Centering Knowledge Along the Responsible LLM Supply Chain: An Empirical Study & Multi-Stakeholder Taxonomy [Supplementary Material]

AGATHE BALAYN, Microsoft Research, USA
FANNY RANCOURT, ServiceNow, Canada
FABIO CASATI, ServiceNow, Switzerland
UJWAL GADIRAJU, Delft University of Technology, The Netherlands

ACM Reference Format:
Agathe Balayn, Fanny Rancourt, Fabio Casati, and Ujwal Gadiraju. 2026. Centering Knowledge Along the Responsible LLM Supply Chain: An Empirical Study & Multi-Stakeholder Taxonomy [Supplementary Material]. *Proc. ACM Hum.-Comput. Interact.* 10, 6, Article CSCW061 (October 2026), 21 pages. https://doi.org/10.1145/3816909

1 DETAILS ABOUT OUR METHODOLOGY

1.1 Semi-structured interviews

Participant recruitment. The interviews happened between October 2023 and January 2024. Participation was voluntary and participants were motivated by the idea of reflecting on their own practices. Before participating in the interviews, the participants were asked to sign a consent form approved by the ethics committee of one of our organizations. While we aimed at interviewing a representative set of stakeholders involved in an LLM supply chain, we could however not be entirely comprehensive, and scoped out stakeholders whose role does not consist directly in envisioning, designing, developing, deploying, or using components of the LLM system. These are for instance employees of provider and deployer organizations working in human resources or accounting departments, or stakeholders outside the AI supply chain who might indirectly impact it (e.g., by influencing the requirements for the AI system) such as external auditors, academic AI researchers, and the public.

Interview participants. The official roles of our participants range from engineer, software engineering manager, to product manager, product owner, or platform leader, to researcher, innovation leader, or chief scientist, to UX designer or researcher, to content designer, etc. The hierarchical levels of our participants vary, from junior, to more senior employees, to managers. The practical tasks of our participants included implementing, testing, and overseeing LLM systems, or strategizing around the requirements of the system and its development (e.g., functional requirements, definition of risks, etc.). The participants span different roles including executives, directors, and individual contributors at different levels of seniority, and different amounts and types of experience with AI.

Authors' Contact Information: Agathe Balayn, Microsoft Research, USA, balaynagathe@microsoft.com; Fanny Rancourt, fanny.rancourt@servicenow.com, ServiceNow, Canada; Fabio Casati, ServiceNow, Switzerland, fabio.casati@servicenow.com; Ujwal Gadiraju, Delft University of Technology, The Netherlands, u.k.gadiraju@tudelft.nl.

ACM 2573-0142/2026/10-ARTCSCW061
https://doi.org/10.1145/3816909

Interview protocol. While two-thirds of the semi-structured interviews lasted on average 45 minutes, for the other interviews the participants did not have as much time to give us. In this case, before the interviews, the participants were asked to answer a questionnaire describing their perceptions of AI and their understanding of AI transparency, explainability, and literacy, and only the remaining questions were asked during the interviews (or further probing questions). The interviews were recorded, interview transcripts were created and pseudo-anonymized, and the recordings were then destroyed.

Additional sources of information. While the 71 interviews were our primary source of information about knowledge needs and practices during the study, we complemented them with additional sources of information. Particularly, we gathered additional insights by investigating grey literature and observing documentation efforts in a few organizations. We asked our interview participants to share 1) any example of documentation they might frequently use. In addition, some authors of this paper are partially affiliated with some of the participants' organizations, and took this opportunity to 2) directly observe the practices within some of these organizations.[1] They investigated the types of documentation shared across employees (e.g., emails from different teams and departments, within or across-department channels of conversation). They also observed cross-functional and cross-organizational meetings discussing LLM system needs, extracting knowledge nuggets relevant to these conversations. 3) We followed the work of two teams whose goal was to develop documentation practices for several LLMs and LLM systems. We observed six sessions they conducted with employees of several organizations, where they discussed LLM-related knowledge nuggets these employees would like to see logged in AI documentation and the knowledge purposes thereof. This informed us about additional knowledge needs across the LLM supply chain.

1.2 Data analysis

Scoping knowledge nuggets. When analyzing the interview transcripts, in this work, we only accounted for knowledge nuggets that revolve around AI, AI systems, and their components. Across our study, we naturally also identified knowledge nuggets that are necessary to the stakeholders of the LLM supply chain, yet that are not specific to AI –they are not the object of our study and we excluded them from our LLM knowledge taxonomy. These are for instance the production knowledge that every stakeholder in the supply chain needs to conduct their respective work, even when they do not focus on AI, e.g., a UX researcher needs expertise in qualitative and quantitative user studies, an LLM developer needs experience coding, a legal actor requires an understanding of the legal system in their environment. Knowing how to work in a team or across organizations is also an expertise important across the LLM supply chain but not specific to AI.

Developing the knowledge taxonomy. The first and second authors took part in all analysis activities to develop the taxonomy, and the other authors participated in the discussions to create the knowledge dimensions. Note that these activities did not purely follow a sequential order: the authors annotated the first ten interviews and developed early themes before conducting the next interviews and coding exercises, and repeated this process twice, reconciling the annotations across all analyzed interviews. While the first thirty five interviews were systematically analyzed, we relied on memos written during the next interviews to identify new types of knowledge nuggets or potential relevant knowledge dimensions to update the taxonomy. For instance, in this way we added considerations about "socio-organizational" knowledge nuggets that had not been pointed out in prior interviews.

[1]Note that, in this work, statements from sampled participants do not necessarily reflect the positions of the organizations associated with the authors.

