Documenting High-Risk AI: A European Regulatory Perspective

Isabelle Hupont, Marina Micheli, Blagoj Delipetrev, Emilia Gómez, and Josep Soler Garrido, European Commission

This article discusses transparency obligations introduced in the Artificial Intelligence Act, the recently proposed European regulatory framework for artificial intelligence (AI). An analysis of the extent to which current approaches for AI documentation satisfy requirements is presented and their suitability as a basis for future technical standards is assessed.

The field of artificial intelligence (AI) has experienced an unprecedented rate of development in the last decade, finding a wide range of practical applications across most, if not all, economic sectors. Driven by the digital transformation of society, AI is bound to bring opportunities, not only as an engine for growth and innovation, but also as a means to address some of the most pressing societal challenges in critical domains, for example, the environment, energy, agriculture or health.

In the European Union, this is reflected in the prominent position of AI and data—the raw material needed for its development—in the policy agenda. A wide range of legislative initiatives and funding programs have been put in place in order to unleash the potential of data-driven innovation in Europe, boosting research and industrial capacity, while ensuring that AI is a force for good in society.

Indeed, the most prominent element of the European approach to AI is arguably its human-centered view. Like is the case with many technologies, adoption of AI comes with certain risks, most notably its potential, if not properly used, to negatively affect the fundamental rights,

Digital Object Identifier 10.1109/MC.2023.3235712
Date of current version: 3 May 2023

safety, and health of human beings. Against this backdrop, and in order to set the necessary regulatory conditions for the adoption of trustworthy AI in the European Union, the European Commission presented in April 2021 its proposal for the regulation of artificial intelligence, the AI Act.[1] The AI Act lays down a set of legal obligations for providers of certain AI systems, defining requirements that depend on their risk profile. The legal text does not, however, mandate any specific technical solutions or approaches to be adopted. Instead, it defines the essential, high-level requirements needed for the protection of public interests. In turn, technical solutions for their fulfillment in practice will be specified primarily in the form of technical standards.

European and international standardization organizations developing the standards to support the AI Act are faced with the task of capturing in their specifications the existing landscape of best practices and state-of-the-art techniques and methods in trustworthy AI. In this article, we contribute to these ongoing efforts, and focus our attention on a subset of the requirements defined in the AI Act, namely those related to transparency, understood in our context as the transfer of relevant information about the AI system—or the data used to build it—to relevant stakeholders. Specifically, we perform an in-depth analysis of existing AI documentation approaches which have emerged in recent years, assessing their potential to operationalize two concrete requirements: 1) the provision of information to users and 2) the provision of technical documentation to national authorities and their designated conformity assessment bodies.

AI TRANSPARENCY INITIATIVES: FROM VOLUNTARY PRACTICES TO LEGAL REQUIREMENTS

Transparency of AI systems has been an area of great interest in industry and academic circles in recent years, starting even before the appearance of AI regulatory frameworks. The rapid pace of development of AI methods and their swift transition from research ideas to operational environments, while fundamentally being a technological success story, have occasionally been accompanied by high-profile failures and resulting controversy. Well-known examples include dark-skinned individuals being wrongfully arrested based on face recognition software[2] and self-driving cars failing to stop for a pedestrian in a crosswalk.[3] In addition to their potentially serious direct consequences, these incidents tend to attract considerable attention, both from the AI community and the general public, affecting the reputation of the developers involved and generally undermining trust in AI.

Transparency plays a fundamental role in mitigating these risks. Indeed, many notable AI incidents could have been avoided had the capabilities and limitations of AI systems been properly communicated in the first place to the respective AI practitioners, users or those ultimately affected by their decisions. Inspired in part by this realization, several research and industry-driven initiatives have emerged in recent years with the aim to define documentation approaches that increase transparency and trust in AI. Among the most successful initiatives we find some that focus on the datasets used for AI, such as *Datasheets for Datasets*[4] and *The Dataset Nutrition Label*,[5,6] as well as some that address the documentation of AI models and algorithms, such as *Model Cards*[7] and *AI Factsheets*.[8] Some of these, despite their short life, and while not being formal standards and having a voluntary nature, have already seen a relevant degree of adoption by the AI community.

At this moment, however, a transition from voluntary practices to hard legal requirements is underway with the adoption of AI regulatory frameworks appearing on the horizon in many parts of the world. In the EU, the AI Act is, at the time of writing this article, being discussed by the colegislators, that is, the European Parliament and the Council of the European Union. The ongoing negotiations are expected to lead to its adoption in the near future, which will result in concrete transparency obligations for the providers of certain AI systems.

