including what it is likely or unlikely to be able to do.
5	Training	Information about protocols and data used in optimizing the parameters of a model	Optional	May describe the training protocol, which may include information about datasets used and any pre-processing applied, training objectives optimized for, or hyperparameter configurations. Any information about training data should be structured according to the dataset template in Clause 5.2, or should reference documentation artifacts that follow that template.
6	Evaluation	Evidence regarding the model's quality and fitness for use	Optional	May provide interested parties with evidence that helps them assess how well the model performs its intended function or potential risks regarding use of the model. This may include descriptions of the evaluation protocols and datasets, documented analyses of risks and potential impacts, quantitative evaluation results, and qualitative analysis. If information about evaluation data is included, this field may incorporate or reference documentation structured according to the dataset template in Clause 5.2, or equivalent information.
7	Maintenance and Monitoring	Information about the processes and responsibilities for monitoring and maintaining the model after release	Optional	May provide information about how, by whom, and when the model is updated, versioned, shared, maintained, and ultimately retired. This may include information about mechanisms, plans, and responsibilities to monitor model performance and behavior in deployment and report and resolve incidents. Maintenance and monitoring information can help interested parties to understand the operational lifecycle of the model—whether it is actively maintained, how changes are tracked and communicated, how issues will be detected and responded to, and what will happen when it is retired—which is often important for assessing suitability and integrating the model into an organization's operations.

Table 2: Model documentation template

ID	Root Field Name	Description	Designation	Additional Guidance and Requirements
8	Governance	Information characterizing organizational practices applied over the dataset lifecycle	Recommended	May describe frameworks or organizational practices that guided the provider organization’s activities over the course of the model’s lifecycle. Also may include, reference, or describe documents or records resulting from the application of these frameworks or practices. This field can help an interested party to assess whether the model was produced in conformance with applicable obligations or policies and to integrate the model and the documentation artifact with the interested party’s own organizational policies. The focus on the provider organization’s activities differentiates this field from Usage Rights and Restrictions (Root Field 3), which addresses the terms under which another entity may work with the model.

6 Profiles

Each template in Clause 5 may be extended via profiles, i.e., documents that can be created by any interested party to adapt the templates to particular contexts of use (e.g., an AI application domain, a technical category of dataset or model, a risk level, or a geographic area with specific policy requirements). A profile shall not change or contradict the structure, requirements, and guidance of the templates in Clause 5, except to elevate a root field's designation, i.e., to change it, possibly conditionally, from optional to recommended or required or from recommended to required. A profile may additionally specify any subset of the following:

- Additional subfields of the root fields provided in Clause 5, i.e., more granular structure with which the fields may be populated. Subfields may also be nested into a hierarchy, with subfields within subfields.
- Additional root fields beyond those provided in Clause 5, which may also include subfields or even a hierarchy of subfields.
- Designations for fields indicating whether they are required, recommended, or optional for the intended context. Fields may also be designated as conditionally required or conditionally recommended, in which case the profile shall clarify the relevant conditions. For example, a profile for medical datasets would likely make a privacy policy field required, while a profile for astronomical imagery datasets would likely make the field optional. There would likely be other domains for which profiles would require or recommend referencing a privacy policy only if one exists.
 - A field may be required under some conditions and recommended under others, in which case “conditionally required” shall be used as the designation, but the conditions under which it is recommended shall also be specified.
- Additional information, guidance, or requirements regarding how to populate fields, which may include nuanced criteria that the contents of one or more fields should or must meet. Guidance and requirements may include conditions under which information should or must be omitted from one or more fields.

For example, if a profile specifies version number and release date as subfields of the Dataset Identifying Descriptors field, it can specify that an artifact must provide either a version number or a release date, and that the release date should be specified in a particular format.

If a profile specifies subfields for a parent field, it can become ambiguous how “required,” “recommended,” and “optional” designations for the parent affect the subfields and vice versa. To resolve these ambiguities, a profile shall follow the following conventions:

- To indicate that a subfield of a given parent field is conditionally or unconditionally recommended or required, a profile shall make this designation explicitly about the subfield. Designating a parent field as recommended or required under some set of circumstances shall indicate only that the field should or must be populated with something, but not necessarily that any particular subfields should or must be populated, nor even that the subfields must be used to structure the contents of the parent field.
- If a subfield is designated as required under some set of circumstances (including unconditionally), its parent field shall be designated as required under at least the same set of circumstances. For example, consider a dataset documentation profile that introduces a “Privacy Policy” subfield of Root Field 3 (Usage Rights and Restrictions) and mandates it if the dataset includes personally identifiable information (PII). Root Field 3 must then also be required if the dataset includes PII (or unconditionally). However, a requirement to populate a subfield can be conditioned on the parent field being populated, in which case the parent field may be designated as recommended or optional. For example, a profile may indicate that Root Field 3 is optional, but that if it is populated, it must include a Privacy Policy subfield.
- Similarly, if a subfield is designated as recommended under some set of circumstances (including unconditionally), the parent field shall be designated as recommended or required under at least the same set of circumstances. However, a recommendation to populate a subfield can be conditioned on the parent field being populated, in which case the parent field may be designated as optional.