Positionality statement. The analysis of the data is influenced by the positionality of the authors of this paper. We all work in European and North American institutions, both in academia and industry. We all have prior experience with developing AI systems and conducting responsible AI research, and most of us have spent at least six months working within LLM provider organizations. We were trained in Computer Science, Artificial Intelligence, and Human-Computer Interaction. These experiences shaped our interest in knowledge as a central resource in responsible AI development, and sensitized us to the practical constraints, incentives, and asymmetries faced by different actors in LLM supply chains. Our positionality enabled us to build rapport with participants, ask informed follow-up questions, and recognize tacit forms of knowledge exchange that might be overlooked by outsiders. We view this positionality as a lens that shaped and constrained our analysis. To avoid biases as much as possible, we adopted reflexive and collaborative analysis practices to surface and question assumptions derived from our own experiences.

1.3 Validation of the taxonomy

We rigorously surveyed Google Scholar (conducting snowball sampling) for publications that use keywords related to AI or machine learning, as well as keywords related to AI transparency, explainability, or literacy. We retained publications that conduct user studies (e.g., [25]) or deal with new tools or new algorithms (e.g., [5, 34]) in relation to these notions, or discuss laws or policies or propose a conceptual understanding thereof (e.g., [4, 90]). We only selected publications that include some description of knowledge nuggets, knowledge purposes, and / or knowledge stakeholders to ensure that we can reflect on our LLM knowledge framework based on these publications. We do not include papers that talk about AI transparency or replicability in the context of AI research (e.g., research on new algorithms, on new applications of AI) [37] but only retain papers focusing on AI used nowadays in organizations. As the goal of our work is not to conduct a literature survey around AI transparency, explainability, or literacy, we also excluded several papers to remain within a tractable number of publications, while ensuring that the selected publications are representative of the landscape of the literature. For that, as we noticed that plenty of publications propose new algorithms for AI explainability for decision-makers, we decided to choose only a few of those that are highly cited. We also excluded publications on AI literacy when these address the literacy of stakeholders that are not relevant to our study, such as publications that discuss the literacy of individuals younger than a university education (e.g., young children), and of individuals evolving in specific application areas outside our LLM supply chain (e.g., AI literacy for journalists). Table 1 lists all the publications we analyzed.

Table 1. List of publications surveyed in this paper.

Publication type	Publication list	Total
Policy study	[2, 4, 9, 13, 14, 23, 24, 28, 31, 33, 38, 45, 50–52, 54, 58, 64, 70, 71, 83, 85, 86, 89, 90, 92, 93]	27
HCI tool	[5–7, 22, 34, 41, 42, 62, 76, 87, 96]	11
HCI study	[8, 10, 11, 15, 16, 18–21, 25–27, 29, 30, 39, 40, 43, 44, 46–49, 55–57, 59, 60, 63, 65, 66, 69, 72, 74, 78–81, 88, 91, 94, 95]	41
Policy & regulation	[3, 61, 67, 68, 73, 75, 82]	7
Company AI principles	[1, 12, 17, 32, 35, 36, 77]	7

Note that although we built our taxonomy based on an empirical study of LLM supply chains, we compare it to publications whose primary target is not necessarily LLM systems but AI systems which rely on discriminative algorithms. We do so for two reasons. There is not enough literature that addresses questions of knowledge in LLM systems specifically yet. There is a reasonable

overlap in stakeholder roles across the two types of technologies, conducting similar work and consequently having similar knowledge needs. Despite certain differences, LLMs and prior AI systems share various technical principles and regulations, many harms they can cause are similar and the related responsible AI notions to prevent them rely on similar ideas. While it is therefore natural to find differences in the knowledge nuggets and dimensions covered, the amount of overlap identified also hints at the generalizability of our taxonomy.

2 EXPANSION OF OUR QUALITATIVE FINDINGS

2.1 Insights about Knowledge Practices and Challenges

2.1.1 Limited Knowledge Exchanges across the Supply Chain. We also found highly-consequential, missed, opportunities for fulfilling the knowledge gaps.

Missed opportunities. Our interviews revealed that knowledge exchanges occur throughout the supply chain—within individual teams, across teams within the same organization, and even between organizations. However, while relevant knowledge exists within the supply chain, participants also described instances where this knowledge was not communicated at all. For example, P2 (responsible AI researcher) reflected: *"Sometimes I learn way after the fact that the team has integrated something about responsible AI in their model development, that is completely different from what we suggest. I ask 'why did you do it that way?' and they say 'we had to do something and we didn't know that you were doing something.' There's a lack of communication. There are so many teams working on so many things, and they don't talk to each other and don't know that others have already tried solving this or done something to alleviate something."* In other cases, although knowledge exchanges occurred, the information was not always interpreted correctly, leading to flawed decisions. This often stemmed from differences in expertise and terminology among actors in the supply chain, which created opportunities for misunderstanding. P24 (developer) reflected on this challenge: *"We had a discussion recently where our designers wanted to show where the hallucination was in the paragraph. But they didn't think it was possible. And us as data scientists, we thought it would be much more valuable to show the hallucination per sentence. But then our PM, who was the bridge in the middle, told us that designers cannot do that in the UI. And then the designers thought it was not possible technically to do that. It was an example of: we need to be so much more in sync, so that designers can know what's possible to do, and that we can know as data scientists also what is possible to do from an interface perspective, so that we can meet in the middle."*

Knowledge gaps as barriers to responsible deployment. Many participants highlighted how having incomplete, unmet, or wrongly met, knowledge needs led to different consequences depending on the role of the practitioner. Customers of LLM providers and end-users may become disappointed if they lack an accurate mental model of what LLMs can and cannot do. Besides, users may unintentionally misuse systems when they are unaware of what the LLM was trained for: (P13) *"We started putting together model cards for customers to understand the use cases [for the LLM system]. Then, the opportunity for misuse goes down. Even for the well intended customers, if you don't tell them, how would they know? We currently tell them 'please use it only for the following'."* As for upstream practitioners, knowledge gaps not only led teams to duplicate and waste efforts, but they also led individuals to make inappropriate design and deployment decisions. For instance, one noted, (P24) *"Sometimes we have executives, PMs or designers that don't have a lot of experience in AI. And if you have data scientists who are not businessy, you have a clash. PMs ask a question to the data scientists and don't get the right answer. So you have bad decision taken and responsible AI is in the cracks. Because PMs don't know, and the data scientists don't understand what was the intent behind the question and don't realize what they could say that would make the decision better."*

2.2 Illustration of the Knowledge Taxonomy

In Tables 2, 3, 4, we illustrate with empirical findings (from our interviews) the various knowledge categories we identified as important for the stakeholders along the LLM supply chain.

