



The landscape of data and AI documentation approaches in the European policy context

Marina Micheli¹ · Isabelle Hupont² · Blagoj Delipetrev¹ · Josep Soler-Garrido²

Published online: 28 October 2023
© The Author(s) 2023

Abstract

Nowadays, Artificial Intelligence (AI) is present in all sectors of the economy. Consequently, both data-the raw material used to build AI systems- and AI have an unprecedented impact on society and there is a need to ensure that they work for its benefit. For this reason, the European Union has put data and trustworthy AI at the center of recent legislative initiatives. An important element in these regulations is transparency, understood as the provision of information to relevant stakeholders to support their understanding of AI systems and data throughout their lifecycle. In recent years, an increasing number of approaches for documenting AI and datasets have emerged, both within academia and the private sector. In this work, we identify the 36 most relevant ones from more than 2200 papers related to trustworthy AI. We assess their relevance from the angle of European regulatory objectives, their coverage of AI technologies and economic sectors, and their suitability to address the specific needs of multiple stakeholders. Finally, we discuss the main documentation gaps found, including the need to better address data innovation practices (e.g. data sharing, data reuse) and large-scale algorithmic systems (e.g. those used in online platforms), and to widen the focus from algorithms and data to AI systems as a whole.

Keywords Trustworthy AI · Transparency · AI policies · Documentation

Introduction

The widespread adoption of Artificial Intelligence (AI) across all sectors of the economy is having an unprecedented impact on society. This technology brings new opportunities for prosperity and development, including in critical application areas such as healthcare, agriculture and the environment. However, it also introduces a number of risks, for example related to the automation of high-stakes decisions such as criminal recidivism (Tolan et al., 2019), law enforcement based on facial analysis (Hupont et al., 2022) or bank

credit scoring (Demajo et al., 2020). Similarly, data, the raw material used to build AI systems, has proven to be an extremely valuable asset, and attention must be given to its collection, distribution and use by the different economic actors, in order to ensure it works for the benefit of society as a whole (Boyd & Crawford, 2012; Caton & Haas, 2020). For this reason, AI and data are attracting the attention of policymakers across many jurisdictions, including the European Union, where they are the subject of a number of recent high-profile legislative and investment initiatives (European Commission, 2021a, 2022d).

Transparency has been acknowledged as one of the key pillars of trustworthy AI by high-level expert groups, from reference institutions such as the European Commission (European Commission, 2019) and the Organisation for Economic Cooperation and Development (OECD, 2022). It affects both AI systems and the data used for their development and during operation. Transparency encompasses several concepts. On the one hand, it covers traceability and explainability mechanisms, so that the use of data, the behaviour and the decisions made by an AI system can be traced, logged and explained to the stakeholders concerned. On the other hand, it refers to a clear, truthful and compre-

✉ Marina Micheli
marina.micheli@ec.europa.eu

Isabelle Hupont
isabelle.hupont-torres@ec.europa.eu

Blagoj Delipetrev
blagoj.delipetrev@ec.europa.eu

Josep Soler-Garrido
josep.soler-garrido@ec.europa.eu

¹ Joint Research Centre, European Commission, Viale Fermi 2749, 21027 Ispra, VA, Italy

² Joint Research Centre, European Commission, Seville, Spain

hensive disclosure of information about the capabilities and limitations of an AI system, including those related to data, e.g. how they were collected, quality issues and potential biases (Díaz-Rodríguez et al., 2023; Hupont & Fernández, 2019; Pasquale, 2015). While the abstract ideal notion of transparency does not guarantee reliability and accountability (Ananny & Crawford, 2018), establishing effective documentation practices and processes for AI can be an important step on the road to trustworthy AI. Data and AI documentations can be crucial tools, e.g., to assist the performance of independent audits of AI systems, gain trust from the general public and provide a framework for authorities to build regulatory requirements (Falco et al., 2021).

Therefore, moving beyond conceptual accounts of transparency, this article examines a particular type of “tool” that can be adopted to substantially “produce transparency” (Hansen & Flyverbom, 2015). Documentation approaches for AI, in fact, are commonly referred to as “transparency artefacts” (Pushkarna et al., 2022; Rostamzadeh et al., 2022) and “transparency tools” (Stoyanovich & Howe, 2019) since they allow transferring information and improving communication between different stakeholders such as AI developers, data providers and users. If these become embedded in standard practices, they could contribute to addressing problems related to the lack of transparency in AI. At the moment, however, “documentation is currently uncommon in the development of algorithmic systems and there is no agreed upon format for what should be included when documenting the origin of a dataset” (Tsamados et al., 2022).

This, however, may be about to change, as the adoption of structured documentation practices for data and AI is shifting from being voluntary to becoming a strict legal requirement. Documentation needs are prominently addressed in several regulations on data and AI currently under development in the European Union, including the AI Act (European Commission, 2021a), the Digital Services Act (European Commission, 2022b), the Data Governance Act (European Commission, 2022c) and the Data Act (European Commission, 2022d). Within the evolving European policy context, documentation is considered a key driver for accelerating the adoption of AI systems in conformity to existing laws and fundamental rights (Hupont et al., 2023a). It is a key element to support a wide range of regulatory objectives, whether these are related to mitigating the risks of AI systems in critical applications, assessing conformity with legal AI trustworthiness requirements, or promoting sharing and reuse of data assets under fair terms. Documentation can indeed contribute towards increased data sharing between EU companies, public administration, universities, citizens and civic organisations, facilitating the discovery and use

of quality datasets for AI. Overall, if widely and efficiently adopted, documentation can promote trust and transparency in data sharing, help alleviate technical and ethical concerns surrounding AI, and guide a fairer digital transformation.

In the last few years, several approaches have been proposed by AI practitioners in academia and industry to document datasets and AI models with the goal to foster responsible and trustworthy AI, and to mitigate potential risks and harms. Previously, significant work has been done in the field of data management, in particular various efforts converged to standardize data documentation and metadata according to FAIR principles (see for instance the Data Stewardship Wizard project¹). The FAIR principles define key characteristics that datasets, algorithms, tools and workflows should have for good data management and stewardship. The guiding principles are Findability, Accessibility, Interoperability and Reusability (Wilkinson et al., 2016). Similarly to some of the most recent AI-oriented documentation approaches examined in this article, FAIR principles’ goal is to increase transparency and foster trustworthy data handling practices to increase trust and promote access and reuse of data among researchers and citizens. In this study we draw also from those principles, and extend analysis to include, among others, concerns such as ethics, accountability, trust and quality, always from the lens of data and its use for Artificial Intelligence. Documentation approaches analysed support developers and other stakeholders to reflect on and to tangibly report key information about datasets, AI models or AI systems. While some of these documentation approaches have already seen some level of adoption by the relevant actors, to the best of our knowledge they have not been the subject of a study analysing them in a holistic manner, identifying their commonalities, complementarities and gaps.

In this article we present the first comprehensive landscape review of documentation approaches for data and AI, and analyse it through the lens of current European policy needs. We have reviewed more than 2200 papers that touch, to a varying extent, on the problem of documenting data and/or AI. From these, we have identified 36 initiatives proposing full original documentation approaches, and categorised them according to a set of parameters. The aim of this work is, on the one hand, to provide scholars with a better understanding of existing documentation approaches, their objectives, how to catalogue them and to what extent they address the transparency needs of current EU policy proposals and regulations on data and AI. On the other hand, it aims at helping practitioners, policymakers and authorities to better navigate this landscape and identify possible sources that meet their needs.

¹ <https://docs.ds-wizard.org/en/latest/>.

Data and AI documentation in support of European policies and initiatives

The increasing availability of data and the resulting adoption of AI-based solutions are having a transformational effect on societies and economies around the world. Consequently, they are also rising to the top of policy agendas in many jurisdictions. This includes the European Union, which is starting to write a new rulebook for its digital future through a number of ambitious legislative and investment initiatives with data and AI taking a central role, as depicted in Fig. 1.

To structure the presentation of different initiatives, the comparison of their transparency needs and how these are covered by current documentation approaches, we group them in two broad categories: innovation-oriented approaches and risk-oriented approaches (European Parliamentary Research Service, 2022).

To the first category belongs the European Commission's Data Strategy, which was announced in 2020 to spur innovation, creating a single market for data within the EU across sectors and Member States. The strategy has the overall objective to increase data flows through the European economic and societal fabric, facilitating data access and (re-)use through a new set of rules and frameworks. It comes with two key regulation proposals, the Data Governance Act (European Commission, 2022c) and the Data Act (European Commission, 2022d), with wide-ranging impacts on the availability of data assets for most societal stakeholders. These are supported by, and are complementary with, digital initiatives funded under the Digital Europe Programme (DEP) (European Commission, 2021b).

Within the "risk-oriented" category we include regulations and proposals aimed at addressing and mitigating risks deriving from AI systems, platforms and data analytics, such as discrimination, impact on fundamental rights or disinformation. A key initiative in this area is the AI Act (European Commission, 2021a), which is currently under negotiation involving the European Parliament, the Council of the European Union and the European Commission. Once adopted, it will set out the regulatory conditions for the adoption of trustworthy AI practices in the European Union. The AI Act is complemented by the recently presented AI Liability Directive (European Commission, 2022a), which aims to ensure that persons harmed by AI systems enjoy the same level of protection as is the case with other technologies. Another key initiative expected to be imminently adopted by the European co-legislators, the Digital Services Act (European Commission, 2022b), imposes obligations for online intermediaries and platforms, seeking to comprehensively address the most pressing societal risks emerging from their use. This includes very large online platforms and search engines offering services underpinned by complex algorithmic systems and AI, e.g. for content moderation and recommendation.