- Any field lacking a designation shall be assumed to be designated as optional, except where this would conflict with Clause 5 or the conventions above.

When an attestation is made that a given documentation artifact conforms to this standard, it shall indicate either conformity to the base templates from Clause 5 or conformity to a specific profile of the templates. The following conditions shall be met for a documentation artifact to be considered conformant to a profile:

- The artifact is conformant to the corresponding base template from Clause 5.
- If the artifact includes a field for which the profile specifies descriptions or guidance, that field is populated with information matching what is specified by the profile, and information not relevant to the stated scope of the field is not included.
- If the artifact includes information matching the descriptions and guidance that the profile specifies for a given field, that field is present in the artifact and contains said information (or a reference to it).
- The artifact includes all fields designated by the profile as required, as well as any fields designated as conditionally required where the relevant conditions apply.
- The artifact meets all other requirements specified by the profile.

A subfield specified by a profile may be populated with further nested subfields not specified by the profile. For example, if a profile provides a single “Risk analyses” subfield, a conforming documentation artifact may populate that subfield with discussions of a variety of risks, with one subfield per risk nested under “Risk analyses.”

When a field is recommended or conditionally recommended by a profile, that field should be populated under any specified conditions. However, if the field is missing or empty, an attestation of conformity to the profile would still be valid.⁴

Annex A provides example profiles that specialize the templates from Clause 5 with subfields, designations, requirements, and guidance that will likely be applicable to many settings.

⁴ Conformity assessment schemes can always specify that a recommendation in the underlying standard or specification is required under the scheme. Thus, a conformity assessment scheme for a profile can specify that a field recommended by the profile is required.

NOTE TO REVIEWERS

NIST welcomes any efforts to pilot the application of the base templates in Clause 5, the default profiles in Annex A, and any other profiles for which organizations might identify a need. NIST invites stakeholders to share learnings from such pilot efforts. Even after the zero draft is published, it will likely remain helpful to have information from real-world experiences as the work is integrated into standards (assuming this draft is developed further by a standards developing organization).

NIST also welcomes any suggestions for clarifying the conventions, guidance, and requirements in this clause, as needed, so that they can be applied as readily as possible by many kinds of actors.

697 **A Default profiles**

NOTE TO REVIEWERS

NIST welcomes input on any changes needed to the default profiles to make them applicable to as wide a range of contexts as possible. In particular:

- If any changes should be made to fields’ **designations**, NIST invites explanations of how those changes would improve both outcomes along the dimensions in Clause 4 and implementation of the guidance from Clause 5 across a wide range of organizations, application domains, geographic regions, and AI dataset or model types.
- NIST also invites input on how best to clarify the intended content of each field so that documenters can reliably and consistently populate them. The approach implemented in the fourth column below is to offer a **description** structured like a definition—a noun phrase describing content that would be appropriate for the field—sometimes with additional **guidance** on how to interpret the designation and description. An alternative would be to frame the fourth column as prescriptions or instructions, with indications of what a field “shall,” “should,” or “may” include (see Foreword regarding these ISO/IEC terms). In either case, as discussed in a note above, it would also be possible to provide illustrative examples, perhaps in the form of a fully populated documentation artifact based on the profile.

Relatedly, NIST welcomes input on whether the fields are sufficiently distinguished from each other, noting that a small amount of semantic overlap may be acceptable (particularly because fields can reference each other).

698 **A.1 General**

699 This Annex provides default profiles for AI datasets and AI models. These profiles specialize the templates from
700 Clause 5 using the profile construct described in Clause 6, and may be used in attestations of conformity.

701 The subfields, designations, requirements, and guidance in these profiles are based on extensive stakeholder
702 input to NIST regarding what fields are commonly needed to meet the documentation objectives established in
703 Clauses 3 and 4. The profiles thus provide a common foundation intended to suffice for many contexts, either as
704 a profile that documentation artifacts conform to directly or as a basis for more specialized profiles with additions
705 or modifications. However, there will likely be some contexts for which profiles will need to differ considerably.

706 The various designations and their implications are defined in Clause 6. Broadly, “Required” means an artifact can
707 only be conformant with the profile if the field is populated; “Recommended” means the field should be populated,
708 but an artifact can be conformant without it; and “Optional” means there is no particular expectation that the field
709 will be populated.

710 In places, these tables elevate the designation of a root field from Clause 5 (e.g., to change it from optional to
711 recommended or required). The rationale for this, as discussed in Clause 5, is that the tables there are designed
712 to maximize flexibility, whereas the default profiles can be more prescriptive.