DOCUMENTATION REQUIREMENTS FOR HIGH-RISK AI SYSTEMS

The obligations defined in the European AI Act are closely linked to defined risk profiles for AI systems, which can be, from highest to lowest: *unacceptable risk*, covering harmful uses of AI or uses that contravene European Union values and that are consequently prohibited; *high risk*, covering AI systems that may create an adverse impact on people's safety, health, or fundamental rights under certain circumstances; *limited risk*, applying to some AI systems that are not considered high risk but whose operation shall be informed to the natural persons exposed to them (for example, chatbots or deepfake videos); and *minimal risk*, covering all other AI systems that can be deployed in the EU without additional legal obligations beyond those already in place.

Reasonably, and as depicted in Figure 1, the strongest documentation obligations apply to high-risk AI systems. This is the case for both stakeholder categories considered, which nevertheless have different transparency needs, ranging from clear but concise and accessible instructions of use for users, to comprehensive technical documentation—demonstrating full compliance with the legal requirements—for authorities and conformity assessment bodies. In this context, the question arises whether existing documentation practices—and their wealth of publicly available examples—could be leveraged when formalizing these needs at the technical level in the form of standards.

To assess this, we have identified, compiled, and classified the main information elements relevant to these stakeholders based on the legal text. These are summarized in Table 1, which also includes for reference the relevant articles where the corresponding information elements are covered.

The information required encompasses not only the AI systems themselves, but also the datasets used to build them. Indeed, the data used for training, validating, and testing these systems is known to have a substantial impact on their operation, and, if not properly managed, can be a major source of failures and incidents. Accordingly, the data-related elements listed should not only be documented but also be part of a suitable data governance practice. It is also important to note that many information elements are common to users, authorities, and conformity assessment bodies, with the main difference being the general level of detail required (for example, in *dataset scope*, *data representativeness* or *purpose*), or the need for some additional considerations to be captured for conformity assessment, such as a description at the technical level of processing steps, implemented features, and assessment details (for example, in *data preparation*, *risks*, or *human oversight*).

It should be noted that Table 1 is not meant to be a final and exhaustive list of information elements needed for compliance with future legal requirements. First and foremost, because the AI regulation is still under negotiation, and is therefore subject to be modified in its road toward adoption. Furthermore, documentation needs arising from the requirements laid down by the AI Act will most certainly include elements not directly connected to technical design and development characteristics, the main focus of this work. Documentation elements that are pertinent for the AI Act but not explicitly considered in our context include, among others, certain product-related aspects (for example, product layouts and illustrations, installation and maintenance instructions), documentation of process-oriented requirements (for example, procedures, roles, and responsibilities in a quality, risk, or data management system), and documentation of postmarket placement measures (for example, monitoring and incident reporting measures).