Table 2. Examples of knowledge nuggets for the specific knowledge.

Knowledge category	Knowledge nuggets	Illustrative quote
Specific technology		
Technical artifact	Artifact description, e.g., data flow.	*"A great degree of transparency in how features are implemented at the level that we can say, if you do this in the user interface, here is what the data flow looks like, would matter a lot in our adoption processes. Here's how the data flows. Here's where your data is used."*
Socio-technical artifact	Artifact quality, e.g., algorithmic bias related to guardrails	*"You should show that your algorithm is behaving appropriately, without bias."* (P19)
Interactional artifact	Artifact description, e.g., documentation in the user interface	*"Algorithmic transparency is telling customers the algorithms used and their limitations."* (P7)
Organizational artifact	Artifact description, e.g., insurance of good use of LLM outputs	*"We tell our employees 'here's the ethical guidelines for AI at [organization]'. Read and attest to them to use these tools. Although people can still do bad things, it's a control that allows us to protect the bank if something bad happens from an employee perspective."* (P57)
Specific production		
Technical artifact	Process, e.g., data creation	*"[A-org]'s responsibility would be for the technical working. So the processing of that data, the transport of that data, and where the learning is done."* (P45)
Socio-technical artifact	Process, e.g., artifact development for the guardrails	*"We also developed our own hallucination detection as a post-processing step."* (P21)
Interactional artifact	Process, e.g., system deployment	*"I would like to see team leaders be able to say that they've actually tried the AI out with the customers first. And to see how the call went, and how helpful it actually was. Before actually deploying the AI more broadly."* (P60)
Organizational artifact	Actors and governance	*"There's so many different organizations in the company, too, like we're not even aware of who knows or doesn't know what what we're working on."* (P3)
Specific use		
Technical artifact	Online usage of specific models	*"What models are being used. For example, for this feature, are you using a [A-org] model that's running in your data center? Are you using external provider model? "* (P33)
Social-technical artifact	Purpose of explanation artifacts	*"Explainability will be something that we could provide per inference in case you're curious to know how this specific inference was generated."* (P7)
Interactional artifact	Online usage of the entire system	*"We provide them with a product intended to work in a specific way, [..] they decide to customize that, or bypass some of the guardrails."* (P30)
Organizational lens	Application of organizational practices, e.g., training	*"If **we educate them appropriately**, but then they decide to customize that, or bypass some of the guardrails, then it should be their responsibility."* (P30)
Specific context		
Technical lens	Requirements	*"In your requirements management system, you write: as a user I want to do X, Y and Z. The best practice that emerged is to provide some acceptance criteria and context, so that the person going to develop the requirement has not only the use-case and the user story."* (P18)
Socio-technical lens	Usage policies of the LLM system	*"The way that our platform works is that you could change things. And that's where the customer responsibility could be involved. We provide them with a product intended to work in a specific way. If they decide to customize that, then it should be their responsibility."* (P30)
Interactional lens	Customer needs	*"We do discovery to understand what customers are looking for, and their team structures and their needs. And then concept testing to see their thoughts."* (P6)
Organizational lens	Organizations' resources	*"You have managers who assign resources to a use-case. Then you have team leads who are specifying which resource should do what, especially for engineers."* (P4)

Table 3. Examples of knowledge nuggets for the abstract knowledge.

Knowledge category	Knowledge nugget	Illustrative quote
Generalized technology		
Technical	Basics of LLM functioning	LLM training: *"At the back-end like fine-tuning algorithms, I don't understand much. I'd be fully transparent here. I have still more learning to do."* (P5)
	Basic LLM artifacts	Functioning of generative AI: *"They say a lot we need to educate people, and they do these courses. They recently did a generative AI basic course. This resonates with me a lot, the goal to educate as many people as possible about how AI works."* (P64)
	Theories that explain LLM systems' (limited) performance and their bounds	Training "rules" *"Is it better to first have a very good model, and then increase bits by bits its capabilities to make it aligned and helpful, or is it better on the contrary to first put some guard rails and train within that limited bound? I'm not sure we know how to do the later one in research."* (P32)
Socio-technical	Responsible AI concepts	Explainability: *"At least when they're simple, you have the advantage that you can understand what the model is doing. It gets removed as soon as you move to a little bit more complex model like boosted trees, this family of models is already harder to understand, and neural networks are even more opaque."* (P4)
Interactional	Human-AI collaboration concepts, e.g., over-reliance	*"when it comes to user research. There can be a tendency of people to sort of over-rely on AI, just assume that whatever AI says -it must be true, because it's this magic thing."* (P37)
Organizational	Governance frameworks	*"We rely on their expertise to establish some sort of governance on the technology and how to bring in AI in the organization itself, and establish a committee internally to help make these decisions."* (P35)
Generalized use		
Technical	Advantages of specific AI techniques	Advantages of logistic regression: *"Logistic regression, at least they're very simple, you have the advantage that you can understand what the model is doing, which gets removed as soon as you move up one level to a little bit more complex model like boosted trees like this family of model is already harder to understand. and neural networks are even more opaque. "* (P50)
Socio-technical	Risks of AI	Unfair LLM outputs: *"There's a lot of them, and it depends on the usage, and how it's connected to the outside world. The models are applied in production and impacting people's lives. So any kind of bias in the training data, or use of things like race, post code that's associated with bad neighborhood, quote unquote - can lead to problematic decisions by these models."* (P50)
	In-depth understanding of the transformations and harms LLMs causes in society.	Inequalities in the benefits of LLMs *"the thing that worries me most about AI is the potential of leaving the underprivileged behind. It's an extremely powerful technology which is very easy to use for communities that have access to it. As a result, in this knowledge-based economy, you get an edge which can lead to a few getting even more powerful, while those who do not have access to that technology, whether because of technical know-how, or GPUs, they can very easily be left behind."* (P16)
Interactional	Use-cases	*"we have our chief technology officer. He's a very technical, and was one of the first to start experimenting and looking around with it, and stood up a little center of excellence team where they review all the different genAI use cases, and start putting work into it and start developing things, trying to educate our users on the safe use of genAI."* (P43)
Generalized context		
Technical	The conventions of the scientific community working on LLMs.	*"openAI at the moment is considered a de facto, large language model provider. So everything that we need to develop, everyone, including the customers, are interested in knowing how well our solution performs against openAI."* (P22)
Socio-technical	AI policies	Data privacy: *"we wanna scrub malicious data from that, we wanna make sure that there's no PII in it, and most of the risk is being driven by GDPR, the EU AI Act and the Canadian AI act that they're developing."* (P29)
Interactional	Perceptions of AI by various audiences	*"The broader public is becoming more and more familiar and using AI in general, and being impacted by it. So there's more and more concern about AI and investment in AI too."* (P20)
Organizational	Organizational players	*"We don't really have a big arm for process change management which is required for a really good implementation. So that's where our partners come in."* (P35)

