*Human Interaction and Emerging Technologies (IHIET-AI 2025),* Vol. 161, 2025, 84–93
https://doi.org/10.54941/ahfe1005899

# The Digital Trust Radar - Identification and Analysis of Guidelines to Support Responsible AI

**Janine Jäger, Jona Karg, Petra Maria Asprion, and Ilya Misyura**

Institute for Information Systems, School of Business, University of Applied Sciences and Arts Northwestern Switzerland FHNW, Basel, Switzerland

## ABSTRACT

The concept of Digital Trust can be utilized to classify and assess the responsible design, implementation and use of artificial intelligence (AI) technologies. Laws, standards, and guidelines are essential as they support the establishment of procedures that promote responsible AI technologies and therefore broad added value, societal acceptance and public confidence in AI. This contribution introduces the 'Digital Trust Radar', a structured digital repository synthesizing seventy-eight guidelines, standards and laws relevant to establish responsible AI in organizations. Through a systematic approach, these documents were categorized and analyzed based on various criteria including authorship, geographic focus, intended audience, AI application domain, AI type, and governance alignment. The findings reveal significant variability in the scope and thematic focus of AI related laws, guidelines, or standards, emphasizing ethical, legal, and technical considerations. Our categorization scheme provides a comprehensive overview of international approaches to support AI governance for responsible AI and serves as a valuable resource for stakeholders navigating the complexities of AI design, integration and usage.

**Keywords:** Artificial intelligence, Digital trust, Digital trust radar, Regulations, Responsible AI

## INTRODUCTION

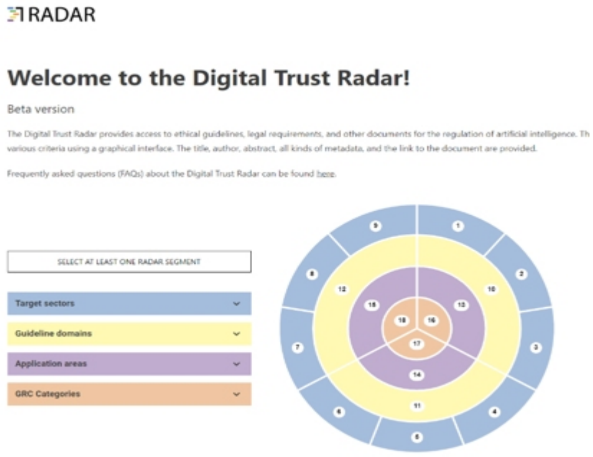
The rapid expansion of artificial intelligence (AI) technologies has introduced various applications in organizational contexts. Organizations across all sectors are seeking to integrate AI technologies into their operations to enhance efficiency and innovation, with several technological, organizational, and individual-related implications (Bankins et al., 2024; Lee et al., 2023). As organizations rapidly adopt AI technologies to boost efficiency and innovation, ensuring "Responsible AI" and "Digital Trust" has become an important concern for governments and organizations worldwide (European Commission, 2024b; UNESCO, 2024; World Economic Forum, 2024a).

"Responsible AI" is a term that encompasses several criteria that AI technologies should fulfil. "Responsible AI capabilities" can encompass the ability of organizations to address ethical challenges specific to AI by effectively utilizing suitable tools, practices, strategies, and processes (Akbarighatar et al., 2023). In our research, we use "Responsible AI" to describe efforts to develop and apply AI technologies responsibly, meaning the emphasis on accountability and the alignment with ethical principles and societal norms. According to Bendel (2021) this includes aspects such as explainability (Explainable AI), trustworthiness (Trustworthy AI), data protection, reliability and security. However, these criteria are not standardized and need further clarification through the community (Bendel, 2021).

"Digital Trust" can be defined as "individuals' expectation that digital technologies and services - and the organizations providing them - will protect all stakeholders' interests and uphold societal expectations and values" (World Economic Forum, 2024a). Trust should be created through a demonstrably trustworthiness of the AI system, which requires a shift of focus from performance-driven to trust-driven AI (Li et al., 2023). AI regulation is a necessary tool for merging the diverging driving forces of AI development and to facilitate the creation of trustworthy and responsible AI systems (Díaz-Rodríguez et al., 2023). A study by Gillespie et al. (2021) indicates that the belief that existing regulations and laws are adequate to ensure AI's safe usage is one of the strongest factors influencing trust in AI.