In this section, we outline the relevance of data and AI documentation approaches in the context of the current European regulatory landscape.

Data documentation in the context of the European strategy for data

According to EU data policies, important benefits, in terms of innovation, economic development and societal well-being, can be achieved through greater availability and circulation

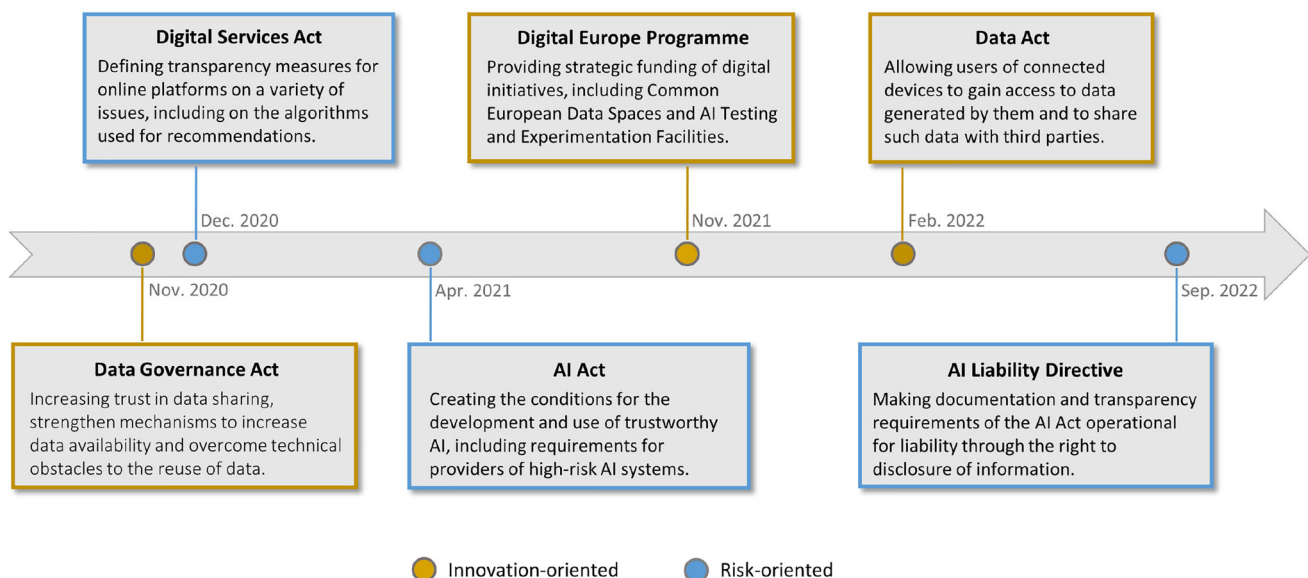


Fig. 1 Relevance of AI and data documentation in the ongoing European regulatory landscape

of data in key economic sectors such as healthcare, the environment, energy, agriculture and others. This is among the objectives of the Data Governance Act—the first legislative proposal released within the European Strategy for Data. This regulation is set to promote trust in data sharing and lead to an increase in data availability, e.g. through the introduction of the new roles of data intermediaries and data altruism organisations. It also paves the way for the creation of new European digital infrastructures for data sharing, the common European data spaces subsequently funded through the DEP (European Commission, 2021b). Data spaces encompass both technical elements to enable data flows to be established between ecosystem participants as well as governance structures to facilitate trusted and sustainable data sharing practices.

Dataset transparency and documentation are bound to play an important role in the functioning of these legislative and investment initiatives. Data producers, intermediaries and consumers are expected to have specific data documentation needs in order to maximise data sharing and the value extracted from data within data ecosystems. As such, data spaces will benefit from documentation approaches that increase transparency of the data made available in them. Participants will then have the necessary controls and mechanisms to find relevant datasets made available by other participants and be able to assess their suitability for specific use cases. Dataset documentation can be part of data spaces governance, supporting data discovery, informing stakeholders about data provenance and quality, ultimately enabling them to generate an added value, e.g. by means of AI services and solutions. These transparency mechanisms could have a strong beneficial impact, especially in domains with a fragmented data landscape and many existing data silos, enabling in particular smaller players, such as small and medium-sized enterprises (SMEs), to thrive and innovate.

Data documentation approaches could also be highly relevant for other stakeholders such as data altruism organisations. These organisations are expected to act in a transparent manner, providing accurate information about their activities (i.e. purposes of data processing, how general interest is pursued, and technical means adopted) to both competent authorities, data subjects and the general public.

Beyond the realm of data sharing ecosystems and related entities and organisations, important data assets emerge from many aspects of daily life. A notable example is data generated through our interactions with machines such as internet of things (IoT) devices. This is an important element in the Data Act, the recent European regulation ensuring fairness in the allocation of value from data. Enhancing innovation through a more competitive data market, e.g. by allowing access to machine-generated data by parties other than the product manufacturer, is one of the objectives of the Data Act. This is one example where a regulation can be effectively supported by data documentation, enabling developers

in innovative companies (e.g. AI system providers) to create new services and connected products, e.g. for smart home appliances or industrial machines, based on data generated through their operation by end-users, who are given the right to access and share it. Given the complexity of these machines, thorough data documentation approaches linked to - and complementing—technical specifications describing the devices and machines themselves would be particularly beneficial, as well as standardised approaches for documenting widely used sensor data in the specific domains of consumer and industrial goods. This is an example of the potential relevance of dataset documentation in the context of the Data Act. Other provisions in this broad-ranging regulation may also benefit from comprehensive data documentation practices, such as scenarios requiring sharing private sector data in the public interest to improve evidence-based decision-making in emergency situations.

AI documentation and EU regulations on AI and digital services

In April 2021, the European Commission presented its proposal for the Regulation of Artificial Intelligence (the AI Act). The AI Act lays down a set of legal obligations for providers of AI products and systems which may bring about new risks or negative consequences for individuals or society, defining requirements that depend on their risk profile. In particular, high-risk AI systems—those with a potential negative impact on the health, safety or fundamental rights of individuals—have to meet specific requirements that ensure a high level of trustworthiness. In order to satisfy these requirements, technical documentation of the AI systems and the datasets used to train them play an essential role.

Most stakeholders involved in the lifecycle of high-risk AI systems are expected to benefit from AI and dataset documentation approaches. This includes those certifying that AI products meet regulatory requirements. Whether concerning internal stakeholder in the AI provider organisations performing self-assessment activities, or external conformity assessment bodies, technical documentation of the AI system will be a basis for assessment regarding compliance with legal requirements. Similarly, authorities will also benefit from technical documentation of high-risk AI systems, e.g. in the context of market surveillance activities. Transparency for the users of AI systems can also be built on the basis of AI documentation approaches. Indeed, existing documentation initiatives are well suited for the provision of information to users of AI systems, and could potentially evolve towards technical standards in support of regulatory needs. In fact, before an AI product is placed on the market, AI documentation approaches and related initiatives such as design checklists can contribute to the adoption of trustworthy AI development practices, e.g. by promoting communication