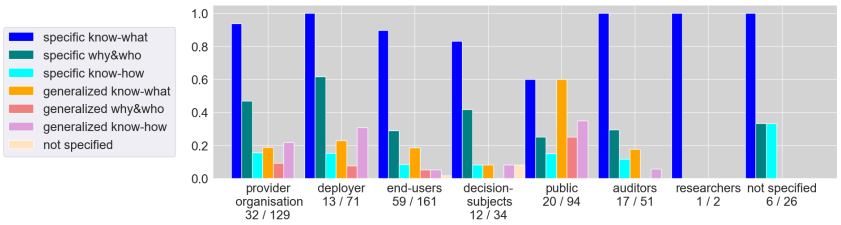
Table 4. Examples of abstract production knowledge.

Knowledge category	**Knowledge nugget**	**Illustrative quote**
Technical artifact	Instrumental knowledge for development	Algorithm training: *"all these models require GPUs, compute power, and storage to do that."* (P15)
Technical artifact	Knowledge for doing, e.g., tricks for algorithm training	*"You don't know in advance exactly what they're going to do. And the experience of selecting the appropriate loss function, it's an experience: you do that a few times, and then hopefully, you're good at it, and you guess right every single time. but thinking in very high-dimensional spaces, people are just really bad at doing that. We don't have the greatest intuitions for it. Which means that the problems can be hard, which means that the likelihood you're going to get it exactly right in a very large parameter space is very low."* (P33)
Technical artifact	Knowledge for reflecting on evaluations	*"We need to pick better metrics. For example, we're aiming to improve accuracy but we cannot just look at accuracy. That's not the only thing that matters. We're still not looking at things like calibration, errors of our models or the data: does it look biased? is it going to be robust?"* (P24)
Socio-technical artifact	Knowledge for doing	Interactions to avoid over-reliance on the system: *"Over-reliance is easy to work on. We can ensure that we keep multiple checkpoints in the product to ensure that they are keeping the human in the loop. And it's not making every single decision. And humans have to really be in there."* (P6)
Socio-technical artifact	Knowledge for reflecting on socio-technical choices	*"One of the big questions these days in frontier AI systems is: how do we evaluate potential risk? One thing is to defend against harm, and the other is to test whether those defenses are good. Some of the way we evaluate is very much algorithmic, but also there are some societal aspects. You have to think of what are the threats from a social perspective."* (P53)
Interactional artifact	Knowledge for doing, e.g., design tricks	*"we have a research project on how to efficiently interact with genAI systems. It's at the research stage. But hopefully, we can get some valuable insights and actionable ones that we can integrate in our products so that our users can use them as efficiently and effectively as they can."* (P49)
Interactional artifact	Knowledge for reflecting on the conditions for interactions with LLM systems	*"AI creates algorithmic capital, and it is not available to everybody to use it to harness more value. And the people who benefit from it, it's the company that gets more capital out of it."* (P2)
Organizational artifact	Instrumental knowledge for development	*"There is a desire of the company to create LLMs to protect customer data. But we need transparency in the training, and set guidelines that govern the collection of data so that we can be confident that what we're building is not transgressing ethical considerations."* (P20)
Organizational artifact	Knowledge for doing, e.g., organizational process design	*"Whenever one thinks there's a gap, let's document that and come up with our mediation plan. And sometimes, by the time you find out that something occurred, it's already been resolved. How did we resolve it last time? Does it show something more systemic? That's the challenge for me: training people in the organization to take those steps of self-identification and self-governance."* (P10)
Organizational artifact	Knowledge for reflecting on the organizational consequences of LLM systems	*"Concentration of power, because those that control AI will control the information, the decisions, and influence people."* (P11)

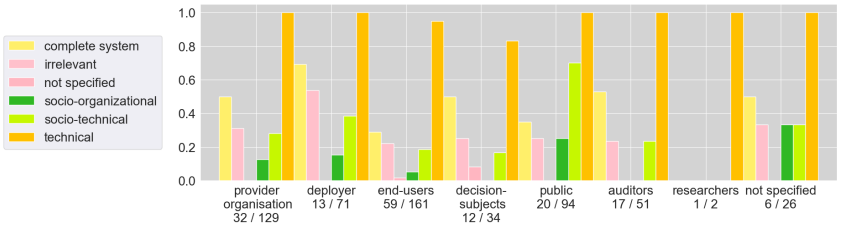
3 ASSESSMENT OF THE KNOWLEDGE TAXONOMY

3.1 Validity of the Taxonomy

3.1.1 Validity for AI practitioners. During the feedback sessions, we presented the taxonomy progressively to the participants. We started by introducing the knowledge categories, then provided examples of knowledge nuggets per category and showed slides with an exhaustive list of these nuggets, and finally zoomed out to the overall taxonomy and the dimensions that connect the nuggets across the categories. In response, the participants did not discuss any nugget that was not already in the taxonomy, which validated the *comprehensiveness* of the taxonomy. the participants primarily discussed how various knowledge nuggets are or could be useful to them and other stakeholders, validating the *importance* of the nuggets included in the taxonomy. For instance, an LLM researcher (P31) discussed the specific production knowledge and explained *"Asking for rationales is important. For instance, why were these metrics and test data used for evaluation? Was it to look at biases? So how were these biases defined?"*



(a) Percentage of publications referring to a knowledge category for each AI supply chain actor.