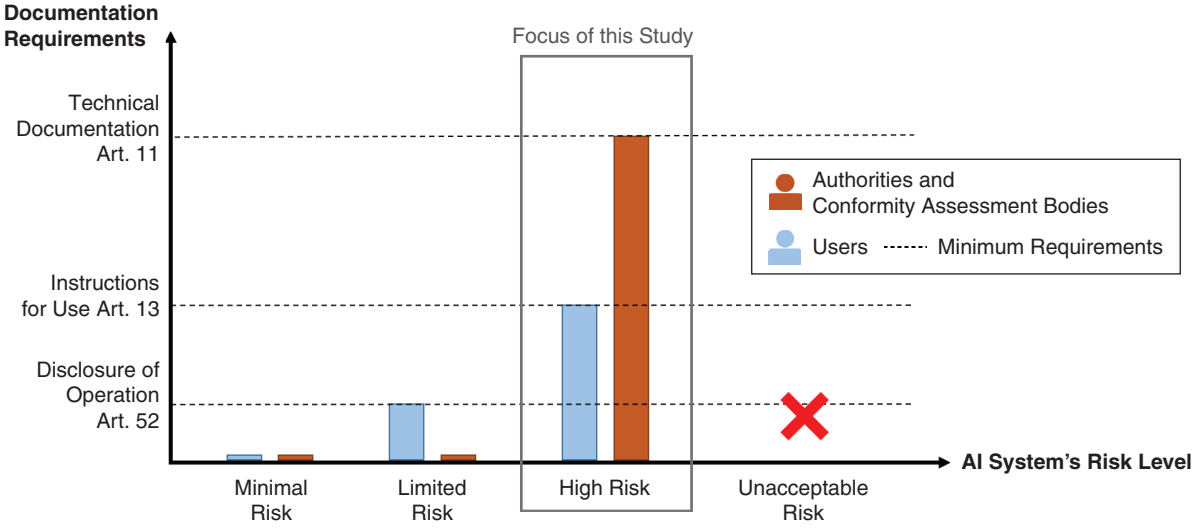

FIGURE 1. Minimum documentation requirements in the European AI Act, depending on the AI system's risk level (minimal risk, limited risk, or high risk) and the intended recipient (users or authorities and conformity assessment bodies). In this article, we cover the operationalization of documentation obligations for high-risk AI systems and both stakeholders.

Despite this, Table 1 provides a detailed overview of the type of technical information elements related to datasets and AI systems that providers of high-risk AI systems are expected to systematically document once the AI Act comes into force and serves as a solid basis for our assessment.

ASSESSMENT OF STATE-OF-THE-ART AI DOCUMENTATION APPROACHES

Selection criteria

For our analysis, we selected a relevant subset from the range of existing AI and dataset transparency and documentation approaches, with the selection criteria following the specific needs of the European AI regulation proposal.

The first key consideration was related to the horizontal nature of the AI Act, meaning that it defines requirements for high-risk AI systems independent of the specific sector in which they operate. Similarly, it covers a wide range of AI methods (for example, types of algorithms), not being fundamentally limited to specific techniques such as machine learning. Considering this, we included documentation approaches that can be broadly applied to most AI systems and techniques.

Another important consideration we made for the selection of documentation methodologies was their alignment with the risk-based approach defined by the European AI regulation. In this regard, methodologies explicitly considering AI-related risks and related attributes (such as robustness, fairness and bias, discrimination, and other potential sources of harm to fundamental rights, safety, or health) were favored.

TABLE 1. Summarized list of relevant technical information elements under the European AI Act, distinguishing between documentation requirements related to AI datasets and systems.

	Information element	Target	Description	Articles
DATASETS	Data provenance	A	Specify the origin and/or source(s) from which the data have been collected (for example Internet, private database, third party public dataset).	10, 11
	Dataset scope	U, A	Assess the type of data contained in the dataset (for example, numerical, categorical, text, image) and its scope (for example, sales data, personal data, medical data of a specific target population). Assess possible data gaps and/or shortcomings (for example, sufficiency of data, appropriate level of granularity).	10, 11, 13
	Data collection	A	Indicate how data was collected from its origin/source(s) (for example, by means of web crawling techniques, by querying a private database). If data was collected from different sources, indicate the aggregation techniques used.	10, 11
	Data preparation	U	Provide information about data format(s). Give proper consideration to the split into training, testing and validation sets, and what their individual characteristics are.	10, 13
		A	Additionally, document data preparation and processing operations, including: annotation, labeling, cleaning, and enrichment.	10, 11
	Data correctness	A	Include a detailed analysis of dataset quality, with concrete metrics and measures, to ensure that data are correct and complete (for example, images have enough resolution, check missing values, outliers, noise level in terms of incorrect labels).	10, 11
	Data representativeness	U, A	Provide evidence, with concrete measures and metrics, on the relevance and representativeness of data with regard to the persons or groups of persons and the geographical, behavioral, or functional setting on which the system is intended to be used.	10, 11, 13
	Data privacy	A	Describe the data safeguards for the fundamental rights and freedoms of natural persons in the form of privacy measures (for example, data pseudonymization, encryption, anonymization, aggregation).	10, 11, 13

(*Continued*)

TABLE 1. (*Continued.*) Summarized list of relevant technical information elements under the European AI Act, distinguishing between documentation requirements related to AI datasets and systems.