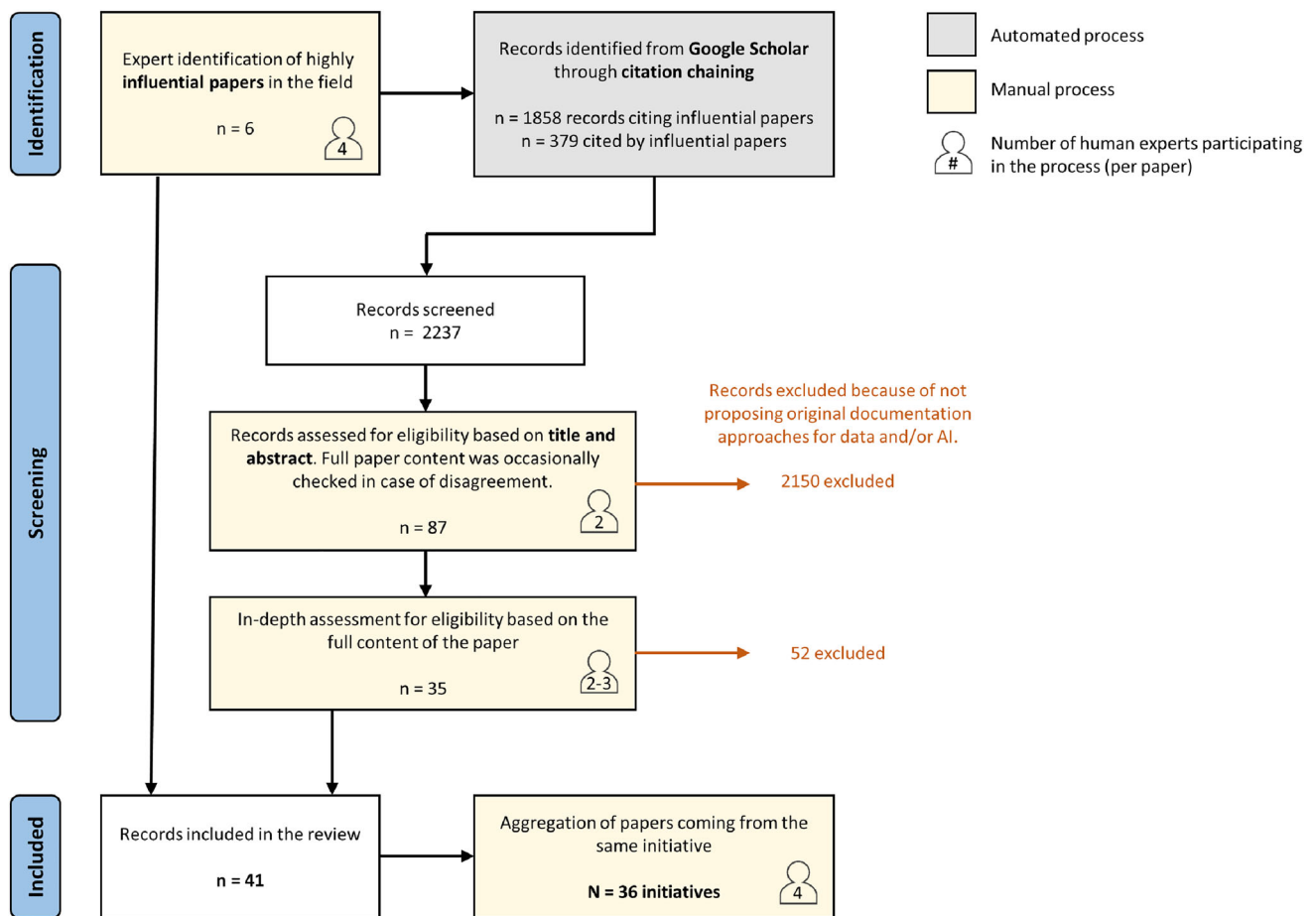


Fig. 2 Protocol followed for the discovery of state-of-the-art documentation approaches for data and AI, which complies with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021)

and reflection about potential AI risks within development teams.

It is not only those stakeholders directly involved in the design, development, verification or use of AI systems, which are the direct beneficiaries of AI transparency approaches. As an example, with the adoption of the AI Liability Directive, society at large will enjoy increased protection from damage resulting from the utilisation of AI systems, with documentation playing a role in addressing the specific difficulties of proof linked with AI, in order to ensure that justified claims are not hindered.

Beyond AI product legislation, AI documentation approaches are expected to play a relevant role in the context of European online platform regulation activities, most notably the Digital Service Act (DSA), which aims to create a safer digital space where the fundamental rights of users are protected. The DSA introduces new rules for online intermediaries and platforms such as social networks, search engines or online marketplaces, including a call for algorithmic accountability and transparency. In particular, concrete transparency and risk management obligations are defined

for very large online platforms and search engines. Many services on these platforms are based on recommendation, search, and advertising algorithms with potentially systemic risks, i.e. impacting the entire society as a whole, such as extremism, manipulation, discrimination or disinformation. Comprehensive documentation approaches for the algorithmic systems in use by these platforms would be useful for independent auditors and authorities involved in regulatory supervision and enforcement, as well as for vetted researchers independently investigating their potential risks. In this regard, documentation approaches tailored to content moderation, recommendation and search algorithms would be particularly suitable, as well as approaches for documenting large-scale algorithmic systems with a highly complex internal architecture, large user bases and potentially systemic societal impact.

Besides legislative acts like the AI Act or the DSA, European investment initiatives are also poised to benefit from AI documentation approaches. A notable example are the AI Testing and Experimentation Facilities (TEFs), also planned under the Digital Europe Programme. These are conceived

as physical or virtual facilities for providers of AI-based solutions to carry out integration, test, experimentation and validation activities in real-world environments, supporting their innovative potential and competitiveness. These facilities can benefit from both dataset as well as AI documentation approaches. To begin with, TEFs are expected to provide relevant datasets, including documentation, to AI developers in order to support them in the creation of field-tested AI solutions at scale. During the early phases of development and piloting, AI documentation has the potential to streamline interactions between AI development and field-testing and AI deployment stakeholders, accelerating iterative solution refinement and advancement towards higher readiness levels. Ultimately, the resulting AI-based applications will be brought to the market, perhaps being made available to end-users through communities such as sector-specific data spaces, which require AI documentation approaches suitable for their respective target users in specific sectors.

Methodology to collect and analyse documentation approaches

In this section we describe how we performed the systematic review of state-of-the-art approaches for the documentation of datasets and AI, explaining the methodology followed to collect and assess existing approaches, summarised in Fig. 2.

Dataset and AI documentation is an emerging practice, which has been shaped by a selected set of recent key publications. For this reason, instead of using a predefined keyword search, we based our review on a cluster of six highly influential papers, namely: *Datasheets for datasets* (first published in 2018) (Gebru et al., 2021), *Data statements for Natural Language Processing (NLP)* (Bender & Friedman, 2018), *The dataset nutrition label* (Holland et al., 2018), *AI FactSheets* (Arnold et al., 2019), *Model cards* (Mitchell et al., 2019) and *Accountability for Machine Learning (ML) datasets* (Hutchinson et al., 2021). These six sources were selected with the objective to achieve a high recall in the retrieval of relevant papers through citations. The selection was based on a combination of criteria, including their high and rapidly increasing number of citations, the wider adoption of the proposed approaches by practitioners, and the prominence of the entities advancing them (including key academic and industry players in the AI space). In addition, our selection has been contrasted and validated using similar analyses from stakeholders in the AI transparency and standardisation communities, such as the Partnership for AI's list of reference documents,² an analysis of ML documen-

tation tools by Hugging Face³ or ISO/IEC standardisation documents currently in development covering transparency of AI.⁴

As a first step, we used Google Scholar⁵ to collect all the publications citing any of the six influential papers and published until October 2022. Google Scholar allows to obtain both peer-reviewed journal articles and content published in online repositories, universities and other websites. This is useful for exploring a rapidly emerging field, for which many key outputs are still available in a pre-print format. As a result of this search, we obtained 1858 records. Additionally, we added other 379 sources that were cited by the influential papers. In total, 2237 papers were identified for screening.

In a second step, all resources in the list were examined by the four authors of this article with the goal to select those potentially relevant for the purpose of the study. We considered as relevant papers those proposing original approaches for dataset and/or AI model documentation, using the following concrete inclusion and exclusion criteria:

- Included papers encompass not only resources proposing horizontal documentation approaches (i.e., approaches applicable to different AI fields), but also those tailoring horizontal approaches to specific fields or types of data as they introduce novel questions, items, structures or modules.
- We exclude from the analysis resources that do not propose novel documentation approaches, such as literature reviews, surveys, commentaries or theoretical pieces.
- We exclude papers presenting software toolkits for the automation of documentation, such as *The Model Card Toolkit*⁶ and the *Symphony* framework (Bäuerle et al., 2022). They are not included in the scope of this survey as rather than describing novel documentation approaches, these contributions focus on the provision computational tools to automate existing ones.
- We exclude contributions that focus mostly on representation formats for AI documentation. Examples include recent proposals for the generation of machine-readable AI and dataset reports based on semantic technologies (Amith et al., 2022; Naja et al., 2022). These approaches are extremely useful, as they enable, for example, complex queries to be run on the documentation elements. However, they are not in the scope of this study as describing specific information elements to document is not their primary focus.

³ <https://huggingface.co/docs/hub/model-card-landscape-analysis>.

⁴ <https://www.iso.org/standard/84111.html>.

⁵ Google Scholar academic papers search engine: <https://scholar.google.com/>.

⁶ The Model Card Toolkit GitHub repository: <https://github.com/tensorflow/model-card-toolkit>.

² <https://partnershiponai.org/paper/about-ml-reference-document/13/#Appendix-A>.

In this second phase of analysis, title and abstract for all papers were examined, while the content was checked occasionally in case of disagreement. Each paper was cross-checked by two researchers and meetings were held to define those to be included or removed from the landscape. After this preliminary screening, 87 papers were flagged as potentially relevant for the review.

In a third step, the 87 resulting approaches were examined in depth. Each paper was scrutinised by at least two researchers to determine whether the article presented a novel approach for documentation of datasets and/or AI. 35 papers were selected after this in-depth screening. The final sample comprises in total 41 manuscripts (the 35 screened plus the 6 seminal ones), belonging to 36 different documentation initiatives.

Finally, we annotated each initiative according to the following set of criteria: (1) main focus of the documentation approach (distinguishing between data, AI model or AI system); (2) type of methodology adopted for documentation; (3) type of personas (i.e. stakeholders) to which the documentation is addressed; (4) scope (i.e. application area), reflecting on whether it is a horizontal approach that can be potentially applied to any field or if it is conceived ad-hoc for a concrete area (e.g. education, health, etc.), type of data (e.g. image data, text data) or AI task (e.g. NLP, ML); (5) whether it has some level of automation (e.g. statistics generated automatically from data) or not; (6) the main goals and concerns addressed by the approach.

All authors examined the final initiatives and proposed how to classify them according to each criterion. The classifications were discussed and consolidated during dedicated meetings. In the next section we present the results of the analysis and in the following one we discuss how they relate with the current policy needs.

Current landscape on documentation approaches for datasets and AI

In this section we present the findings of our systematic review of the current state-of-the-art approaches for the documentation of datasets and AI. Annex A provides the detailed annotation of each of the 36 documentation approaches and Fig. 3 shows the most relevant statistics. Below, we summarise the main results by each analytical criterion.