(b) Percentage of publications referring to a knowledge type for each AI supply chain actor.

Fig. 1. Analysis of publications based on the knowledge stakeholders they refer to (the stakeholders in the AI/LLM supply chain are in bold). We also show the number of publications mentioning the different stakeholders, and the number of knowledge nuggets mentioned by the publication across type of stakeholder.

3.1.2 Generalizability of the Taxonomy. Figure 1 provides a fine-grain description of the knowledge needs reported in existing publications, as a function of the supply chain actors they pertain to. The scopes of actors included in our taxonomy and covered by prior works overlap, but bear differences. Our literature survey confirms that existing publications primarily refer to actors outside the LLM (or AI) supply chain –these can be end-users of the AI systems, decision-subjects, the public, auditors, or researchers. While we found 110 mentions of these actors and 368 mentions of relevant knowledge nuggets for these actors, we only found respectively 45 mentions of actors and 200 of nuggets *within* the AI supply chain (more often of technical background, and working for provider organizations as opposed to deployer ones). As our taxonomy covers all previously mentioned knowledge nuggets, it is applicable to stakeholders outside the AI supply chain.

However, this does not mean that all knowledge nuggets are actually needed by these stakeholders. For instance, the socio-organizational knowledge highlighted by our taxonomy is primarily mentioned in publications that relate to stakeholders within the AI supply chain, but never for decision-subjects or auditors who might have less use for it (see Figure 1b).

3.1.3 Depth of the Knowledge Taxonomy. Our goal was primarily to identify and structure the categories of knowledge needs that researchers or policymakers might neglect, and hence our framework does not detail the knowledge nuggets as deeply as certain prior publications (especially those [13, 34] that focus on a single technical artifact used in multiple AI supply chains, instead of focusing on an entire supply chain as we do).

Note that the depth we use in the knowledge taxonomy follows the one employed by most of our interview participants. Oftentimes, they report their needs at the level of the smallest knowledge sub-categories our taxonomy allows (e.g., they talk about the performance of the LLM model or the pre-processing of the dataset which refer respectively to the specific technology knowledge around the technical artifact output quality or around the technical artifact development), and sometimes at a deeper level (especially when they are of a technical background), e.g., they talk about the GPUs used to train the LLM model which is one level deeper the specific technical production knowledge about the LLM model development. Most publications instead report needs following one single level of depth, typically the deeper one. We recommend explicitly exploiting these different depth levels in the future depending on one's goal. The category-level is easily reported on by stakeholders in a reasonable amount of time, higher levels can serve to reflect on existing literature but are not necessarily actionable, and lower levels used in the literature are more comprehensive and operationalizable but time-consuming to explore.

Mapping the nuggets present in prior publications to our knowledge categories and sub-categories to populate them would be easy and could make the taxonomy even more actionable for certain users (e.g., practitioners developing AI documentation tools). For instance, datasheets for datasets [34] and data-centric explanations [5] contain fine-grained knowledge nuggets (e.g., dataset recommended data splits, potential sources of data noise) that our interview participants with technical development and governance roles discussed as useful. These nuggets, although not explicit in our taxonomy, would fit in the specific, technology knowledge (e.g., data split used, data quality) and production knowledge (e.g., recommended data split) related to the artifact dataset.

3.1.4 Comparison to Other Taxonomies. Table 5 shows how the dimensions in our taxonomy compare to those in prior taxonomies: our taxonomy covers a larger number of dimensions and sub-dimensions.

Table 5. Comparison of the structure of our knowledge taxonomy with other taxonomies.

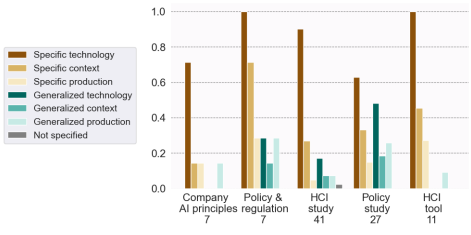
Dimension	Existence in prior taxonomies	Example from prior taxonomies
Abstraction	Conserved since most taxonomies focus on a single sub-dimension.	The AI literacy framework [59] only contains **generalized knowledge**. Datasheets [34] only contain **specific knowledge**.
Theme	Conserved implicitly in the categories of some publications.	Datasheets [34] contain "motivation", 'use", "composition", and "process" categories, respectively reflecting the **context**, **use**, **technology**, and **production** sub-dimensions.
	Not reflected by any category.	Model cards [62] or transparency index [13] do not differentiate by **theme**.
Lens	Conserved since most taxonomies focus on a **single lens**.	Data-centric explanations [5] only contain knowledge related to **technical aspects**.
	A few publications implicitly reflect multiple **lenses**.	The "composition" and "maintenance" categories in datasheets [34] respectively match **technical artifacts** and **organizational processes**.
	Not reflected by any category.	The transparency index [13] includes **organizational processes** such as "model documentation" but does not separate them from knowledge of other **lenses**.