To summarize, "Digital Trust" and "Responsible AI" are interconnected concepts focused on building user confidence and ensuring ethical technology use. To achieve these objectives, organizations depend on laws, guidelines, and regulations for designing and utilizing AI technologies. While regulatory bodies are emphasizing AI ethics and are creating guidelines for responsible AI, these are primarily voluntary and so far lack enforcement when actual harm occurs (Smuha, 2021). Additionally, the complex and growing landscape of diverse laws, guidelines, or standards, hereafter referred to collectively as "documents", poses a challenge for developers and users of AI (World Economic Forum, 2024b).

Those challenges led to the following research question "*What relevant documents exist to establish responsible AI in organizations and how can they be systematically categorized*?"



RADAR

Welcome to the Digital Trust Radar!

Beta version

The Digital Trust Radar provides access to ethical guidelines, legal requirements, and other documents for the regulation of artificial intelligence. Th various criteria using a graphical interface. The title, author, abstract, all kinds of metadata, and the link to the document are provided.

Frequently asked questions (FAQs) about the Digital Trust Radar can be found here.

**Figure 1**: Landing page of the digital trust radar (Competence Center Digital Trust, 2024).

The results from our research were integrated into a structured and filterable digital web-based repository referred to as Digital Trust Radar (see Fig. 1). The radar aims at supporting various stakeholders in identifying relevant AI-related documents that guide them in their AI design, deployment, or use. In the following we discuss the applied methodology, key insights, and the implications for responsible AI.

## ANALYSIS APPROACH

We designed a structured digital repository and associated front-end of pertinent documents focused on responsible and trustworthy AI – the Digital Trust Radar. English or German were chosen as relevant languages based on the researchers being in a German-speaking context and English being the dominant language of international academia, regulation, and technological development. We conducted a systematic key word-based internet search from October 2023 to March 2024. For the search we used specific keywords in both English and German. The English keywords included "AI Guideline", "AI Regulation", "AI Law", "AI Recommendation", "AI Norm", and "AI Standard", with corresponding German translations to ensure comprehensive data retrieval. We identified 96 documents encompassing both legally binding and non-binding documents. Out of these documents, thirteen were excluded due to not meeting the inclusion criteria and another five being duplicates. Accordingly, 78 documents are included in the final analysis. Inclusion criteria for AI guidelines involved selecting documents and websites that directly address AI design, deployment and operation or usage, are issued by recognized bodies, and are publicly accessible. Exclusion criteria filtered out documents that only tangentially mention AI, are inaccessible due to paywalls, or are research-oriented contributions.

The attributes and categories for the document description and analysis were developed through a structured and methodical process aimed at covering a comprehensive range of aspects relevant to AI guidelines. Initially, a set of primary attributes was identified to describe the scope and relevance of each guideline. Attributes such as authorship, year of publication, type of publishing organization, origin, regional focus, and type of guideline were selected to understand the authority and geographical context of the documents. Attributes related to the medium, size, and available languages of the documents were included to assess accessibility for diverse audiences. For a more in-depth analysis of the applicability of the documents the following four categories with related sub-categories were developed:

1) **Target Sector:** This category was chosen to assess the applicability of documents across different sectors. Those sectors include: (a) *business* and economy, (b) *society and citizens*, (c) *policymakers and administration*, (d) *science and research*, (e) *healthcare*, (f) *arts*, (g) *non-profit and non-governmental organizations*, (h) *education*, and (i) *cross-sectoral*, describing documents not focusing on any specific sector.