The increase in the number of documentation approaches in the years 2021–2022 is remarkable: from 3–4 initiatives per year in 2018–2020 to 11–14 in 2021–2022.⁷ The community is therefore recognising the importance of transparency and is counting more and more strongly on the matter.

Most documentation approaches (n. 18) focus on data exclusively, followed by 14 focusing on both data and models (AI system), and finally by only 4 that cover AI models exclusively. Many initiatives for documentation are based on the premises that the impact of data (its quality, representativity, origin, etc.) in AI systems is still underestimated. As claimed by the authors of *The Dataset Nutrition Label* (Chmielinski et al., 2022), data used to train models are “an often-overlooked site of harm” because problematic, incomplete or biased datasets cause models to replicate issues embedded in the training dataset.

We also examined the scope of the initiatives, e.g. whether they apply to a specific field or data domain. We found that most approaches (n. 15) are horizontal in scope. Therefore those documentation approaches are appropriate to all kinds of datasets and models as they work cross-sectors. The remaining approaches, instead, elaborate documentations for specific fields/sectors or types of data. Several derive from the pioneering *Datasheets for datasets* (Gebru et al., 2021), adapting it to a particular domain such as the arts (Srinivasan et al., 2021), health (Rostamzadeh et al., 2022) and journalism (Showkat, 2022). Initiatives for distinct data types include Natural Language Processing (NLP) (Bender & Friedman, 2018; Lhoest et al., 2021; McMillan-Major et al., 2021; Shimorina & Belz, 2021), population data (Anik & Bunt, 2021), or images datasets (Prabhu and Birhane, 2020). The advantage of tailoring documentation to sectors is expressed for instance by Richards et al. (2020), which claim that in this way documentation can be appropriate for different industries and the different regulatory schemes within which these industries operate. Other studies (DrivenData, 2022) suggest to adopt custom documentation, by customising a default one (horizontal), in order to address domain specific concerns which are not included in it. These more specialised initiatives have been published more recently, compared to the influential papers we draw from, and thus it can be suggested that this trend towards sector-specific documentation approaches is likely to continue in the near future. It is also worth noting that many documentation approaches, although general in scope, are designed to document from the lens of a specific use case, with publications providing examples of specific scenarios and applications.

Regarding the methods proposed for documentation, we have identified the four different types defined in Table 1: (1) questionnaires; (2) information sheets; (3) composable widgets; and (4) checklists. These methods vary in terms of length, level of detail, visualisations and in the guidance they provide to those who have to fill them in. Questionnaires (n. 15) and more concise—typically one-pagers—and visual information sheets (n. 15) are the most popular approaches. Fewer initiatives propose other documentation methods, such as widgets (n. 4) and checklists (n. 2).

⁷ Please note that for 2022 we do not cover the whole year, but the period January–October.



Fig. 3 Most relevant statistics resulting from the systematic review of the 36 state-of-the-art approaches for the documentation of datasets and AI

The four methods satisfy different transparency needs. Questionnaires usually provide more in-depth coverage, fostering reflection and an understanding of the ethical implications. Their open-ended nature, however, also poses challenges. For instance questions need to be very clear and narrowly formulated if the documentation is meant to increase the reproducibility of studies (Ramírez et al., 2021) or regulatory compliance, otherwise it would risk generating a “transparency fallacy”, an illusion of remedy with no meaningful impact (Edwards & Veale, 2018). To address this shortcoming, questionnaires could be accompanied by other methods such as information sheets and widgets, that are more well-defined and structured, at times automatically generated, which can additionally increase machine-to-machine (M2M) data and model sharing. Both questionnaires and checklists are meant to spur reflection about ethics during development and to operationalise high-level principles for ethical AI (Madaio et al., 2020). Yet, checklists, more than questionnaires, support developers as

“to do lists” to adopt during development. However, they might lack of depth as they only provide binary information (“yes” or “no” answers). In fact, Madaio et al. (2020) suggested that checklists are development guides and do not provide sufficient detailed content for a thorough documentation. Widgets, instead, are often automatically or semi-automatically derived as a side effect of the computational process (Chmielinski et al., 2022; Stoyanovich & Howe, 2019). Although they also prompt critical questions concerning the data/AI system, they focus more on producing a comprehensible output that offers a quick overview and is easily comparable.

A limited number of documentation initiatives have some level of automation (n. 6), of which 4 are in the form of widgets (Afzal et al., 2021; Anik & Bunt, 2021; Chmielinski et al., 2022; Stoyanovich & Hoew, 2019; Sun et al., 2019). Partially- or fully-automated documentation methods typically allow the user to select which widgets or visual components they want to include, and compose them to cre-

Table 1 Types of documentation methods found in our literature review

Method	Description
Questionnaire	Set of structured questions to gather information about specific aspects of an AI model/dataset. They typically require a strong documentation effort and encourage a thorough internal reflection about the use and potential misuse of datasets or AI systems
Information sheet	Grid for structured collection and presentation of information. It typically follows a concise and visual schema that summarises information on different technical and non-technical aspects. These are often referred to as “cards” or “factsheets”
Composable widgets	Modular information elements that can be freely composed to build the final documentation. A user interface or some other level of automation might be provided to fill and compose the widgets
Checklist	List of items required, things to do or points to consider that are to be systematically checked throughout the different design phases. While not conceived as documentation approaches, they can inform and guide the documentation of AI systems

ate the final document. Some components merely provide a user-friendly way to fill in information about a dataset or model in text mode. Others are data-driven, which requires the user to upload data in tabular format (e.g. dataset samples and metadata, outputs generated by a model) to automatically obtain insights about what a dataset contains or how a model behaves. Examples of facts and metrics that can be computed include: data duplicates, outliers, biases in data distributions, mislabelled data instances, cross-correlations, classification errors and confusion matrices. Data-driven components are nevertheless generally limited in terms of the type of data that can be analysed, not being able to process complex, unstructured or multimodal data formats like image, video, audio and sensor data.

Different stakeholders are expected to benefit from dataset documentation. Our analysis led to identifying nine different types of “personas”, which are presented in Fig. 4 together with their functional roles and varying documentation needs. We annotated for each approach the main personas targeted as consumers of documentation. We found that the most common targeted persona is the “AI developer”, addressed by 32 initiatives. Stakeholders belonging to this persona (e.g. engineers, researchers, data scientists, product teams, model validators, etc.) are expected to benefit from documentation in various ways: for selecting a dataset that best suits their purposes, performing more robust training of AI models, and being more aware of risks and ethical implications (see more on the objectives of documentation approaches at the end of this section).

Documentation approaches, however, are not addressed only to developers. Other stakeholders frequently targeted include the “AI auditor” (addressed in 15 of the reviewed

initiatives), the “Data governance officer” (n. 14) and the “Authority/Regulator” (n. 12). These personas have different transparency needs than developers, which depend on their specialised roles and expertise in relation with AI systems. The data governance officer is mainly interested in data management activities, being the person in charge of an organisation data strategy, including managing the life-cycle of datasets, tracking their use in different applications and defining data sharing agreements. The “AI auditor” and “Authority/Regulator”, instead, have different objectives: they need to be informed about an AI system’s features, risks and limitations in order to verify whether it performs as claimed and according to established rules, technical standards and legal requirements.

Documentation initiatives also address non-technical audiences, such as: “AI users” (n. 13), who are the final users of a system in operation and as such need to be aware of its capabilities, limitation and risks; “data or AI subjects” (n. 6), whose data is used in the system and/or who may experience effects from a model; and “society” (n. 6), which includes a broader range of societal actors ranging from the general public, to journalists or researchers. Initiatives targeted at these stakeholders share objectives with algorithmic impact assessments, which are aimed at informing the public and at engaging in productive conversations with individuals and communities about how an AI system might impact their lives or the lives of those around them (Moss et al., 2021).

The findings highlight that documentation approaches could be targeted to a wide range of stakeholders, beyond developers (Arnold et al., 2019). Considering differences in backgrounds and needs, more formats and versions could be created, which are tailored to specific types of stakehold-