Table 6 compares the specific knowledge categories that our dimensions form with categories in prior taxonomies. Only minor differences are identified, that revolve around the hierarchical level of the categories: our knowledge categories might be split or merged. Such variations do not strongly impact future analyses since our dimensions —knowledge lens, theme, and abstraction— subsume others' dimensions or categories.

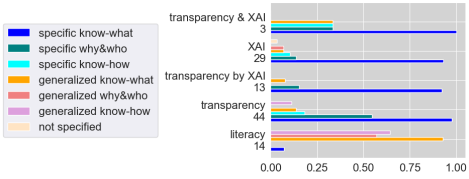
Table 6. Comparison of the categories reflected in our taxonomy and other taxonomies.

Discrepancy	Example
Categories more fine-grained than in our taxonomy	In datasheets [34] and data-centric explanations [5], data processing and data collection are two explicitly separate categories that we left undetailed in our specific production knowledge category (within the design and development nuggets).
Categories with a broader scope than in our taxonomy	The "how does AI work?" category of the literacy framework [59] contains nuggets that we split into multiple sub-categories in the generalized technological knowledge, e.g., differentiating between basic understanding of AI technology and theories behind AI (specifically LLMs in our case).
Same categories split using one dimension of our taxonomy	Sokol et al. [84] discuss three "explanation targets" on the same level (data, model, predictions), whereas they are distinguishable with an additional dimension in our specific technological knowledge category: data and model are two entities while predictions refer to the entity output quality.
Same categories split using multiple dimensions	The transparency index [13] is organized based on the "domains" of foundation models, i.e., whether the knowledge refers to an AI model itself (our specific technological knowledge for LLM models), or downstream consumption activities (our specific contextual knowledge for LLM models), or upstream development activities (our specific production and technological knowledge for LLM models to describe technical artifacts such as datasets used to develop the models).

3.2 Exploitation of the Taxonomy



(a) Percentage of publications of a specific type that mention a category of knowledge nuggets.



(b) Percentage of mentions of a knowledge category among the publications discussing one of the responsible AI notions (AI transparency, explainability, and literacy), two of the notions without distinction (transparency & XAI) or two notions in combination (transparency by XAI refers to publications where XAI is discussed as a means to achieve transparency). The number of publication discussing each responsible AI notion is indicated under the publication typology.

3.2.1 Identification of under-researcher knowledge nuggets and categories. Figure 2a and the more detailed plots (Figure 3) show to what extent the various knowledge categories that we identified as relevant in our interviews are covered by prior AI publications and regulations. Specific technology knowledge is predominant, while all the other categories lack exploration.

3.2.2 Investigation of the conceptualization of responsible AI notions. Our taxonomy enables researchers to disentangle [53] how existing publications define the notions of AI literacy, transparency, and explainability, with significant overlaps across notions and publications. Figure 2b

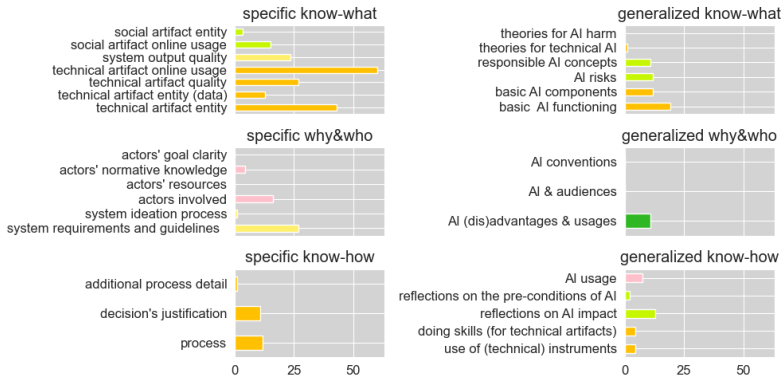


Fig. 3. Percentage of publications that mention each knowledge nugget among all publications. Knowledge nuggets that are never mentioned do not appear in the bar charts.

and the more detailed plots (Figure 4) for instance reveals that both transparency and explainability publications discuss specific knowledge, while both literacy and transparency publications discuss generalized production and technological knowledge, pointing out the potential lack of clear distinction between the two concepts. The lack of clear distinction might be explained by the conceptual ambiguities in how these notions relate to each other: some publications explicitly present explainability as a means for transparency; while others discuss them potentially as interchangeable notions, or independently –sometimes as means, sometimes as goals in themselves.

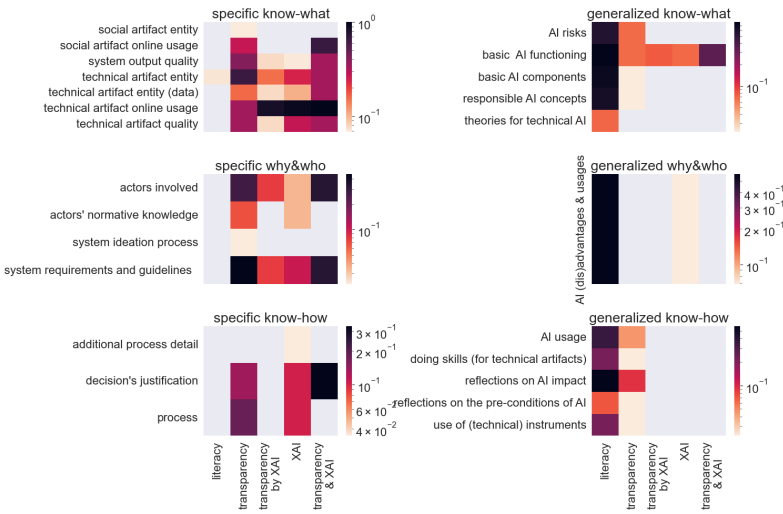


Fig. 4. Percentage of mentions of a knowledge nugget among the publications discussing one of the responsible AI notions (AI transparency, explainability, and literacy), two of the notions without distinction (transparency & XAI) or two notions in combination (transparency by XAI refers to publications where XAI is discussed as a means to achieve transparency). Note that certain knowledge nuggets do not appear at all in the plots because they were not referred to by any publication.