713 Each template includes a subfield about risk and impact analysis (Field 5.1 in Table A.2.1 and Field 6.6 in Table
714 A.3.1). Most of the other subfields can also in principle be linked to risks of a model or dataset: they can describe
715 risk assessments or risk treatments. For example, evaluation procedures can include evaluation instruments
716 aimed at assessing risk, and training procedures can include mechanisms intended to reduce an identified risk.
717 Populating a subfield can also constitute a risk treatment: stating an out-of-scope use can mitigate the risk of
718 someone inadvertently using a dataset or model in that way. The templates are organized in terms of technical
719 characteristics and procedures; however, if documentation authors instead wish to organize information in terms
720 of risks, this can be accomplished by placing much of the information in Fields 5.1 and 6.6, and then referencing
721 those fields from the others.

722 **A.2 Default AI dataset profile**

723 Table A.2.1 provides a list of fields from Clause 5.3 along with subfields and designations. Where further
724 explanation is required to clarify a designation or description, it is provided in the “Additional Guidance and
725 Requirements” column.

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
1	Identifying Descriptors	Information that uniquely identifies the dataset or describes its conditions of creation	Required	Shall provide enough information to determine whether a given dataset is the one described by the documentation artifact.
1.1	Identifier	One or more canonical names of the dataset, or dynamic datasets bundle, if applicable	Recommended	
1.2	Description	Brief human-readable summary of the dataset’s characteristics	Required	
1.3	Authorship Information	Information about who created the dataset and documentation	Recommended	
1.3.1	Authors, Creators, and Developers	One or more names of responsible parties, i.e., individuals or entities who created or developed the dataset	Recommended	
1.3.2	Documentation Authors	One or more names of entities who prepared the documentation artifact	Optional	
1.4	Point of Contact	Point of contact for feedback or questions related to the dataset	Conditionally recommended	Recommended if the dataset can be accessed by entities outside the provider organization; otherwise optional.
1.5	Version Information	Information uniquely associated with a specific state of the dataset	Conditionally required	See field 1.5.1 and 1.5.3.
1.5.1	Version Identifier	Formal identifier for the iteration of the dataset or the dynamic dataset bundle being documented	Conditionally required	Required if the dataset can be accessed by entities outside the provider organization; otherwise optional.

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
1.5.2	Cryptographic Signature	Cryptographically signed manifest containing one or more cryptographic hashes for the dataset instances	Optional	
1.5.3	Release Date	Date on which the documented version of the dataset was first made available in its complete form	Conditionally required	Required if the dataset can be accessed by entities outside the provider organization; otherwise optional.
1.5.4	Decommissioning Date	Date by which the provider organization ceased, or plans to cease, supporting or circulating the dataset	Conditionally required	Required if the provider organization has planned or acted upon such a date; otherwise optional.
1.6	Release Methods	Information describing methods by which the dataset can be accessed or obtained (e.g., cloud service providers, product surfaces, websites where weights can be downloaded)	Conditionally recommended	See field 1.6.1.
1.6.1	Source URL	URL from which the dataset can be obtained (or which provides information on how to obtain it), if applicable	Conditionally recommended	Recommended if the dataset can be obtained via the internet; otherwise optional.
1.7	Citation	Technical citation format for referencing the dataset or the dataset bundle	Conditionally recommended	Recommended if the dataset can be accessed by entities outside the provider organization; otherwise optional.
1.8	Attributions	Notices or attributions that are to be presented in association with the dataset, if applicable	Conditionally required	Required if Usage Rights and Restrictions specifies a mandate to attribute; otherwise optional.
1.9	Documentation Version Identifier	Formal identifier for the iteration of the present documentation artifact	Required	Even for a single dataset version, different versions of the documentation artifact sometimes need to be published (e.g., to include newly available information or correct errata). To address such situations,

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
				this field shall provide an identifier uniquely associated with the present version of the artifact.
2	Intended Use	Information characterizing the applications for which the dataset was designed	Conditionally recommended	See fields 2.1 and 2.3.
2.1	Intended Task	Information describing one or more applications or use cases that the dataset is intended to support (e.g., image segmentation, translation, summarization)	Conditionally recommended	Recommended if there are one or more defined tasks that the dataset is intended to be used for; otherwise optional.
2.2	Applicable Domains	Sectors or domains in which the dataset is designed to be applied	Optional	
2.3	Limitations	Information describing design constraints, coverage gaps, or other limitations that would likely constrain use of the dataset	Conditionally recommended	Recommended if the dataset provider knows of constraints, gaps, or other limitations that would not necessarily be obvious to potential dataset users; otherwise optional.
2.4	Anticipated Use	Information describing ways in which the dataset is expected to be used in support of the intended tasks	Optional	May include prior applications or deployments of the dataset.
2.5	Out-of-Scope Use	Information describing notable uses for which, or combinations in which, the dataset is known to be inapplicable or inadvisable	Optional	
3	Usage Rights and Restrictions	Information about terms, conditions, and rights governing the use, sharing, and modification of the dataset	Conditionally required	Required under conditions described in 3.1; otherwise recommended.
3.1	License	Text of or link to license terms governing use of the data, including any limitations on how the data can be used, modified, or shared	Conditionally required	If the dataset can be accessed by entities outside the provider organization under specific conditions or terms, this field is