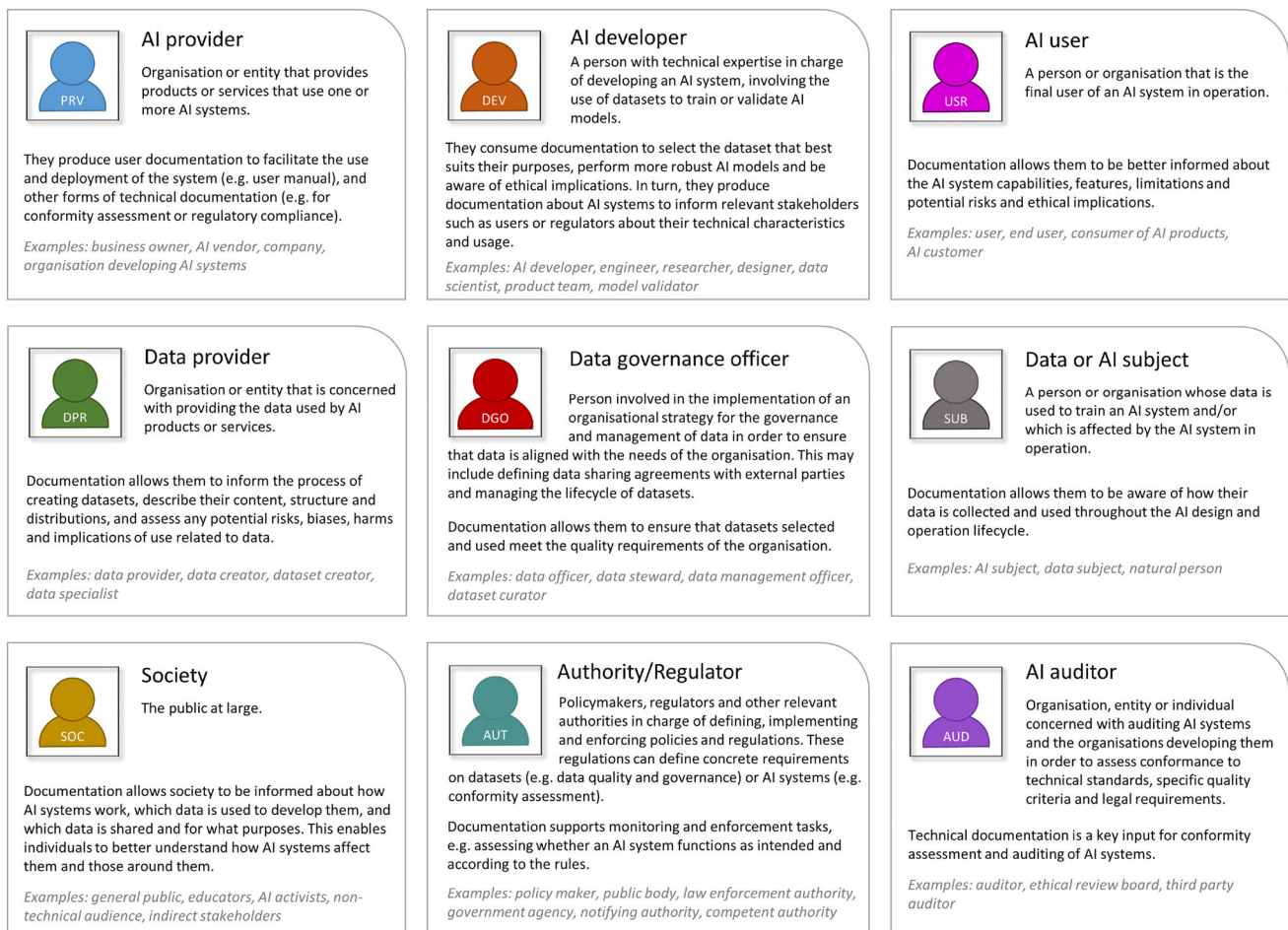


Fig. 4 Personas identified in our review, including their functional roles and documentation needs. Definitions are partly inspired by AI stakeholders in the ISO/IEC 22989:2022 standard (ISO/IEC, 2022)

ers (Anik & Bunt, 2021). For instance Bender and Friedman (2018) suggest to distinguish between “short form” and “long form” documentation. For academia, industry and government, long-form data statements should be a requirement, but for end users and other non-expert audiences, documentation approaches should not be lengthy. Mohammad (2021) proposes information sheets to be tailored to stakeholders—one version for society at large and one for researchers and developers. The former should be without technical or policy jargon and with a narrower focus on how systems can impact people and how they can contribute or push back. To include non-expert perspectives, some studies propose that engaged public, activists and communities could be involved in determining the format of documentations. Examples include the “policy-focused toolkit” based on community participation (Krafft et al., 2021), and the “ethics sheets” created through community and group efforts (Mohammad, 2021).

Documentation approaches are designed to increase transparency, but different nuances exist in the overall goals and concerns of the various initiatives. We identified six

high level concerns or objectives and classified them as: (1) ethics, (2) quality, (3) reproducibility, (4) discoverability, (5) accountability and (6) trust. An often-mentioned goal of documentation is to increase ethical awareness among developers and other stakeholders (n. 23). In such case, documentation is adopted to communicate goals, provenance, curation procedures, shortcomings and caveats of dataset dissemination, circulation and use. The main goal is to inquire whether datasets have been fairly sourced, ethically used and that they do not cause harm when used in AI applications. Several ethical issues are raised in the papers, such as fairness, discrimination, consent and data protection, responsible data collection, tagging and use of data, justice and social impact. For example, in the context of image datasets ethical issues are raised for taxonomies adopting inappropriate and/or sensitive labels or the use of images of real people without consent (Prabhu & Birhane, 2020).

Another relevant concern is accountability (n. 16), which focuses on providing evidence that a correct protocol has been followed and on communicating reliable information

to the public and decision-makers. Within this concern, there might also be the objective to support certification for regulatory compliance, although we did not find an explicit focus on compliance given the relatively early stage of AI and data regulation.

Another group of documentation approaches aims of providing detailed information to enhance reproducibility of models or studies (n. 11). This concern is mainly targeted at researchers and other developers from the machine learning community, but might apply as well to other stakeholders that require comprehensive disclosure (e.g. monitoring bodies) (Kühl et al., 2021).

A few initiatives place emphasis on the issue of quality (n. 6), addressed beyond other considerations of ethical issues, such as provenance of data or implications of its use. In these cases, the attention is placed on providing information about data attributes that can help increase the performance of an AI model, including data preparation, representativeness, correlations and coverage of the sample. Quality also extends beyond data and encompasses broader issues of efficiency, reliability and performance of an AI model or system.

Finally, a small number of initiatives emphasise also other concerns such as discoverability (n. 2), intended as enhancing discovery and access to data/model assets (e.g. finding of appropriate datasets/models for AI and facilitating sharing), or building trust (n. 3) with the general public, data subjects and AI end users.

When literature meets regulation: challenges, opportunities and recommendations

In the following we make some key considerations on how the current landscape of dataset and AI documentation approaches is developing and to what extent it addresses the transparency needs of the recent EU policy initiatives described in section “[Data and AI documentation in support of European policies and initiatives](#)” section.

In their focus, a considerable number of initiatives align with the transparency needs of risk-based EU policies—such as the AI Act and the Digital Services Act—covering AI systems used in scenarios presenting risks to the fundamental rights of individuals (e.g., Chen et al. (2022), Hupont and Gómez (2022), Krafft et al. (2021)). Documentation initiatives focusing on data for AI, AI models and AI systems could effectively support these regulations. It should be noted, however, that documentation approaches may tend to be more focused on potential risks posed by AI systems to individuals (e.g. discrimination) than larger systemic risks at societal level (e.g. amplification of disinformation), and this could represent a relevant gap to fill by the AI documentation community.

On the other hand, transparency needs of innovation-oriented EU policies, e.g. targeted towards data ecosystems for enhanced data sharing, and supporting new access rights to data, are only tangentially addressed by the initiatives reviewed. This fact was also apparent in other dimensions of our analysis, such as personas and concerns, addressed in more detail below. In general, the emergence of approaches describing datasets in a use-case agnostic manner, with the objective of facilitating discovery and reuse, would be highly positive in the context of the European strategy for data.

Another gap can also be identified in relation to technical coverage. While data and AI models are often the subject of documentation proposals, those with a focus on complete AI systems are fewer (e.g., Arnold et al. (2019), Germain Lee (2022), Madaio et al. (2020), McMillan-Major et al. (2021)). In many cases, however, AI models are just components of potentially much larger systems that integrate and orchestrate multiple software and hardware elements having an effect on outcomes as well as risks. This is relevant both in the context of risk mitigation (e.g. AI used as safety components of machinery, or AI components within complex content recommendation algorithms), but also in the context of fostering innovation, e.g. documenting datasets from connected devices to enable reuse should be inevitably linked to design and operation aspects of the overall system.

A further finding from the review is a trend towards the specialisation of documentations. The more general and horizontal nature of a first wave of proposals (e.g., DrivenData (2022), Gebru et al. (2021), Mitchell et al. (2019)) paved the way for the emergence of more specialised ones (e.g., Chaudhry et al. (2022), Rostamzadeh et al. (2022), Zheng et al. (2022)). While most of the influential papers upon which we based our search were broad in scope, more recent proposals adopt a narrower focus, either towards specific domains (such as healthcare, surveillance systems or journalism) or specific types of data sources (e.g. NLP, population data or image data). This trend is aligned with current policy needs. A focus on specific sectors is key for innovation-oriented data policies, as they have a strong sectoral component, such as the creation of European Data Spaces for specific economic sectors. Focused approaches can also support the objectives or risk-oriented regulations as they complement horizontal ones, allowing the specific needs of certain high-risk AI systems (e.g. some uses of face recognition) to be addressed.

From a regulatory perspective, the most suitable method for documenting will depend on several factors. A key element is whether a regulation has specific provisions on AI and data documentation and, in turn, whether these are voluntary or mandatory. The AI Act proposal, for instance, defines documentation requirements for AI providers including technical elements that should be included (European Commission, 2021a; Hupont et al., 2023). Rigorous questionnaire layouts, well-structured information sheets and a

set of detailed metrics captured in the form of widgets could contribute towards providing the information requested for authorities and conformity assessment bodies. The adoption of detailed checklists can also facilitate compliance assessment and auditing tasks, capturing concrete practices and technical methods adopted by AI providers and online platforms towards regulatory compliance. Finally, more concise information sheets and other visual documentation approaches could be suitable elements of instructions of use for users of AI systems.