4 DETAILED RESULTS ABOUT STAKEHOLDERS AND KNOWLEDGE GOALS

In this section, we present additional results that we made use of to decide on the dimensions and sub-dimensions of our taxonomy.

4.1 Categorizations of Knowledge Stakeholders & Purposes

4.1.1 Categories of Knowledge Purposes. We identified four types of knowledge purposes, showing the types of interaction the knowledge requester aims to have with other stakeholders involved in their task.

Collaborating: Working with others by consuming and sharing knowledge. One needs to consume and share knowledge in order to work with others. We found several ways in which knowledge is used in this context. One might need knowledge from others to coordinate efforts, avoid repeated efforts and increase the efficiency of an organization; as well as to make sure that expected work has been conducted successfully by the relevant individuals or teams; or they might require knowledge to identify how they can rely on others and use their support, e.g., utilizing their relevant knowledge, expertise, or resources. Knowledge can also be used to enter into discussions with others and help or challenge their own decision-making process. Finally, certain participants discussed using knowledge nuggets to decide whether to enter into a collaboration with others, whether to start working for others, and/or whether to trust others at all.

Informing other's decisions by sharing knowledge. Knowledge nuggets can be shared with others to inform them and impact their work across the LLM supply chain. Particularly, we identified several ways in which a knowledge sharer aims at impacting a potential knowledge consumer. This might be done without any ulterior motives (e.g., they simply have the will to educate another stakeholder because they deem a topic important), or in order to impact further decisions by the knowledge consumer. Particularly, this impact can revolve around managing the consumer's expectations, especially when a potential AI consumer starts using an AI system or its component for the first time; or fostering trust of the consumer, such as when an AI user might have to choose which AI provider to work with. More broadly, one might also aim at developing the image of their organization to the greater public by sharing knowledge, such as displaying and attesting of responsible AI efforts, especially in comparison to competing organizations. It can also be about implicitly fostering active efforts in one direction such as fostering activities to ensure responsible AI development and use; or explicitly convincing one to make a decision, such as the decisions to put in place new protocols to ensure privacy preservation during data collection.

Informing one's own decisions by consuming knowledge. In the context of their LLM supply chain activities, participants mentioned requiring various types of knowledge to conduct their usual tasks. This can for instance be about knowledge to decide how to develop an AI component or to assemble multiple components into one AI system, knowledge to make strategic decisions around the development of future AI components or the adoption or commercialisation of such AI components, or knowledge to decide when to rely on the outputs of an AI system (for its users). It can also be knowledge in order to set up governance efforts within or across junctions of the LLM supply chain. In some cases, we found that our participants desired to access knowledge to make decisions that go beyond their typical activities, and expand on their expected work. That was especially the case for participants who, for instance, were not tasked to reflect about responsible AI topics by their employer, but, deeming it important, decided on learning about the topic to incorporate new elements and reflections in their own practices and the practices of their colleagues.

Ensuring traceability by communicating knowledge. Keeping traces of activities and reflections happening across the supply chain is not necessarily intended specifically for oneself or others, and does not always serve a specific and immediate purpose –some of our participants mentioned that a knowledge nugget might be important for them to know while admitting that they could not express the purpose of the nugget just yet, and that it could become relevant in the future. In some cases, ensuring the traceability comes from the intent of the participant to support their future work or the future work of others, and in other cases, it has been explicitly suggested or requested by another entity, especially for accountability purposes and debugging of future issues the AI system could cause, or simply to comply to general requirements such as around transparency. For instance, the consumer of the knowledge nugget or any other entity such as a regulation might require specific information to be shared around responsible AI topics such as the evaluation of the outputs of an AI system.

4.1.2 Categories of knowledge stakeholders. We find that our participants, regardless of their designation within the supply chain, might have several of the knowledge purposes described above. We also find that the specific knowledge nuggets needed by our participants differ based on the type of task they have to conduct within a specific knowledge purpose mentioned (this is easily imaginable with the purpose related to informing one's own decisions for which the types of decisions to be made are directly related to one's designation). Particularly, we find that stakeholders vary based on the *goal of the activities* they conduct. They can either have a strategizer, do-er, or evaluator role. Besides, the *disciplinary orientation of these activities* matters. Indeed, we find that stakeholders' activities vary based on the kind of discipline their role fits in. This can be technical, user-centered, responsible AI-focused, customer-centered (i.e., business), or governance-related. These two dimensions can be combined into the types of roles the stakeholders occupy across the LLM supply chain, and subsequently the type of knowledge they might require. For instance, a technical do-er role corresponds to an AI developer, data engineer role, or machine learning engineer who would develop an AI system, while a technical strategizer role might fit more the software architect designation, or potentially the product manager designation which would be accompanied by customer-centered strategizers as well. Such customer-centered strategizer roles could also include employees of organizations that adopt AI who make the adoption decisions on the organization to work with, as well as employees of organizations that develop AI and make decisions based on business variables on which technical infrastructure to use, etc.

4.2 Patterns of Knowledge Needs

Based on our interview results, an approximate mapping can be outlined between the categories of knowledge nuggets needed (the what), the categories of knowledge purposes (the why), and the knowledge stakeholders (the who) that might want to consume or share this knowledge.

Primary trends identified. Naturally, general trends appear between knowledge categories and knowledge purposes and their stakeholders' activities. For instance, regardless of the specific stakeholder, we find that the more work is conducted in interaction with others (particularly collaborating with others and informing others' decisions), the more the production and context knowledge are important, in order to first identify who to work with and how. Then, the knowledge that will be shared with the identified knowledge consumer often revolves around specific technology knowledge for both needs, or translation knowledge to support others in conducting an activity (collaboration), or abstract knowledge for informing other's decisions.