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
				required, and it shall note any prohibited uses or data combinations, if applicable. This helps shield the provider organization and future users from both liability and potential misuse. In any other circumstances, recommended.
3.2	Payments	Information describing payments required or requested from those using, modifying, or sharing the dataset	Conditionally recommended	Recommended if payments are required to use, modify, or share the dataset, in which case information should be given on how to provide payment; otherwise optional.
3.3	Rightsholders	Information referencing parties who have ownership or other interests in the dataset	Conditionally recommended	Recommended if there are parties who have ownership or other interests in the dataset; otherwise optional.
4	Composition and Provenance	Information characterizing what the dataset contains and how it was built	Required	Should carefully trade off informativeness vs. judiciousness, as discussed in Clause 4.4.1.
4.1	Data Characteristics	Information describing the dataset's form and structure	Recommended	
4.1.1	Dataset Size	Information describing the number of instances included in the dataset (e.g., in tokens, bytes, or images with classification labels)	Optional	
4.1.2	Data Modalities	Information describing the type or types of data included (e.g., text, audio, image, video)	Recommended	
4.1.3	Dataset Splits	Information describing dataset partitioning and rationale for the splits	Optional	

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
4.1.4	Attributes	Information describing features or fields present in each data instance	Optional	
4.1.5	Data Structure	Information describing the underlying structure of the data (e.g., formats, directories, files)	Optional	
4.1.6	Data Types and Formats	Information describing encoding and representation of data instances, including any attribute values	Optional	
4.2	Sensitive Data	Information describing what personally identifiable information (PII) or confidential information, if any, the dataset is known to include	Conditionally required	Required if the dataset is known to contain PII; otherwise optional. May describe efforts to protect confidentiality (e.g., use of privacy-enhancing technologies). May reference, or be referenced by, other fields (e.g., 4.6 and 4.8.3) that are relevant to privacy and other sensitive data.
4.3	Data Collection	Information describing methods and circumstances of data acquisition or generation	Recommended	
4.3.1	Data Sources	Information describing from where the data was obtained	Recommended	If specific sources are not known or infeasible to list (e.g., because there are too many), should still give general information about what kinds of sources were used and how a source would have become included (e.g., via a web crawl). May include whether source data are publicly accessible, and if so, on what websites or other internet properties.
4.3.2	Data Acquisition Methods	Information describing the methods used to obtain the data (e.g., licensing agreement, publicly accessible download, web scraping)	Optional	

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
4.3.3	Data Collection Time Span	The temporal range during which the data instances in the dataset were acquired, produced, recorded, or observed	Optional	May be indicated as approximate or unknown. May be separated into multiple time spans for different stages of the data collection process (e.g., acquiring source data vs. enrichment).
4.4	Time Span Described	The temporal range described by the dataset, if applicable	Conditionally required	Required if the data describes events that occurred during a specific time span. May be indicated as approximate or unknown.
4.5	Sampling Procedure	Information describing the methods used to select data instances	Optional	
4.6	Consent Procedures	Information describing the process for obtaining informed consent for any collection of human-generated data	Conditionally recommended	Recommended if data was collected using procedures that would require informed consent; otherwise optional.
4.7	Data Flows	Information describing the chain of custody, i.e., how the data passed through different entities and processes (e.g., source, collection, processing, storage, and the data outflow)	Optional	May indicate geographic areas where data was or is processed or stored.
4.8	Data Enhancement	Information describing the data augmentations to improve or enrich coverage for data (e.g., to stretch, skew, add features to, or impute missing values in the data)	Optional	
4.8.1	Annotation	Information describing annotation processes, including annotation guidelines or any enrichment performed on the data	Recommended	If applicable, may simply indicate that there was no annotation process to enrich the source data.
4.8.1.1	Annotators	Information describing the role of annotators and methodology regarding inter-annotator agreement	Recommended	If applicable, may simply indicate that there was no annotation process to enrich the source data.