2) **Application Area:** This category was chosen to understand whether the documents are aiming at (a) *AI design and development*, (b) *AI integration, deployment and operation*, or (c) *AI usage*.
3) **GRC (Governance, Risk, and Compliance):** This category was chosen to describe whether the documents focus on aspects such as (a) *governance* structures, (b) *risk management* procedures, and (c) *compliance* with certain regulations, standards, or similar policies.
4) **Guideline Domain:** This category was chosen to classify documents based on their thematic focus, helping to understand the specific areas of AI regulation that the documents cover. Three thematic focuses were selected for analysis which were (a) *ethics and law*, (b) *cybersecurity*, and (c) *technology and methods*.

All selected documents were then analyzed independently by two raters to ensure consistency in categorization and evaluation. Cohen's Kappa inter-rater reliability has been calculated for all four categories (see Table 1) in multiple iterations. The results of each iteration were then discussed by the two raters to improve the categorization of the next iteration. Thereby, in accordance with Landis and Koch (1977), Cohen's Kappa values greater than 60 were considered satisfactory and therefore indicate a reliable categorization of the documents. In addition, incoherent ratings were examined in detail and a mutual result was determined.

The initial round of analysis involved examining ten documents, with one being excluded. Despite satisfactory inter-rater reliability scores across all categories (see Table 1), the decision was made to conduct independent analyses by both raters in subsequent iterations due to insufficient reliability in specific sub-categories, particularly in the GRC and guideline domain sub-categories, which had to be refined for clarity. In the second iteration, another set of ten documents was analyzed, with 1 failing to meet inclusion criteria. There were mixed results in inter-rater reliability. Since it was not reasonable to assume that the inter-rater reliability would be consistently acceptable, it was decided to have both raters conduct the entire analysis process independently and then compare their results. The third iteration also analyzed ten documents, with one exclusion, resulting in stabilized inter-rater reliabilities. The fourth iteration followed a similar pattern, with four exclusions, and satisfactory reliability across all categories. Improvements in reliability were noted in three out of four categories during the fifth iteration. Nevertheless, four documents did not meet inclusion criteria. However, the sixth and final iteration presented a decline in reliability for the previously improved categories, potentially attributed to seven exclusions. Despite this, overall reliability remained within acceptable ranges, avoiding the need for further iteration. Subsequent discussions between the raters revealed significant deviations in the final ratings from the initial independent assessments, highlighting sharper distinctions in the categories during these deliberations.

**Table 1.** Cohen's Kappa inter-rater reliability.

| Documents | κ Target Sector | κ Application Areas | κ GRC | κ Guideline Domain |
|---|---|---|---|---|
| **01–10** | .818 | .629 | .868 | .729 |
| **11–20** | .733 | .808 | .650 | 1.00 |
| **21–30** | .838 | .792 | .762 | .874 |
| **31–50** | .885 | .795 | .795 | .759 |
| **51–73** | .880 | .977 | .886 | .929 |
| **74–96** | .933 | .649 | .780 | .643 |

## RESULTS

Regarding the descriptive attributes of the documents (see Table 2) the investigation revealed frequencies of various document characteristics that provide insights into authority and geographical context. The main type of publishing organizations for the documents were governmental organizations ($n = 28$), followed by non-governmental or non-profit organizations (NGO/NPO) ($n = 26$), for profit organizations (n = 16) and research organizations ($n = 8$). The documents reflect a diverse geographical origin with documents originating predominantly from international organizations ($n = 30$), followed by Germany ($n = 10$), and Switzerland ($n = 4$), the United States ($n = 15$), and the European Union ($n = 14$). The remaining documents originated in the United Kingdom ($n = 3$), Canada ($n = 1$), and the Vatican City ($n = 1$).

The date of publication of the included documents spans from 2017 to 2024. The majority were published in 2023 ($n = 24$), with fewer in 2020 ($n = 12$), 2024 ($n = 11$), and equal numbers in 2022 and 2021 ($n = 8$). The remaining documents were published in the years 2017 to 2019 ($n = 10$) or could not be assigned to a specific year of publication ($n = 5$). However, it is important to note that documents were only identified until March 2024; therefore, not all documents published in 2024 were included in the analysis.