When the goal is to spur innovation, encouraging data sharing and AI development, documentation can support both the data discovery and development stages. For the former, documentation should be easily accessible and comparable for practitioners, facilitating comparison and selection of relevant datasets for specific applications. For this goal, information sheets and widgets are probably more appropriate as they offer readily accessible information. Once suitable datasets are found, more thorough documentation approaches, e.g. based on information sheets or questionnaires, can provide the level of detail required by developers to use third-party datasets in their own AI solutions.

A key finding of the review is that documentation approaches cover the needs of a wide range of stakeholders (e.g., Geburu et al. (2021), Pushkarna et al. (2022), Tagliabue et al. (2021)). This is significant in relation to regulatory transparency needs, which concern not only developers, but also auditors, authorities, users, public sector organisations, data subjects and society at large (see Fig. 1). Unsurprisingly, however, the majority of initiatives are meant to facilitate the work of developers. Developers are an obvious recipient of dataset documentation approaches. Yet, AI documentation methods are also targeted at developer audiences, as these are often conceived as a vehicle to guide implementation tasks. Therefore, a large set of initiatives are meant to play a key role in facilitating the adoption of trustworthy AI/data sharing practices by the developer community. Their standardisation and widespread adoption could be greatly beneficial, especially for smaller players with limited resources, which represent a key policy priority in Europe, as this could contribute to levelling the playing field with larger and less resource-constrained AI providers, which have started to define and adopt own documentation practices.

The results highlight that many initiatives are also suitable for authorities and auditors (e.g., Baracaldo et al. (2022), Bender and Friedman (2018), Grasso et al. (2020), Stoyanovich and Howe (2019)), which in practice require specific information elements at a sufficient level of detail to carry out assessment tasks. The needs of authorities and auditors are less covered by current documentation approaches than those of AI developers but are still reasonably represented. These approaches have the potential to evolve into standards, serving as a basis for assessing compliance of AI products

and AI-based digital services with legal obligations. Furthermore, these could support the range of assessment modalities considered in EU regulations, including self-assessment performed by the providers themselves, conformity assessment of high-risk AI systems by notified bodies, or third-party audits of online platforms. With respect to the latter, specific approaches for the documentation of AI systems for content moderation, recommendation and searches have not been identified, and would be beneficial in the context of the Digital Services Act.

A limited number of approaches, less than one-third, specifically cater to the needs of users and societal stakeholders affected by the AI systems or members of data ecosystems. Current policies, however, demand transparency towards those that use and operate AI systems developed by others, as well as towards a broad range of societal non-expert stakeholders that may be impacted by their operation. In the case of the AI Act, the provision of information to users of high-risk AI systems is a key requirement, and some of the approaches reviewed provide a good basis for regulatory compliance. Similarly, the Data Governance Act sets the expectation of greater transparency about data sharing, especially towards society, data subjects and public sector operators. For non-expert societal stakeholders, documentation formats should be adopted that achieve higher levels of clarity, conciseness and accessibility, and preliminary work in this direction has been identified in a small subset of the approaches examined.

In terms of the concerns addressed by the documentation approaches, we observe that those more often covered in the landscape—e.g. ethical concerns and, to a lesser extent, accountability—align closely with EU policies with a risk-oriented approach. More than half of the papers address ethical concerns inquiring into the provenance of datasets and implications of the use of AI systems, and delving into how to address and mitigate biases and risks, a main objective of the AI Act proposal. Accountability is another key objective, as a number of initiatives explicitly consider the need to hold providers accountable for the functioning and impact of their datasets and AI systems. This concern will likely be more prominent in future iterations of these initiatives, especially if they take into account upcoming regulatory requirements, as well as needs arising from emerging AI auditing practices. On the other hand, far fewer initiatives cover concerns attuned to policies focused on data ecosystems, data access rights and data innovation more broadly. EU data policies in that area highlight transparency needs for increased discoverability of data assets and building trust in data sharing, which rarely belong to the concerns explicitly mentioned in the papers reviewed. Trust and discoverability, and to some extent reproducibility, are key considerations within innovation-oriented approaches as they eliminate barriers to data sharing, and foster collaboration and innovation in the

Table 2 Details on state-of-the-art data and AI documentation approaches

Short title	Year	Focus	Method	Scope	Auto.	Main personas	Main concerns
Datasheets for datasets (Gebru et al., 2021)	2018		Questionnaire		×	AI developer, Data governance officer, AI user, AI auditor, Authority/Regulator, Data or AI subject, Society	Ethics
An ethics checklist for data scientists (Driven-Data, 2022)	2018		Questionnaire		×	AI developer	Ethics, Accountability
Data statements for NLP (Bender & Friedman, 2018)	2018		Information sheet	NLP	×	AI developer, AI user, AI auditor, Authority/Regulator	Ethics, Reproducibility
The dataset nutrition label (Holland et al., 2018; Chmielinski et al., 2022)	2018		Widgets		✓	AI developer	Quality, Ethics
Nutritional labels for data and models (Stoyanovich & Howe, 2019; Sun et al., 2019)	2019		Widgets		✓	AI developer, Society, Data or AI subject, AI user, Authority/Regulator	Accountability, Reproducibility, Discoverability
FactSheets (Arnold et al., 2019; Richards et al., 2020)	2019		Questionnaire		×	AI user, AI auditor, Authority/Regulator	Accountability, Trust
Model cards for model reporting (Mitchell et al., 2019)	2019		Information sheet		✓	AI user, AI auditor, Authority/Regulator	Ethics, Accountability

Table 2 continued














Short title	Year	Focus	Method	Scope	Auto.	Main personas	Main concerns
Algorithmic accountability frameworks (Grasso et al., 2020)	2020		Questionnaire	Ecology	×	AI developer, AI auditor, Society	Ethics, Accountability
Co-designing checklists (Madaio et al., 2020)	2020	 	Checklist		×	AI developer, AI auditor	Ethics, Accountability
The LiQuID metadata model (Oppold & Herschel, 2020)	2020		Questionnaire		×	AI developer, AI auditor	Accountability, Ethics
Large image datasets (Prabhu & Birhane, 2020)	2020		Information sheet	Image data	✓	AI developer	Ethics
Ethics sheet for AI tasks (Mohammad, 2021, 2022)	2021	 	Information sheet	Affective computing	×	AI developer, AI auditor, Society, Data or AI subject	Ethics
Artsheets for art datasets (Srinivasan et al., 2021)	2021		Questionnaire	Art	×	AI developer, Society, Authority/Regulator, Data governance officer	Ethics
Checklist for crowdsourcing experiments (Ramírez et al., 2021)	2021		Checklist	Crowdsourced data	×	AI developer	Reproducibility
Data readiness report (Afzal et al., 2021)	2021		Widgets		✓	AI developer, Data governance officer	Reproducibility, Quality

Table 2 continued

Short title	Year	Focus	Method	Scope	Auto.	Main personas	Main concerns
Comprehensive and comprehensible data catalogs (Subramaniam et al., 2021)	2021		Questionnaire		×	AI developer, Data governance officer	Discoverability
Building legal datasets (Soh, 2021)	2021		Questionnaire		×	AI developer, Data governance officer	Ethics, Accountability
DAG card (Tagliabue et al., 2021)	2021		Information sheet		✓	AI developer, AI auditor	Reproducibility, Quality
Accountability for ML datasets (Hutchinson et al., 2021)	2021		Questionnaire		×	AI developer, AI auditor, Authority/Regulator, Data governance officer	Ethics, Accountability
Rigorous supervised ML (Kühl et al., 2021)	2021		Information sheet	Machine learning	×	AI developer	Reproducibility
The human evaluation datasheet 1.0 (Shimorina & Belz, 2021)	2021		Questionnaire	NLP	×	AI developer	Reproducibility
Reusable templates (McMillan-Major et al., 2021)	2021		Information sheet	NLP	×	AI developer, Data governance officer, AI user, Data or AI subject	Ethics, Reproducibility
A community library for NLP (Lhoest et al., 2021)	2021		Information sheet	NLP	×	AI developer, Data governance officer, AI user	Ethics, Reproducibility

Table 2 continued












Short title	Year	Focus	Method	Scope	Auto.	Main personas	Main concerns
Data-centric explanations (Anik & Bunt, 2021)	2021		Widgets	Population data	×	AI developer, AI user	Ethics, Accountability
Action-oriented AI policy toolkit (Krafft et al., 2021)	2021	 	Questionnaire	Surveillance	×	Society, Data or AI subject	Ethics, Trust
FairEd (Vasquez Verdugo et al., 2022)	2022		Information sheet	Education	×	AI developer, Data governance officer	Ethics
Federated learning: A FactSheets approach (Baracaldo et al., 2022)	2022	 	Information sheet	Federated learning	×	AI provider, AI developer, AI auditor, Authority/Regulator	Accountability, Trust
Healthsheet (Ros-tamzadeh et al., 2022)	2022		Questionnaire	Health	×	AI developer, Data governance officer	Ethics
Responsible data and algorithmic practices in the news media (Showkat, 2022)	2022	 	Questionnaire	Journalism	×	AI Developer, AI user, Data governance officer	Accountability, Reproducibility
Data cards (Pushkarna et al., 2022)	2022		Information sheet		×	AI developer, Authority/Regulator, AI auditor, Data or AI subject, Data governance officer	Ethics, Accountability

Table 2 continued

Short title	Year	Focus	Method	Scope	Auto.	Main personas	Main concerns
Method cards (Adkins et al., 2022a, 2022b)	2022	AI	Information sheet	Hor	×	AI developer	Ethics, Quality, Reproducibility
Network report (Zheng et al., 2022)	2022	D	Information sheet	Network data	×	AI developer	Quality
Transparency index framework for AI in education (Chaudhry et al., 2022)	2022	D AI	Questionnaire	Education tools	×	AI developer, AI user, Data governance officer, Authority/Regulator, AI auditor	Ethics, Quality, Accountability
Documenting affective computing use cases (Hupont & Gómez, 2022)	2022	D AI	Information sheet	Affective computing	×	AI developer, AI user, Authority/Regulator, AI auditor	Ethics, Accountability
Collections as ML data (Germain Lee, 2022)	2022	D AI	Questionnaire	Cultural heritage	×	AI developer, AI user, Data governance officer	Ethics
Transparency and standardization in ophthalmologic AI (Chen et al., 2022)	2022	AI	Information sheet	Ophthalmology	×	AI user, Authority/Regulator, AI auditor	Ethics, Accountability

field of AI, with the potential to unlock massive economic and societal benefits. Although only a very small number of the reviewed documentation approaches explicitly mention “trust” as the main concern addressed (n. 3), the overall and most widespread objectives are attuned with EU’s push for trustworthy AI (European Commission, 2019, 2021a), as they deal with tackling ethical concerns and promoting accountability.