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
4.8.2	Synthetic Generation of Data	Information describing the algorithms and software used to generate any synthetic data, including how the synthetic data was incorporated into the dataset and what portion of the dataset consists of synthetic content vs. real-world data	Conditionally recommended	Recommended if there is synthetic data in the dataset; otherwise optional.
4.8.3	Data Preprocessing	Information describing transformations that the source data underwent before inclusion in the dataset (e.g., decisions on data curation, data cleaning, filtering of undesired or prohibited content, formatting, normalization of null values or outliers)	Conditionally recommended	Recommended if pre-processing was done on the source data; otherwise optional.
4.9	Data Quality Control	Information describing methods to ensure consistency and reliability, including to detect and handle errors, mislabeled instances, and noise	Optional	
5	Evaluation	Evidence regarding the dataset's quality and fitness for use	Conditionally required	See 5.1 and 5.3.
5.1	Data Analysis	Information describing analyses performed to gain insight into the nature and composition of the data (e.g., label consistency, data balance)	Conditionally required	See 5.1.1. Recommended even if 5.1.1 is not populated. May include information on data minimization assessment.
5.1.1	Data Validation	Information describing TEVV (testing, evaluation, validation and verification)—i.e., methods for and results from measuring or otherwise assessing the dataset's quality and suitability for its intended uses, including assessments of the adequacy of the data distribution	Conditionally required	Required if there is at least one defined task, application, or use case that the dataset is intended to support; otherwise optional.

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
5.2	Data Visualization	Visual summaries of dataset structure, key properties, and distribution	Optional	
5.3	Risk and Impact Analyses	Analyses of known vulnerabilities, risks, and impacts related to the dataset, including evidence supporting those analyses	Conditionally required	Required if negative impacts have previously been publicly documented and traced to this dataset or similar datasets. If this field is populated, it should include analysis of risks from possible inclusion of illegal material (e.g., child sexual abuse material, or CSAM), as applicable. If multiple risk analyses are provided (e.g., for distinct risks), they may be split into separate subfields. If populated, should characterize known residual risks. May include or reference results of empirical evaluations or measures taken to treat identified risks.
6	Maintenance and Monitoring	Information about processes and responsibilities for dataset oversight and curation after release	Conditionally recommended	See 6.4, 6.5, and 6.7.
6.1	Distribution	Information describing how the dataset is shared and published by the provider organization	Optional	
6.2	Maintainer	Point of contact or organization responsible for dataset maintenance	Optional	May include prior holders of this role.
6.3	Data Storage	Information describing how and where the dataset is stored	Optional	
6.4	Incident Reports	Incident reports involving the dataset, potentially including resolutions of such incidents	Conditionally recommended	Recommended if the dataset was in use before publication of this version of the documentation artifact; otherwise optional.

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
6.5	Versioning Policy	Information describing how changes are tracked and documented, including criteria for or frequency of dataset updates	Conditionally recommended	Recommended if the dataset has undergone or is expected to undergo periodic updates; otherwise optional.
6.6	Change Log	List of changes made to the dataset after initial release	Optional	May include why changes were made (e.g., new data available or undesired data identified in prior release); what evidence, feedback, or analysis (e.g., of trade-offs) informed the updates; or what impacts the updates are expected to bring.
6.7	Deprecation Policy	Information describing the provider organization's conditions and procedures for ceasing to support or circulate the dataset	Conditionally recommended	Recommended if the provider organization plans ongoing efforts to support the dataset; otherwise optional.
7	Governance	Information characterizing organizational practices applied over the dataset lifecycle	Recommended	
7.1	Framework Adherence or Compliance	Information describing policies, compliance and governance frameworks, and codes of conduct that have governed the provider organization's provision of the dataset; may include documentation of adherence	Optional	
7.1.1	Review Processes	Information describing efforts to ensure consistent application of policies, frameworks, practices, and codes of conduct, particularly in the course of organizational decision-making	Optional	
7.2	Consent and Data Rights Obligations	Information describing the process for obtaining informed consent for any collection of human-generated data	Optional	

Table A.2.1: Default AI dataset profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
7.3	Documentation Practices	Information describing the documentation artifact, and if applicable, the organizational practices that yielded the artifact	Recommended	May include how standard operating procedures are documented and how the documentation artifact is versioned, if applicable.
7.3.1	Documentation Quality Control	Information describing efforts to ensure that the documentation artifact accurately reflects the state of the dataset	Recommended	May include a record of when and by whom the accuracy of the documentation was reviewed.
7.4	Funding	Information disclosing sources of financial support for the development and maintenance of the dataset	Optional	
7.5	Terminology	Definitions or explanations of terms of art or other technical vocabulary used in the documentation	Optional	

728
729
730
731
732
733
734
735
736
737
738
739
740

A.3 Default AI Model Profile

Table A.3.1 provides a list of fields from Clause 5.3 along with subfields and designations. Where further explanation is required to clarify a designation or description, it is provided in the “Additional Guidance and Requirements” column.

Fields 5.1-(*) and 6.1-(*) in Table A.3.1 are meant to refer to datasets. To maintain consistency across documentation artifacts, these fields shall document the datasets using the default dataset documentation profile given in A.2 (though requirements from that profile may be relaxed if the information is not available to or cannot be revealed by the model provider). Where feasible, dataset documentation artifacts should be referenced rather than included in full. If the model developers made changes to a dataset for training or evaluation purposes, the corresponding field should still reference the original dataset’s documentation artifact, but it should provide any overrides or addenda to that artifact only for those fields that differ from those in the base dataset’s documentation.