Regarding the type of guidelines, most documents are non-binding guidelines or recommendations ($n = 61$), out of which two are additionally classified as codices. Only one document is a law, which is the EU Artificial Intelligence Act (European Commission, 2024a). The remaining documents are so-called codices from professional groups or industries ($n = 6$), including those two that were classified as guidelines as well, norms or standards ($n = 4$), and "other" documents ($n = 8$), comprising, amongst others, whitepapers and AI principles of large international companies.

Additionally, regarding document size, there was a relatively balanced distribution among small documents ($n = 21$) with up to 25 pages, medium-sized documents ($n = 18$) with 25 to 50 pages, and large documents ($n = 24$) with more than 50 pages, whereby another two documents could not be assigned a size due to their file formats. Of these 65 documents, a notable amount of twelve were additionally published on websites, and beyond that thirteen documents were published exclusively on websites. Most documents (n = 70) were published in English.

**Table 2.** Frequencies of selected attributes (N = 78).

| Attributes | Characteristics | *n* |
|---|---|---|
| Type of publishing organization | For profit organization | 16 |
| | Governmental organization | 28 |
| | NGO/NPO | 26 |
| | Research organization | 8 |
| Regional origin of the guideline | International | 30 |
| | Germany | 10 |
| | Switzerland | 4 |
| | USA | 15 |
| | European Union | 14 |
| | Other regions or countries | 5 |
| Year of publication | 2024 | 11 |
| | 2023 | 24 |
| | 2022 | 8 |
| | 2021 | 8 |
| | 2020 | 12 |
| | 2017–2019 | 10 |
| | Without identifiable publication date | 5 |
| Type of guideline | Guideline or recommendation | 61 |
| | Law or ordinance | 1 |
| | Codex | 6 |
| | Norm or Standard | 4 |
| | Other | 8 |

Regarding the analysis of the categories (see Table 3), the results show a predominant focus on the business and economic sector ($n = 50$). Another strongly represented target sector is policymakers and administration ($n = 32$), followed by science and research ($n = 21$) and, society and citizens ($n = 14$), In contrast, only eleven documents refer to education, ten to non-profit or non-governmental organizations and six documents refer to the arts sector. It should be noted that most documents mention several target sectors ($n = 48$), and less than half of the documents refer to just one target sector ($n = 30$), with most of these documents being assigned to the category business and economy ($n = 20$). Sixteen documents did not mention any specific target sector but declared a general applicability of the document, with another six documents declaring a general applicability but also mentioning specific sectors.

In terms of application areas, most documents ($n = 58$) are assigned to the sub-categories “Design and development” as well as “Integration, deployment, and operation”. Fewer documents address the usage of AI technologies ($n = 40$). However, most documents ($n = 54$) address more than one application area and only twenty-four documents are exclusively assigned to one of the three application areas, with eleven of these specifically relating to AI usage, seven to AI design and development, and six to AI integration, deployment and operation.

Regarding the GRC category, governance is the most frequently addressed sub-category with 55 documents, followed by risk ($n = 36$) and compliance

($n = 31$). Ten documents did not fit any of the GRC sub-categories, in contrast twenty documents spanned two of the three and seventeen all three GRC categories.

In the guideline domain category, ethics and law is the most prevalent sub-category with sixty-five documents in total, and twenty-one documents exclusively dedicated to this sub-category. This is followed by technology and methods ($n = 40$), and cybersecurity ($n = 33$). Whereby only ten documents are exclusively dedicated to one of these two sub-categories.

**Table 3.** Document ratings for each category and sub-category ($N = 78$).

| Categories | Sub-Categories | $n_{\text{total}}$ ($n_{\text{exclusive}}$) |
|---|---|---|
| Target Sector | Business and economy | 50 (20) |
| | Society and citizens | 14 (1) |
| | Policymakers and administration | 32 (5) |
| | Science and research | 21 (1) |
| | Healthcare | 14 (0) |
| | Arts | 6 (1) |
| | NPO and NGO | 10 (0) |
| | Education | 11 (2) |
| | Cross-sectoral | 22 (16) |
| Application Areas | Design and development | 58 (7) |
| | Integration, deployment and operation | 58 (6) |
| | Usage | 40 (11) |
| GRC | Governance | 55 (22) |
| | Risk | 36 (4) |
| | Compliance | 31 (5) |
| | Not GRC specific | 10 (10) |
| Guideline Domain | Ethics and law | 65 (21) |
| | Cybersecurity | 33 (2) |
| | Technology and methods | 40 (8) |