Conclusions and future work

This article presents a overview of 36 state-of-the-art documentation approaches for data, AI models and AI systems identified in the scientific literature. We have systematically analysed these approaches using a specific set of criteria (focus, scope, methods, personas, concerns) with the goal to increase understanding of the existing landscape and discuss its coverage of transparency-related needs of European policy initiatives for trustworthy AI and data-driven innovation.

Our review shows that a majority of initiatives align with the transparency demands of current risk-oriented approaches in EU policy, especially with precepts contained the draft AI Act. Nevertheless, we identified three main gaps when it comes to covering current regulatory needs. First, documentation for data innovation purposes (e.g. through the sharing and reuse of data as promoted in the European Strategy for Data) is not prominently represented within the approaches found. Second, most documentation methodologies focus on data and AI models, not covering important aspects related to the AI system as a whole (e.g. interaction and oversight mechanisms provided to the user, hardware and software architectures on which the system runs, operational context of use), which is at the core of regulations such as the AI Act. Third, there are no documentation methodologies to date tailored or adapted to the specific characteristics of large-scale algorithmic systems such as those used in online platforms. Furthermore, these are often associated with risks of a systemic nature, e.g. the amplification of disinformation, which deserve specific consideration. Therefore, the definition of documentation approaches supporting their transparency would be beneficial and useful in light of the Digital Services Act.

Beyond these trends, it is worth observing that the number of proposals for documentation is constantly increasing and this article provides a snapshot of a dynamic ecosystem. For instance, we observe that new approaches with a narrow focus constantly emerge, e.g. dedicated to specific sectors or types of data and AI systems. This resonates well with current regulations that, although being largely horizontal, envision key

application sectors, either due to their economic relevance or to their potential risks to fundamental rights.

This research has some limitations and future areas of work can already be identified. The findings of the study are determined by the cluster of six papers adopted as initial seeds to search for the literature. This strategy was devised with the objective to cast a wide net that includes as many relevant references in this rapidly emerging field as possible. However, it is likely that certain strands of work are not fully represented in the review, such as those concerning GDPR documentation for data subjects or the the adoption of FAIR principles in data management. A number of these approaches are already seeing early adoption by AI practitioners, and future studies could assess whether adoption is leading to increased maturity and eventual standardisation. Another relevant question for future work concerns producers of documentation. In this work we have only examined intended audiences, yet a discussion on effective processes for documentation production and management deserves attention (e.g. Tahaei et al. (2021)). It is expected that various stakeholders with different skill sets should be involved in the production of effective AI documentation artefacts (Ibáñez & Olmeda, 2021). Further work should also zoom in on a selected subset of approaches, focusing on those that provide the most comprehensive coverage of regulatory needs. Analysis could focus on how these approaches are being implemented by practitioners, providing guidance aiming to facilitate consistent adoption, ensuring that the level of detail, relevance and accuracy of the information elements documented are fit for purpose. Future research could also investigate the effectiveness of these approaches for specific stakeholders. This could involve, for instance, conducting user studies to characterise how dataset and AI information is interpreted and leveraged by non-expert audiences.

Acknowledgements The authors would like to thank their colleagues and coordinators in the Digital Economy Unit of the Joint Research Centre for the fruitful discussions and their support.

Disclaimer The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

Funding This research received no specific grants or funding.

Data availability The data generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations





Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as

long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Comprehensive list of current documentation approaches for data and AI

Table 2 compiles comprehensive information about the final 36 state-of-the-art documentation approaches identified in this work. For each approach, we provide the following information:

- **Short title:** short title assigned to the paper, from original ones.
- **Year:** year of publication of the paper.
- **Focus:** its main focus, which can be data , AI model  or AI system .
- **Method:** the type of method it proposes, from the ones listed in Table 1.
- **Scope:** domain for which it has been conceived, where  indicates horizontal approaches.
- **Automation:** whether it offers some level of automation (✓) or not (×).
- **Main personas:** main type of personas (i.e. stakeholders) targeted as consumers of the produced documentation, from the ones in Fig. 4.
- **Main concerns:** main concerns addressed in the approach.

References

- Adkins, D., Alsallakh, B., Cheema, A., Kokhlikyan, N., McReynolds, E., Mishra, P., Procope, C., Sawruk, J., Wang, E., & Zvyagina, P. (2022a). Method cards for prescriptive machine-learning transparency. In *2022 IEEE/ACM 1st international conference on AI engineering—software engineering for AI (CAIN)* (pp. 90–100).
- Adkins, D., Alsallakh, B., Cheema, A., Kokhlikyan, N., McReynolds, E., Mishra, P., Procope, C., Sawruk, J., Wang, E., & Zvyagina, P. (2022b). Prescriptive and descriptive approaches to machine-learning transparency. In *CHI conference on human factors in computing systems extended abstracts* (pp. 1–9).
- Afzal, S., Rajmohan, C., Kesarwani, M., Mehta, S., & Patel, H. (2021). Data readiness report. In *2021 IEEE international conference on smart data services (SMDS)* (pp. 42–51). IEEE.
- Amith, M. T., Cui, L., Zhi, D., Roberts, K., Jiang, X., Li, F., Yu, E., & Tao, C. (2022). Toward a standard formal semantic representation of the model card report. *BMC Bioinformatics*, *23*(6), 1–18.
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989.
- Anik, A. I., & Bunt, A. (2021). Data-centric explanations: Explaining training data of machine learning systems to promote transparency. In *CHI conference on human factors in computing systems* (pp. 1–13).
- Arnold, M., Bellamy, R. K., Hind, M., Houde, S., Mehta, S., Mojsilović, A., Nair, R., Ramamurthy, K. N., Olteanu, A., & Piorkowski, D. (2019). AI FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development*, *63*(4–5), 1–6.
- Baracaldo, N., Anwar, A., Purcell, M., Rawat, A., Sinn, M., Altkrouri, B., Balta, D., Sellami, M., Kuhn, P., & Schopp, U. (2022). Towards an accountable and reproducible federated learning: A factsheets approach. [arXiv:2202.12443](https://arxiv.org/abs/2202.12443)
- Bäuerle, A., Cabrera, Á. A., Hohman, F., Maher, M., Koski, D., Suau, X., Barik, T., & Moritz, D. (2022). Symphony: Composing interactive interfaces for machine learning. In *CHI conference on human factors in computing systems* (pp. 1–14).
- Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, *6*, 587–604.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.
- Caton, S., & Haas, C. (2020). Fairness in machine learning: A survey. [arXiv:2010.04053](https://arxiv.org/abs/2010.04053)
- Chaudhry, M. A., Cukurova, M., & Luckin, R. (2022). A transparency index framework for AI in education. [arXiv:2206.03220](https://arxiv.org/abs/2206.03220)
- Chen, D. K., Modi, Y., & Al-Aswad, L. A. (2022). Promoting transparency and standardization in ophthalmologic artificial intelligence: A call for artificial intelligence model card.
- Chmielinski, K. S., Newman, S., Taylor, M., Joseph, J., Thomas, K., Yurkofsky, J., & Qiu, Y. C. (2022). The dataset nutrition label (2nd gen): Leveraging context to mitigate harms in artificial intelligence. [arXiv:2201.03954](https://arxiv.org/abs/2201.03954)
- Demajo, L. M., Vella, V., & Dingli, A. (2020). Explainable ai for interpretable credit scoring. [arXiv:2012.03749](https://arxiv.org/abs/2012.03749)
- Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., 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. [arXiv:2305.02231](https://arxiv.org/abs/2305.02231)
- DrivenData. (2022). An ethics checklist for data scientists. <https://deon.drivendata.org/>
- Edwards, L., & Veale, M. (2018). Enslaving the algorithm: From a “right to an explanation” to a “right to better decisions”? *IEEE Security & Privacy*, *16*(3), 46–54.
- European Commission. (2019). Ethics guidelines for trustworthy AI. <https://ec.europa.eu/digital-single-market/en/news/ethicsguidelines-trustworthy-ai>
- European Commission. (2021a). The Digital Europe Programme. <https://digital-strategy.ec.europa.eu/en/activities/digital-programme>
- European Commission. (2021b). Proposal for a Regulation on Artificial Intelligence. <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>
- European Commission. (2022a). Proposal for a Directive on adapting non contractual civil liability rules to artificial intelligence. https://ec.europa.eu/info/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en
- European Commission. (2022b). Proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending