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
1	Identifying Descriptors	Information that uniquely identifies the model or characterizes its conditions of creation	Required	Shall provide enough information to determine whether a given model is the one described by the documentation artifact.
1.1	Identifier	One or more canonical names of the model or model series, including the model family, if applicable	Recommended	
1.2	Description	Brief human-readable summary of the model's characteristics, which may include its size (e.g., in parameters) and input and output modalities	Required	
1.3	Authorship information	Information about who created the model and documentation	Recommended	
1.3.1	Developer	One or more names of individuals or entities who created or developed the model	Recommended	
1.3.2	Documentation Authors	One or more names of entities who prepared the documentation artifact	Optional	
1.5	Point of Contact	Point of contact for feedback or questions related to the model	Conditionally recommended	Recommended if the model can be accessed by entities outside the provider organization; otherwise optional.
1.6	Version Information	Information uniquely associated with a specific configuration or iteration of the model	Conditionally required	See 1.7.1 and 1.7.3.

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
1.6.1	Version Identifier	Unique identifier associated with the release of the model being documented, with identifiers structured to enable easy determination of whether a version is current	Conditionally recommended	Recommended if the model can be accessed by entities outside the provider organization; otherwise optional.
1.6.2	Cryptographic Signature	Cryptographic signature or signatures of the serialized model	Optional	
1.6.3	Release Date	Date on which the documented version of the model was first made available in its complete form	Conditionally recommended	Recommended if the model can be accessed by entities outside the provider organization; otherwise optional.
1.6.4	Decommissioning Date	Date by which the provider organization ceased, or plans to cease, supporting or circulating the model	Conditionally required	Required if the provider organization has planned or acted upon such a date; otherwise optional.
1.7	Release Methods	Information describing the methods by which the model can be accessed or obtained (e.g., cloud service providers, product surfaces, websites where weights can be downloaded)	Conditionally recommended	See 1.8.1.
1.7.1	Source URL	URL from which the model can be obtained (or which provides information on how to obtain it), if applicable	Conditionally recommended	Recommended if the model can be obtained via the internet; otherwise optional.
1.9	Citation	Technical citation format for referencing the model	Conditionally recommended	Recommended if the model can be accessed by entities outside the provider organization; otherwise optional.
1.10	Attributions	Notices or attributions that are to be presented in association with the model, if applicable	Conditionally required	Required if Use Rights and Restrictions specifies a mandate to attribute; otherwise optional.

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
1.11	Lineage	Information describing prior models from which the model was derived (e.g., by fine-tuning or distillation) and what mechanisms were used	Optional	
1.12	Documentation Version Identifier	Required	Formal identifier for the iteration of the present documentation artifact	Even for a single model version, different versions of the documentation artifact sometimes need to be published (e.g., to include newly available information or correct errata). To address such situations, this field shall provide an identifier uniquely associated with the present version of the artifact.
2	Intended Use	Information characterizing the applications for which the model was designed	Recommended	
2.1	Intended Task	Information describing one or more applications or use cases that the dataset is intended to support (e.g., image segmentation, translation, summarization)	Conditionally recommended	Recommended if there are one or more defined tasks for which the model is intended to be used; otherwise optional.
2.2	Applicable Domains	Sectors or domains in which the model is designed to be applied	Optional	
2.3	Intended Stakeholders	Description of the primary individuals, groups, or organizations that can benefit from or be impacted by the model or its outputs	Optional	
2.4	Deploying Entities	Information identifying individuals, groups, or organizations deploying the model	Optional	

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
2.5	Limitations	Information about technical constraints on the use of the model or its performance or reliability (e.g., dataset limitations or known failure modes)	Conditionally recommended	Recommended if the model provider knows of uses someone would likely attempt to use the model for which it is technically ill suited; otherwise optional.
2.6	Anticipated Use	Information describing ways in which the model is expected to be used in support of the intended tasks	Optional	May include prior applications or deployments of the model.
2.7	Out-of-Scope Use	Information describing notable uses for which the model is known to be inapplicable or inadvisable	Optional	
3	Usage Rights and Restrictions	Information about terms, conditions, and rights governing the use, sharing, and modification of the model	Conditionally required	Required under conditions described in 3.1; otherwise recommended.
3.1	License	Text of or link to license terms governing use of the model, including any limitations on how the model may be used, modified, or shared; may describe the degree of openness (e.g., proprietary or open access) of various components of the model (e.g., weights, code, and data)	Conditionally required	Required if the model can be accessed by entities outside the model provider under specific conditions or terms; otherwise recommended.
3.2	Payments	Information describing whether payment is required to use, modify, or share the model, and if so, how and to whom to make payment	Conditionally recommended	Recommended if payments are required to use, modify, or share the model, in which case information should be given on how to provide payment; otherwise optional.
3.3	Rightsholders	Information referencing parties who have ownership or other interests in the model	Conditionally recommended	Recommended if there are parties who have ownership or other interests in the dataset; otherwise optional.