## DIGITAL TRUST RADAR DEVELOPMENT

After the analysis and categorization, the collected documents were then integrated into a digital repository called "The Digital Trust Radar", which is accessible via the website https://radar.digitaltrust-competence.ch. This website was built on a technology stack optimized for the needs of the radar providers and users. Required components for this stack included a) a content management system with a built-in secure user management that is easy to deploy and customize, b) a NoSQL database to address frequent data structure changes, and c) a website frontend that is built using a widely supported framework. For the content management, Payload CMS was chosen. For data storage, we selected MongoDB, a NoSQL solution. The front end was built using Next.js, a React framework. The radar-like widget for the initial user interaction was built with the database using Chart.js.

The first prototype has been evaluated by various stakeholders and potential users (n = 30). Significant improvements in the front-end design and usability have already been implemented and the current prototype is largely stable and available for further validation. A second round of improvements including the integration of further up-to-date documents is planned for spring 2025 and will then be continued on an ongoing basis.

## CONCLUSION, DISCUSSION AND FUTURE OUTLOOK

In this contribution, we described the methodology and key findings of a collection and analysis of international guidelines related to the design, application and usage of AI technologies. Most documents in our sample are published by governmental as well as non-governmental or non-profit organizations, followed by for-profit organizations, suggesting a multi-stakeholder approach to AI regulation. The significant role of governmental organizations and NGO/NPOs in publishing AI guidelines might indicate an interest and concern regarding the ethical, societal, and regulatory impacts of AI technology, which should be explored in further research. Furthermore, our research shows a significant involvement from international, European and US-based organizations in developing AI guidelines, highlighting that technologically advanced regions are proactive in the development of AI governance, thereby shaping international AI-related norms. Future research could explore the diverse global impacts of AI across all regions more in depth. The increase in publications, particularly in 2023 with continued activity into 2024, indicates an evolving regulatory landscape responsive to new technological and ethical challenges. A key characteristic of current AI governance is the existence of non-binding guidelines over binding legal frameworks, potentially rooted in the fast evolution of AI technology to which traditional legislative processes cannot adapt. However, non-binding guidelines are more flexible and can be updated or revised quickly to keep pace with new developments and discoveries.

Our research also shows a focus on guidelines addressing the design, development, and operational stages of AI systems, resembling the rapid technological advancements and extensive integration of AI across various sector, but especially in the business domain. Furthermore, a substantial number of documents is dedicated to AI governance as well as ethical and legal aspects, highlighting the effort to ensure responsible AI development and utilization. The prevalence of documents focused on the “Business and Economy” sector underscores the critical importance of AI in economic activities and suggests a strong alignment of AI guidelines with economic growth and innovation priorities. However, most documents are not sector-specific and guidelines for certain sectors, such as healthcare and education, although critical with regards to the need for responsible AI, are underrepresented in our sample, presenting a need for guidelines with a focus on these sectors as well as further research on AI regulation in these sectors.

Governance, risk, and compliance aspects of AI are well represented in our sample, with governance being the focus. This underscores the critical importance placed on establishing robust frameworks for overseeing AI operations, which are essential for maintaining public trust and legal compliance. However, the underrepresentation of guidelines specific to

"Risk Management" and "Compliance", with the exception of the EU AI Act (European Commission, 2024a), suggests a need for enhanced focus on such frameworks. The strong emphasis on ethical and legal aspects underscores the ongoing effort to align AI technologies with human values and legal standards. Cybersecurity aspects seem to be integrated within broader guidelines, rather than being the sole focus. Given the evolving nature of cyber threats in the context of AI, there remains a pressing need for updated, specific guidelines that address the unique cybersecurity challenges of AI. Furthermore, the significant representation of documents addressing technological and methodological aspects of AI highlights the importance of ensuring that AI systems are reliable, efficient, and capable of performing their intended functions.