	Information element	Target	Description	Articles
AI SYSTEMS	Purpose	U, A	Document the intended purpose of the AI system and any potential misuse of the system that could be reasonably foreseen.	8, 9, 11, 13
	Risks	U	Describe situations where the specific AI system may lead to risks, including those related to health, safety, and fundamental rights.	9, 13
		A	In addition, describe design features and mitigations adopted to address risks, as well as the description of any residual risks.	9, 11
	Interpretation	U	Include sufficient information in order to enable the user to interpret the outputs of the system and to use them appropriately.	13
		A	Additionally, include details about the assessment of measures implemented to facilitate interpretation of the system outputs.	11, 13
	Human oversight	U	Detail the human oversight measures in place to understand the operation and internal decision-making of the system, identify anomalies, monitor and control its operation, and prevent over-reliance.	11, 14
		A	In addition, provide detailed information about the assessment of human oversight measures in place.	11, 14
	Architecture	A	Description of the AI architecture, including the type of algorithm/model, the processing steps, software components involved, and how they integrate, and computational resources used across different steps, for example, training, testing, validation, and operation.	11
	Development	A	Detailed description of the methods and processes used for AI system development, including the components utilized, such as pretrained models or tools.	11
	Training	A	Describe how the AI system was trained, including an explanation of optimization targets and objective functions with the relevant parameters and tradeoffs involved, the training techniques employed, and any relevant assumption or choices.	11
	Accuracy	U, A	Report the level of accuracy achieved and which can be expected, including on the specific persons or groups targeted by the system.	13, 15
	Robustness	U	Include a description of situations and circumstances under which the performance of the system may be impacted or impaired, and any potential source of errors, faults, or inconsistencies.	13
		A	Furthermore, describe the specific robustness and resilience measures adopted, whether by design in the algorithm or system, or through fail-safe or redundancy.	11, 15
	Cybersecurity	U, A	Document measures adopted to ensure the cybersecurity of the AI model and/or system against specific threats such as adversarial attacks or data poisoning, including evaluation results and metrics.	13, 15
	Test	A	Detailed testing and verification protocol and test logs showing levels of accuracy, robustness, and cybersecurity achieved, including metrics and thresholds defined according to the intended purpose of the system and the identified risks.	11, 15
	Changes	U	Describe any changes that the system may be subjected to after it is placed in the market, for example, if it continues to learn during operation.	13
		A	In addition, describe measures in place to maintain performance as those changes take effect, with due consideration to the potential effects of any feedback loops.	11, 15
	System aspects	A	Consider system-level aspects beyond the AI component itself, including the integration in the overall system, firmware and hardware components, and interactions with non-AI subsystems.	11

Target audience is denoted with A=*authorities & conformity assessment bodies* and U=*users*.

Last but not least, we decided to favor open and well-structured approaches (for example, in the form of research papers) which have already attracted attention and been subject to the scrutiny of the AI research and practitioner communities. Peer-reviewed and field-tested approaches are expected to be more mature, having potentially gone through multiple rounds of improvement. Furthermore, they are typically accompanied by multiple examples and in some cases even software tools that simplify and partially automate their adoption.

Given these considerations, we selected the six documentation approaches listed in the columns of Table 2. Three of them focus on datasets and another two on AI systems. The final one is the very recent *Organisation for Economic Co-operation and Development (OECD) Framework for the Classification of AI Systems*[9] which, while not being a research paper and not being widely adopted due to its novelty, was considered essential to our study given its international and open nature, the high-caliber institutions involved and the large number of contributing AI researchers and practitioners.

Assessment methodology

Five scientific experts on AI from the European Commission's Joint Research Center participated in an assessment exercise following a Delphi method.[10] Initially, the experts went through all selected approaches and, for each of

TABLE 2. Suitability of state-of-the-art documentation methodologies with respect to the technical information elements identified in the European AI Act.

		Target stakeholder	*Datasheets for datasets* Gebru et al.[4]	*Dataset nutrition label* Chmielinski et al.[6,5]	*Accountability for machine learning datasets* Hutchinson et al.[11]	*Model cards* Mitchell et al.[7]	*AI factSheets*[+] Arnold et al.[8,12]	*Framework for the classification of AI systems* OECD[9]
Publication year			**2018**	**2022**[*]	**2021**	**2019**	**2019**	**2022**
Proponent institution(s)			**Black in AI, Microsoft, and several universities (industry + academia)**	**Harvard and MIT (academia)**	**Google (industry)**	**Google (industry)**	**IBM (industry)**	**OECD (intergovernmental organization)**
Documentation Method			**Questionnaire**	**Visual**	**Template**	**Information sheet**	**Information sheet**	**Questionnaire**
DATASETS	Data provenance	A	●	●	●	◐	◐	◐
	Dataset scope	U, A	●	●	●	○	◐	◐
	Data collection	A	●	●	◐	x	○	◐
	Data preparation	U	●	●	◐	◐	◐	○
		A	●	●	◐	○	○	x
	Data correctness	A	◐	◐	◐	x	○	○
	Data representativeness	U, A	◐	○	◐	◐	◐	○
	Data privacy	A	◐	●	◐	x	○	◐

(*Continued*)

TABLE 2. (*Continued.*) Suitability of state-of-the-art documentation methodologies with respect to the technical information elements identified in the European AI Act.