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
4	Design	Information describing the model's architecture and technical assumptions	Recommended	
4.1	Algorithm and Architecture	Information describing the underlying computational paradigm (e.g., supervised learning, reinforcement learning), the structure of the model, and how it was implemented	Optional	May clarify training data requirements implied by the model design.
4.2	Performance Metrics Targeted	Information about the performance metrics (e.g., precision and recall) that system design targeted, and any ensuing design choices	Optional	
4.3	Explainability and Interpretability	Information describing any efforts to ensure that the model produces its outputs in a way that can be understood, interpreted, and counterfactually reasoned about by humans	Optional	
5	Training	Information about protocols and data used in optimizing the parameters of a model	Recommended	
5.1	Training Protocol	Information describing the procedures and configurations used to train the model, including training objectives and hyperparameters	Optional	Where relevant, may contain information on training procedures that follow the main training algorithm ("post-training"), such as efforts to train a large language model to refuse harmful requests.
5.2- (*)	Training Dataset Reference	Dataset documentation artifacts describing the datasets used in training	Recommended	One entry per referenced dataset, with indices substituting for '(*)'. If included, a separate dataset entry should be provided for each distinct use of a given dataset (e.g., if the same dataset was used in pre-training and fine-tuning). Where the reference

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
				dataset cannot be publicly disclosed, the documentation artifact should still describe characteristics that readers likely need to understand, to the extent consistent with the dataset provider's objectives.
5.2- (*).1	Dataset Point of Use in Training	Information describing the part of the training pipeline for which the referenced dataset was used	Conditionally recommended	Recommended if model training involved multiple stages; optional otherwise. E.g., training stages for a large language model often include pre-training, fine-tuning, preference tuning, reinforcement learning from human feedback, and others.
5.2- (*).2	Training Data Role	Information describing the purpose of the referenced dataset as used in the training data relative to the intended use of the model	Optional	
5.2- (*).3	Training Data Limitations	Information describing any quality issues in the referenced dataset that impact model generalizability or performance	Optional	
6	Evaluation	Evidence regarding the model's quality and fitness for use	Recommended	Should communicate the objectives, findings, and limitations of TEVV activities and metrics to enable suitability assessment.
6.1	Evaluation methodology	Information about the approaches and measurement methods used for evaluating the model	Recommended	Should clarify which evaluation methods are fully automated, which are performed by humans, and which involve both.
6.1.1- (*)	Evaluation Dataset Reference	Dataset documentation artifacts describing the datasets used in evaluation	Recommended	One entry per referenced dataset, with indices substituting for '(*)'. If included, a separate dataset entry should be provided for each distinct use of a given dataset (e.g., if the same dataset was used for benchmarking and as part of an AI red teaming

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
				exercise). Where the reference dataset cannot be publicly disclosed, the documentation artifact should still explicitly describe characteristics readers will likely need to understand, to the extent consistent with the model provider's objectives.
6.1.1-(*).1	Dataset Point of Use in Evaluation	Information describing the part of the evaluation pipeline for which the referenced dataset was used	Conditionally recommended	Recommended if multiple types or stages of evaluation were performed; otherwise optional. E.g., evaluation of a large language model often includes benchmark evaluation, AI red teaming, and potentially other methods.
6.1.1-(*).2	Evaluation Data Role	Information describing the purpose of evaluation data relative to intended use (e.g., generalization assessment and performance consistency across a variety of use cases)	Optional	May include splits or percentages of data for test and validation, if applicable
6.2	Third-Party Evaluation	Information describing evaluation activities conducted by parties external to the model developer and the conditions under which evaluation was performed	Optional	May include information about the third-party identity and organizational affiliation, the nature of evaluation (e.g., conformity assessment, AI red teaming), the scope of evaluation (e.g., risks, use cases, capabilities assessed), and access conditions (e.g., black-box, white-box, API only). May reference other subfields of field 6 for technical details, or vice versa.
6.3	Quantitative and Qualitative Analyses	Information describing model performance, including on aggregate and disaggregated performance metrics	Optional	May include results of third-party testing results or red team findings (e.g., a link to an evaluation report).