Our research does have some limitations. Firstly, the temporal scope of the analysis did not cover all documents published in 2024 due to the cutoff in March 2024, potentially missing later developments that could impact the analysis outcomes. Expanding the research scope to include further underrepresented sectors and regions would provide a more holistic view of the regulatory landscape. Investigating the interconnections between sectors more deeply as well as the applicability of guidelines could also generate deeper insights into the multi-faceted nature of AI applications and the regulatory challenges they present. Longitudinal studies could monitor how the focus and frequency of topics evolve over time, providing ongoing feedback to policymakers and stakeholders.

To conclude, our research not only provides a snapshot of the current landscape of AI related guidelines and their characteristics but also suggests areas for further research and exploration, particularly in underrepresented sectors and application areas. The dynamic increase in guideline publications and ongoing updates suggest an actively evolving regulatory landscape that must continuously adapt to new technological challenges and opportunities, ensuring responsible AI development across all sectors. Our insights can be valuable for stakeholders seeking to understand and influence the regulatory environment surrounding AI, guiding strategic decisions, and policy formulation. Mapping the global landscape of AI policies in the format of our Digital Trust Radar, which will continuously evolve, not only helps organizations navigate the complex field of AI governance, but also serves as a foundational tool for future research and development of responsible and trustworthy AI systems.

## ACKNOWLEDGMENT

The authors would like to acknowledge that the project was funded by the Stiftung FHNW (https://www.fhnw.ch/de/die-fhnw/stiftung-fhnw) and the Institute for Information Systems of the School of Business FHNW (https://www.fhnw.ch/en/about-fhnw/schools/business/iwi). The project was co-led by Prof. Dr. Oliver Bendel and Prof. Dr. Petra Maria Asprion.

## REFERENCES

Note: the 78 documents in our sample are not listed here due to limited space, but are documented in the Digital Trust Radar (https://radar.digitaltrust-competence.ch).

Akbarighatar, P., Pappas, I., & Vassilakopoulou, P. (2023). A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review. *International Journal of Information Management Data Insights*, *3*(2), 100193. https://doi.org/10.1016/j.jjimei.2023.100193

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*, *45*(2), 159–182. https://doi.org/10.1002/job.2735

Bendel, O. (2021). *Definition: Responsible AI*. Springer Fachmedien Wiesbaden GmbH. https://wirtschaftslexikon.gabler.de/definition/responsible-ai-123232

Competence Center Digital Trust. (2024). *Digital Trust Radar*. FHNW School of Business. https://radar.digitaltrust-competence.ch/

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

European Commission. (2024a). *EU Artificial Intelligence Act—Up-to-date developments and analyses of the EU AI Act*. https://artificialintelligenceact.eu/

European Commission. (2024b). *Trustworthy artificial intelligence (AI)*. Excellence and Trust in Artificial Intelligence. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en

Gillespie, N., Lockey, S., & Curtis, C. (2021). *Trust in artificial Intelligence: A five country study*. The University of Queensland; KPMG. https://doi.org/10.14264/e34bfa3

Lee, M. C. M., Scheepers, H., Lui, A. K. H., & Ngai, E. W. T. (2023). The implementation of artificial intelligence in organizations: A systematic literature review. *Information & Management*, *60*(5), 103816. https://doi.org/10.1016/j.im.2023.103816

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023). Trustworthy AI: From Principles to Practices. *ACM Comput. Surv.*, *55*(9), 177:1-177:46. https://doi.org/10.1145/3555803

Smuha, N. A. (2021). From a 'race to AI' to a 'race to AI regulation': Regulatory competition for artificial intelligence. *Law, Innovation and Technology*, *13*(1), 57–84. https://doi.org/10.1080/17579961.2021.1898300

UNESCO. (2024). *Ethics of Artificial Intelligence—The Recommendation*. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

World Economic Forum. (2024a). *Digital Trust—Decision-Making for Trustworthy Technology*. https://initiatives.weforum.org/digital-trust/home

World Economic Forum. (2024b, October 14). *Why corporate integrity is key to shaping future use of AI*. https://www.weforum.org/stories/2024/10/corporate-integrity-future-ai-regulation/