		Target stakeholder	*Datasheets for datasets* Gebru et al.[4]	*Dataset nutrition label* Chmielinski et al.[6,5]	*Accountability for machine learning datasets* Hutchinson et al.[11]	*Model cards* Mitchell et al.[7]	*AI factSheets*[†] Arnold et al.[8,12]	*Framework for the classification of AI systems* OECD[9]
Publication year			**2018**	**2022**[*]	**2021**	**2019**	**2019**	**2022**
Proponent institution(s)			**Black in AI, Microsoft, and several universities (industry + academia)**	**Harvard and MIT (academia)**	**Google (industry)**	**Google (industry)**	**IBM (industry)**	**OECD (intergovernmental organization)**
Documentation Method			**Questionnaire**	**Visual**	**Template**	**Information sheet**	**Information sheet**	**Questionnaire**
AI SYSTEMS	Purpose	U, A	x	x	x	●	●	●
	Risks	U	◕	○	◕	●	●	●
		A	○	x	x	◕	◕	x
	Interpretation	U	x	x	x	○	◕	○
		A	x	x	x	x	◕	x
	Human oversight	U	x	x	x	x	◕	○
		A	x	x	x	x	◕	x
	Architecture	A	x	x	x	◕	○	◕
	Development	A	x	x	x	○	◕	○
	Training	A	x	x	x	○	○	○
	Accuracy	U, A	x	x	x	●	◕	x
	Robustness	U	x	x	x	○	●	○
		A	x	x	x	x	◕	x
	Cybersecurity	U, A	x	x	x	x	◕	x
	Test	A	x	x	○	◕	◕	○
	Changes	U	x	x	x	x	◕	◕
		A	x	x	x	x	◕	x
	System aspects	A	x	x	x	x	○	○

[*] The *dataset nutrition label* was first published in 2018 and then revisited in 2022. We assess here its latest version at the time of writing this article.
[†] AI FactSheets are based on templates that can be tailored to different stakeholders. Our assessment takes into account various templates made available by the authors.
Colored spheres indicate the degree of coverage for each requirement, namely: white ○ low coverage; yellow ◕ medium coverage; green ● high coverage. A missing sphere (x) indicates that the element is not considered.

them, annotated the level of coverage of each information element using a color scheme, where white represents low coverage, yellow medium coverage, green high coverage, and a missing sphere (X) indicates that the approach does not consider the requirement. Thereafter, two experts aggregated the responses and identified disagreements. Finally, a group session was carried out with all experts to discuss and resolve disagreements until an overall consensus was reached. A single Delphi round was sufficient to reach consensus.

DISCUSSION

Table 2 presents the assessment of selected documentation approaches resulting from the Delphi exercise described. We considered two dimensions in our analysis: *coverage*, indicated by the presence of a sphere on the table, and *depth*, indicated by its color. In both cases, our quantitative assessment should be taken as an approximate one. Indeed, while there was substantial a priori agreement between individual experts during the entire process, some disagreements had to be resolved, both related to data information elements (for example, *data preparation*, *data correctness*, *data representativeness*, and *data privacy*) as well as AI system-related ones (including *risks* and *test* aspects). Alignment rounds between experts included discussions aiming to define, for example, what would constitute practical and useful documentation levels regarding data distributions of target populations, risk mitigation measures, testing protocols, and data safeguards for privacy and fundamental rights. This might indicate that some information elements required are particularly subjective and require careful attention.

In terms of dataset-related information, our overall assessment is very positive. All three dataset-centered approaches, namely *datasheets for datasets*, *dataset nutrition label*, and *accountability for machine learning*, provide in-depth coverage of AI Act data-related documentation needs, with the first of them being the most complete overall. Still, some specific information elements could be captured in greater detail, namely those related to data correctness, representativeness, and privacy. This is especially the case when considering documentation needs for authorities and conformity assessment, which would benefit from concrete measures and metrics (for example, distributions of data across demographic groups, checks for missing values, outliers, and privacy measures for data), found only to be partially covered.