- Directive 2000/31/EC. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0825>
- European Commission. (2022c). Proposal on a Regulation on European data governance (Data Governance Act). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767>
- European Commission. (2022d). Proposal on a Regulation on harmonised rules on fair access to and use of data (Data Act). <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2021:559:FIN>
- European Parliamentary Research Service. (2022). Governing data and artificial intelligence for all. Models for sustainable and just data governance. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU\(2022\)729533](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729533)
- Falco, G., Shneiderman, B., Badger, J., Carrier, R., Dahbura, A., Danks, D., Eling, M., Goodloe, A., Gupta, J., & Hart, C. (2021). Governing AI safety through independent audits. *Nature Machine Intelligence*, 3(7), 566–571.
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92.
- Germain Lee, B. C. (2022). The “collections as ML data” checklist for machine learning & cultural heritage.
- Grasso, I., Russell, D., Matthews, A., Matthews, J., & Record, N. R. (2020). Applying algorithmic accountability frameworks with domain-specific codes of ethics: A case study in ecosystem forecasting for shellfish toxicity in the gulf of maine. In *ACM-IMS on foundations of data science conference* (pp. 83–91).
- Hansen, H. K., & Flyverbom, M. (2015). The politics of transparency and the calibration of knowledge in the digital age. *Organization*, 22(6), 872–889.
- Holland, S., Hosny, A., Newman, S., Joseph, J., & Chmielinski, K. (2018). The dataset nutrition label: A framework to drive higher data quality standards. [arXiv:1805.03677](https://arxiv.org/abs/1805.03677)
- Hupont, I., & Fernández, C. (2019). Demopairs: Quantifying the impact of demographic imbalance in deep face recognition. In *2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019)* (pp. 1–7). IEEE.
- Hupont, I., Fernández-Llorca, D., Baldassarri, S., & Gómez, E. (2023a). Use case cards: A use case reporting framework inspired by the european AI act. [arXiv:2306.13701](https://arxiv.org/abs/2306.13701)
- Hupont, I., & Gómez, E. (2022). Documenting use cases in the affective computing domain using unified modeling language. In *2022 10th international conference on affective computing and intelligent interaction (ACII)* (pp. 1–8). IEEE.
- Hupont, I., Micheli, M., Delipetrev, B., Gómez, E., & Garrido, J. S. (2023). Documenting high-risk AI: A European regulatory perspective. *Computer*, 56(5), 18–27.
- Hupont, I., Tolan, S., Gunes, H., & Gómez, E. (2022). The landscape of facial processing applications in the context of the European AI act and the development of trustworthy systems. *Scientific Reports*, 12(1), 10688.
- Hutchinson, B., Smart, A., Hanna, A., Denton, E., Greer, C., Kjartansson, O., Barnes, P., & Mitchell, M. (2021). Towards accountability for machine learning datasets: Practices from software engineering and infrastructure. In *ACM conference on fairness, accountability, and transparency* (pp. 560–575).
- Ibáñez, J. C., & Olmeda, M. V. (2021). Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study. *AI & Society*, 1–25.
- ISO/IEC.D (2022). ISO/IEC 22989:2022 standard on Information Technology—Artificial Intelligence—Artificial Intelligence concepts and terminology.
- Krafft, P., Young, M., Katell, M., Lee, J. E., Narayan, S., Epstein, M., Dailey, D., Herman, B., Tam, A., & Guetler, V. (2021). An action-oriented AI policy toolkit for technology audits by community advocates and activists. In *ACM Conference on Fairness, accountability, and transparency* (pp. 772–781).
- Kühl, N., Hirt, R., Baier, L., Schmitz, B., & Satzger, G. (2021). How to conduct rigorous supervised machine learning in information systems research: The supervised machine learning report card. *Communications of the Association for Information Systems*, 48(1), 46.
- Lhoest, Q., del Moral, A. V., Jernite, Y., Thakur, A., von Platen, P., Patil, S., Chaumond, J., Drame, M., Plu, J., & Tunstall, L., ... others (2021). Datasets: A community library for natural language processing. [arXiv:2109.02846](https://arxiv.org/abs/2109.02846)
- Madaio, M. A., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020). Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In *CHI conference on human factors in computing systems* (pp. 1–14).
- McMillan-Major, A., Osei, S., Rodriguez, J. D., Ammanamanchi, P. S., Gehrmann, S., & Jernite, Y. (2021). Reusable templates and guides for documenting datasets and models for natural language processing and generation: A case study of the HuggingFace and GEM data and model cards. [arXiv:2108.07374](https://arxiv.org/abs/2108.07374)
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Conference on fairness, accountability, and transparency* (pp. 220–229).
- Mohammad, S. M. (2021). Ethics sheets for AI tasks. [arXiv:2107.01183](https://arxiv.org/abs/2107.01183)
- Mohammad, S. M. (2022). Ethics sheet for automatic emotion recognition and sentiment analysis. *Computational Linguistics*, 48(2), 239–278.
- Moss, E., Watkins, E. A., Singh, R., Elish, M. C., & Metcalf, J. (2021). Assembling accountability: Algorithmic impact assessment for the public interest. Technical report.
- Naja, I., Markovic, M., Edwards, P., Pang, W., Cottrill, C., & Williams, R. (2022). Using knowledge graphs to unlock practical collection, integration, and audit of AI accountability information. *IEEE Access*, 10, 74383–74411.
- OECD. (2022). OECD Framework for Classification of AI Systems: A tool for effective AI policies. <https://oecd.ai/en/classification>
- Oppold, S., & Herschel, M. (2020). Accountable data analytics start with accountable data: The liquid metadata model. In *ER Forum/Posters/Demos* (pp. 59–72).
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 1–11.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Prabhu, V. U., & Birhane, A. (2020). Large image datasets: A pyrrhic win for computer vision? [arXiv:2006.16923](https://arxiv.org/abs/2006.16923)
- Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data cards: Purposeful and transparent dataset documentation for responsible AI. [arXiv:2204.01075](https://arxiv.org/abs/2204.01075)
- Ramírez, J., Sayin, B., Baez, M., Casati, F., Cernuzzi, L., Benatallah, B., & Demartini, G. (2021). On the state of reporting in crowdsourcing experiments and a checklist to aid current practices. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–34.
- Richards, J., Piorkowski, D., Hind, M., Houde, S., & Mojsilović, A. (2020). A methodology for creating AI factsheets. [arXiv:2006.13796](https://arxiv.org/abs/2006.13796)
- Rostamzadeh, N., Mincu, D., Roy, S., Smart, A., Wilcox, L., Pushkarna, M., Schrouff, J., Amironesei, R., Moorosi, N., & Heller, K. (2022). Healthsheet: Development of a transparency artifact for health datasets. [arXiv:2202.13028](https://arxiv.org/abs/2202.13028)
- Shimorina, A., & Belz, A. (2021). The human evaluation datasheet 1.0: A template for recording details of human evaluation experiments in nlp. [arXiv:2103.09710](https://arxiv.org/abs/2103.09710)

- Showkat, D. (2022). Supporting responsible data and algorithmic practices in the news media. In *6th HUMANIZE Workshop*.
- Soh, J. (2021). Building legal datasets. [arXiv:2111.02034](https://arxiv.org/abs/2111.02034)
- Srinivasan, R., Denton, E., Famularo, J., Rostamzadeh, N., Diaz, F., & Coleman, B. (2021). Artsheets for art datasets. In *35th conference on neural information processing systems datasets and benchmarks track*.
- Stoyanovich, J., & Howe, B. (2019). Nutritional labels for data and models. *A Quarterly bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering*, 42(3).
- Subramaniam, P., Ma, Y., Li, C., Mohanty, I., & Fernandez, R. C. (2021). Comprehensive and comprehensible data catalogs: The what, who, where, when, why, and how of metadata management. [arXiv:2103.07532](https://arxiv.org/abs/2103.07532)
- Sun, C., Asudeh, A., Jagadish, H., Howe, B., & Stoyanovich, J. (2019). Mithralabel: Flexible dataset nutritional labels for responsible data science. In *28th ACM international conference on information and knowledge management* (pp. 2893–2896).
- Tagliabue, J., Tuulos, V., Greco, C., & Dave, V. (2021). Dag card is the new model card. [arXiv:2110.13601](https://arxiv.org/abs/2110.13601)
- Tahaei, M., Frik, A., & Vaniea, K. (2021). Privacy champions in software teams: Understanding their motivations, strategies, and challenges. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–15).
- Tolan, S., Miron, M., Gómez, E., & Castillo, C. (2019). Why machine learning may lead to unfairness: Evidence from risk assessment for juvenile justice in catalonia. In *Proceedings of the seventeenth international conference on artificial intelligence and law* (pp. 83–92).
- Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). The ethics of algorithms: Key problems and solutions. *AI & Society*, 37(1), 215–230.
- Vasquez Verdugo, J., Gitiaux, X., Ortega, C., & Rangwala, H. (2022). Faired: A systematic fairness analysis approach applied in a higher educational context. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 271–281).
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., & Bourne, P. E. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 1–9.
- Zheng, X., Rossi, R. A., Ahmed, N., & Moritz, D. (2022). Network report: A structured description for network datasets. [arXiv:2206.03635](https://arxiv.org/abs/2206.03635)