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
6.4	Performance Limitations	Information about evaluation results that reveal limits to the model's performance and any relevant disclosures about timing and frequency of evaluations	Conditionally recommended	Recommended if evaluation procedures have revealed constraints, gaps, or other limitations that would not necessarily be obvious to potential model users; otherwise optional.
6.5- (*)	Evaluation Metrics and Results	Information about TEVV results, including methods and metrics for measuring or otherwise assessing the model's effectiveness for the intended use, and including the model's performance against those metrics, particularly on established benchmarks	Recommended	If applicable, may be disaggregated to have one entry per discrete task or application for which the model was evaluated, with indices substituting for '(*)'.
6.5- (*).1	Intended Use Evaluated	Indication of what intended task or application this group of evaluation metrics and results assessed fitness for	Conditionally required if parent is populated	Required if evaluations are documented separately for more than one task or application; otherwise optional.
6.5- (*).2	Robustness Metrics and Results	Information about how the model was evaluated to demonstrate consistent performance under distributional shifts and adversarial conditions, and results of such evaluation	Optional	
6.5- (*).3	Evaluation Reliability Metrics and Results	Information about the stability and reproducibility of evaluation results	Optional	May include any metrics that reflect uncertainty estimates, including confidence intervals. This information enables interested parties to determine whether performance claims are stable enough to support deployment decisions.
6.6	Risk and Impact Analyses	Analyses of known vulnerabilities, risks, and impacts related to the model, including evidence supporting those analyses	Conditionally recommended	Recommended if negative impacts have previously been publicly documented and traced to this model or similar models. If this field is populated for a generative AI model, it should include analysis of

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
				risks from possible generation of illegal material (e.g., child sexual abuse material, or CSAM), as applicable. If multiple risk analyses are provided (e.g., for distinct risks), they may be split into separate subfields. Risk and impact analyses may refer to other subfields for risk-relevant evaluation results or risk treatment measures taken, or those fields may refer to this field. If populated, should characterize known residual risks. May include or reference results of empirical evaluations or measures taken to treat identified risks.
7	Maintenance and Monitoring	Information about the processes and responsibilities for monitoring and maintaining the model after release	Recommended	
7.1	Monitoring Mechanisms	Information describing mechanisms that the model owner uses to monitor model performance (e.g., analyzing logs, classifiers, feedback), including methods for detecting drift and otherwise re-evaluating	Recommended	
7.2	Update Mechanisms	Information describing how models are retrained, including in light of monitoring results	Optional	May include plans for how, how often, and by what channels updates will be communicated to stakeholders.
7.3	Change Log	List of changes made to the model after deployment	Optional	May include why changes were made (e.g., incidents or performance degradation); what evidence, feedback, or analysis (e.g., of trade-offs) informed the updates; or what impacts the updates are expected to bring.

Table A.3.1: Default AI model profile

ID	Field Name	Description	Designation	Additional Guidance and Requirements
7.4	Maintainer	Point of contact or organization responsible for dataset maintenance	Optional	May include prior holders of this role.
7.5	Post-Deployment Reports	Information about the model's performance after it is deployed, including major use cases, quantitative analyses, and user feedback	Optional	
7.6	Incident Reports	Incident reports involving the model, potentially including resolutions of such incidents	Conditionally recommended	Recommended if the model was in use before publication of this version of the documentation artifact; otherwise optional.
7.7	Versioning Policy	Information describing how changes are tracked and documented, including criteria for or frequency of model updates	Recommended	
7.8	Decommissioning Procedures	Information describing the provider organization's conditions and procedures for retiring the model from deployment in a manner that minimizes risk	Conditionally recommended	Recommended if the model provider plans to deploy and operate the model; otherwise optional. May include data deletion procedures, stakeholder notification protocols, and transition plans for dependent systems.
7.9	Deprecation Policy	Information describing the model provider's conditions and procedures for ceasing to support or circulate the model	Conditionally recommended	Recommended if the model provider plans ongoing efforts to support the dataset; otherwise optional.
8	Governance	Information characterizing organizational practices applied over the dataset lifecycle	Recommended	
8.1	Framework Adherence or Compliance	Information describing policies, compliance and governance frameworks, incident management practices, and codes	Optional	May include documentation of adherence (e.g., certifications).

Table A.3.1: Default AI model profile				
ID	Field Name	Description	Designation	Additional Guidance and Requirements
		of conduct that have governed the model provider's provision of the model		
8.1.1	Review Processes	Information describing efforts to ensure consistent application of policies, frameworks, practices, and codes of conduct, particularly in the course of organizational decision-making	Optional	
8.2	Documentation Practices	Information describing the documentation artifact, and if applicable, the organizational practices that yielded the model documentation	Recommended	May include how standard operating procedures are documented and how the documentation artifact is versioned, if applicable.
8.2.1	Documentation Quality Control	Information describing efforts to ensure that the documentation artifact accurately reflects the state of the model	Recommended	May include a record of when and by whom the accuracy of the documentation was reviewed.
8.3	Stakeholder Consultation	Information describing consultation activities conducted with interested parties impacted by the model's development and deployment	Optional	May describe any methods used to consult with relevant interested parties to account for their perspectives and feedback, findings from consultations, and subsequent actions taken.
8.4	Funding	Information disclosing sources of financial support for the development and maintenance of the model	Optional	
8.5	Terminology	Definitions or explanations of terms of art or other technical vocabulary used in the documentation	Optional	

B Example of a Machine-Readable Documentation Artifact

(Under consideration for the final draft.)

744 **C Relationships with Other ISO/IEC Standards on AI**

745 (Under consideration for the final draft.)

746 **Bibliography**

747 (To be completed with the final draft.)