Regarding AI system documentation, the assessment is similarly positive, with all of the required information elements being covered, albeit with a somewhat lower depth. Notably, all of the documentation approaches for AI systems reviewed provide substantial coverage of the intended purpose of the system as well as associated risks and potential misuse scenarios. Overall, *AI Factsheets* is found to be the most complete initiative assessed, covering all relevant technical concerns through the entire AI system lifecycle and taking a system- and service-level view beyond individual AI model considerations. This approach is a template-based one, and a methodology is described to tailor it to the specific needs of concrete stakeholders. During our study, we jointly assessed several factsheet templates available in the literature, partially explaining the broad coverage observed. The other AI system documentation approaches assessed, while presenting some gaps, are very complementary in terms of content as well as presentation, notably by providing information in a user-friendly manner, for example, by visual means. In particular, documentation of performance in *model cards*, conveying disaggregated evaluation results, that is, reported separately for the different relevant groups, was highly rated and perceived as very effective.

Collectively, the documentation approaches reviewed appear not to fully achieve the level of technical detail needed by authorities and those responsible for assessing compliance with legal requirements, at least for a subset of the relevant information elements. Indeed, this may not be their objective, as current adopters often link to external supplementary material

INDEED, THIS MAY NOT BE THEIR OBJECTIVE, AS CURRENT ADOPTERS OFTEN LINK TO EXTERNAL SUPPLEMENTARY MATERIAL IN THE FORM OF RESEARCH PAPERS AND CODE REPOSITORIES.

ABOUT THE AUTHORS

ISABELLE HUPONT is a senior researcher at the Joint Research Center, European Commission, 41092 Seville, Spain. Her research interests include affective computing, facial analysis, and biometrics. Hupont received a Ph.D. in artificial intelligence from the University of Zaragoza, Spain. Contact her at isabelle.hupont-torres@ec.europa.eu.

MARINA MICHELI is a scientific project officer at the Joint Research Center, European Commission, 21027 Ispra (VA), Italy. Her research interests include data governance and digital inequality. Micheli received a Ph.D. in information society from the Department of Sociology and Social Research at the University Milan-Bicocca, Italy. Contact her at marina.micheli@ec.europa.eu.

BLAGOJ DELIPETREV is a scientific project officer at the Joint Research Center, European Commission, 21027 Ispra (VA), Italy. His research interests include artificial intelligence and data policies. Delipetrev received a Ph.D. in computer science from TU Delft, The Netherlands. Contact him at blagoj.delipetrev@ec.europa.eu.

EMILIA GÓMEZ is a principal investigator on human behavior and machine intelligence at the Joint Research Center, European Commission, 41092 Seville, Spain, and a guest professor at Universitat Pompeu Fabra. Her research interests include music information retrieval to human-centered artificial intelligence. Contact her at emilia.gomez-gutierrez@ec.europa.eu.

JOSEP SOLER GARRIDO is a team leader at the Joint Research Center, European Commission, 41092 Seville, Spain. His research interests include artificial Intelligence and algorithm auditing and transparency and regulatory inspection. Garrido received a Ph.D. in electrical and electronic engineering from the University of Bristol, U.K. Contact him at josep.soler-garrido@ec.europa.eu.

in the form of research papers and code repositories. However, they appear to be highly suitable to provide concise and accessible technical information for users of high-risk AI systems according to the requirements in the AI Act and have the potential to evolve into formal standards supporting the regulation.

To achieve this, some points may require the attention of standardizers and the wider AI transparency and documentation community. First, some of the identified gaps in terms of depth should be addressed. To this end, approaches beyond those considered in our work may be helpful. This includes not only AI documentation approaches, but other types of documents such as checklists for AI practitioners, for example, the *Assessment List on Trustworthy Artificial Intelligence* (ALTAI[13]) by the High-Level Expert group on AI from the European Commission.[14] The design checks in this or similar works, for example,[15] could be further developed into sections for documentation templates.

Another point potentially requiring attention is promoting consistency in their use. The technical depth of the information provided by different adopters appears to vary significantly. It may be beneficial to further detail the content to be included. This may be facilitated by the use of tools that automate the provision of information, and by detailed guidance for the selection and calculation of relevant metrics, for example, for accuracy, robustness or bias assessment. In this respect, some emerging AI system analysis toolkits, for example,[16] may play a role.

Finally, considering the examples analyzed, existing AI documentation approaches most often seem to describe general-purpose AI systems, for example, those performing generic computer vision and text processing tasks. Tailoring of existing templates to better describe AI systems with more concrete and nuanced risks and operation contexts—as those expected for high-risk AI systems—may be required.

Standards development organizations may be best suited to address these and any other necessary considerations for the evolution of existing AI documentation approaches into technical specifications for transparency and provision of information to the users of high-risk AI systems.

In this article, we presented an analysis of state-of-the-art approaches for documenting AI systems and datasets, assessing them from the point of view of the requirements set out in the European AI regulation

proposal. Our analysis shows that, while not fully aligned with the information elements mentioned in the legal text, these approaches represent a solid and useful basis toward operationalizing documentation requirements for high-risk AI systems at the technical level.

From the point of view of providing full documentation for authorities and conformity assessment bodies, the approaches reviewed could potentially demand additional depth in the description of various relevant technical information elements, for example, through the inclusion of detailed metrics and assessment results. Despite this, in their current form they appear to be very effective at providing concise and accessible information for AI practitioners and users, and could, with a moderate effort, be collectively extended to cover most of the technical information required to ensure proper understanding and use of high-risk AI systems.

The analysis and recommendations presented in this work are intended to inform the evolution—and potentially formal standardization—of AI documentation approaches in support of regulatory needs.

ACKNOWLEDGMENT

The authors would like to thank their colleagues and hierarchy in the Digital Economy Unit of the Joint Research Center for the fruitful discussions and their support. The opinions expressed are those of the authors only and should not be considered as representative of the European Commission's official position.

REFERENCES

1. "Proposal for a Regulation laying down harmonised rules on artificial intelligence," European Commission, Brussels, Belgium, 2021. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence
2. "Another arrest, and jail time, due to a bad facial recognition match," *NY Times*, 2021. [Online]. Available: https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html
3. "Full self-driving clips show owners of Teslas fighting for control, and experts see deep flaws," *Washington Post*, 2022. [Online]. Available: https://www.washingtonpost.com/technology/2022/02/10/video-tesla-full-self-driving-beta/
4. T. Gebru et al., "Datasheets for datasets," *Commun. ACM*, vol. 64, no. 12, pp. 86–92, Nov. 2021, doi: 10.1145/3458723.
5. S. Holland, A. Hosny, S. Newman, J. Joseph, and K. Chmielinski, "The dataset nutrition label: A framework to drive higher data quality standards," 2018, *arXiv:1805.03677*.
6. K. S. Chmielinski et al., "The dataset nutrition label (2nd gen): Leveraging context to mitigate harms in artificial intelligence," 2022, *arXiv:2201.03954*.
7. M. Mitchell et al., "Model cards for model reporting," in *Proc. Conf. Fairness, Accountability, Transparency (FAT)*, Jan. 2019, pp. 220–229, doi: 10.1145/3287560.3287596.
8. M. Arnold et al., "FactSheets: Increasing trust in AI services through supplier's declarations of conformity," *IBM J. Res. Develop.*, vol. 63, no. 4/5, pp. 6:1–6:13, Jul./Sep. 2019, doi: 10.1147/JRD.2019.2942288.
9. "OECD Framework for the Classification of AI Systems: A tool for effective AI policies," OECD, Paris, France, 2022. [Online]. Available: https://oecd.ai/en/classification
10. H. A. Linstone and M. Turoff, *The Delphi Method: Techniques and Applications*. Reading, MA, USA: Addison-Wesley, 1975.
11. B. Hutchinson et al., "Towards accountability for machine learning datasets: Practices from software engineering and infrastructure," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Mar. 2021, pp. 560–575, doi: 10.1145/3442188.3445918.
12. D. Piorkowski, J. Richards, and M. Hind, "Evaluating a methodology for increasing AI transparency: A case study," 2022, *arXiv:2201.13224*.
13. "Ethics guidelines for trustworthy AI," European Commission, Brussels, Belgium, 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
14. "ALTAI - The assessment list on trustworthy artificial intelligence," European Commission, Brussels, Belgium, 2019. [Online]. Available: https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence
15. M. A. Madaio, L. Stark, J. Wortman Vaughan, and H. Wallach, "Co-designing checklists to understand organizational challenges and opportunities around fairness in AI," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2020, pp. 1–14, doi: 10.1145/3313831.3376445.
16. R. K. Bellamy et al., "AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias," 2018, *arXiv:1810.01943*.