Similarly, for keeping track of the LLM supply chain activities for oneself or others, we found that the context and production knowledge are useful and potentially mandatory (according to certain regulations, such as the AI Act [61] that requires to reveal the identity of the LLM system

developer), as they allow for traceability and accountability, e.g., in case of issues caused by the LLM system. As a result, the organizational abstract knowledge is also relevant, especially in identifying obligations for reporting. Additionally, the precise specific technology knowledge related to the conducted activities is required to precisely report on these activities.

The knowledge for informing one's own decisions is the one that requires further differentiation between the different types of stakeholders to better understand their knowledge needs. Below, we describe some additional correlations we identified between the who, what, and why of our LLM knowledge framework.

Trends for informing one's own decisions —technical doer roles. We found that technical do-er roles primarily require technical knowledge across the two knowledge abstractions and knowledge of AI as a research field in order to respect well-accepted conventions. Here the translation knowledge is often an intermediary. It is typically used in combination with specific knowledge in order to develop new abstract knowledge (this is especially the case for AI researchers), or instead is used in combination with abstract knowledge in order to make new decisions for developing new LLM components (and hence generating new specific technical technology knowledge).

Trends for informing one's own decisions —governance roles. Instead, we found that governance roles are primarily interested in specific and abstract knowledge, in order to establish new organizational processes based on their understanding of current processes in their organization (specific production knowledge) and their knowledge of potential other organizations (e.g., precise contextual knowledge about competitors or generalized knowledge of research advances, i.e., generalized technology knowledge), as well as their own understanding of potential issues with current I systems and components and potential risks they can cause –understanding that is developed through reflecting about the current systems via technical and socio-technical technology knowledge, as well as based on their potential understanding of the inherent limitations and harm that LLMs can cause (generalized technology knowledge), their own reflections (generalized production knowledge), and their knowledge of regulations (AI technology consumption by society, i.e., generalized contextual knowledge).

Trends for informing one's own decisions —technical evaluator and strategizer roles. We found that technical evaluators might require technical technology knowledge solely combined with socio-technical technology knowledge and AI technology consumption by society in order to evaluate LLM systems and their components according to relevant requirements. Technical strategizers, additionally to such knowledge required by the evaluators, might also need translation knowledge or generalized technical technology knowledge to plan new efforts and estimate what is feasible to concretely do.

Trends for informing one's own decisions —user-centered and customer-centered roles. User-centered and customer-centered roles revolve around identifying user and customer needs, strategizing over these, and developing relevant LLM components, or making meaningful adoption decisions. We found that the identification activities primarily rely on hands-on work based on socio-technical production knowledge or generalized socio-technical technology and context knowledge to identify who to investigate. Instead, evaluating the LLM components for the requirements identified primarily requires generalized socio-technical production knowledge. Planning the development of the relevant LLM components (often by the product managers) then relies on a combination of the acquired knowledge, their own understanding of the technical and economic feasibility of the work (generalized, technical, technology and production knowledge), and their prior knowledge and reflections of what could be developed (again translation knowledge or generalized knowledge of existing products).

Individual differences in the connection. We found that the connections described do not solely depend on the knowledge purpose of a specific stakeholder role. These are only general trends with exceptions due to differences among individual stakeholders. For instance, a technical strategizer having to decide whether to adopt an LLM component of one organization or another might investigate various properties of the performance of the component (e.g., its accuracy, latency, memory-usage, etc.), or they might make a decision based on the organization itself and their historical collaborations. Different types of knowledge (specific technical technology or production knowledge) are involved in this same decision. These variations seem to be impacted by the organization employing the individual and its practices and goals, as well as by the individual's prior knowledge, acquired practices, goals, and opinions, as well as on their expertise and skills. For instance, two different strategizers might have different backgrounds, with one having more experience critically reflecting on the technical technology knowledge (e.g., the limitations of benchmarks), while the other might have less technical experience but a better understanding of the employees in their teams and their capabilities (context knowledge) and more or less opportunity and time to delve more in-depth into the benchmark.

REFERENCES

[1] 2022. Microsoft Responsible AI Standard, v2. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl?culture=en-us&country=us

[2] David Adkins, Bilal Alsallakh, Adeel Cheema, Narine Kokhlikyan, Emily McReynolds, Pushkar Mishra, Chavez Procope, Jeremy Sawruk, Erin Wang, and Polina Zvyagina. 2022. Prescriptive and descriptive approaches to machine-learning transparency. In *CHI conference on human factors in computing systems extended abstracts*. 1–9.

[3] NIST AI. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0).

[4] Gloria Andrada, Robert W Clowes, and Paul R Smart. 2023. Varieties of transparency: Exploring agency within AI systems. *AI & society* 38, 4 (2023), 1321–1331.

[5] Ariful Islam Anik and Andrea Bunt. 2021. Data-centric explanations: explaining training data of machine learning systems to promote transparency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[6] Matthew Arnold, Rachel KE Bellamy, Michael Hind, Stephanie Houde, Sameep Mehta, Aleksandra Mojsilović, Ravi Nair, K Natesan Ramamurthy, Alexandra Olteanu, David Piorkowski, et al. 2019. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development* 63, 4/5 (2019), 6–1.

[7] Vijay Arya, Rachel K Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C Hoffman, Stephanie Houde, Q Vera Liao, Ronny Luss, Aleksandra Mojsilovic, et al. 2021. One Explanation Does Not Fit All: A Toolkit And Taxonomy Of AI Explainability Techniques. In *INFORMS Annual Meeting*.

[8] Agathe Balayn, Natasa Rikalo, Christoph Lofi, Jie Yang, and Alessandro Bozzon. 2022. How can explainability methods be used to support bug identification in computer vision models?. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[9] Iain Barclay, Harrison Taylor, Alun Preece, Ian Taylor, Dinesh Verma, and Geeth de Mel. 2021. A framework for fostering transparency in shared artificial intelligence models by increasing visibility of contributions. *Concurrency and Computation: Practice and Experience* 33, 19 (2021), e6129.

[10] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, et al. 2021. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 401–413.

[11] Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José MF Moura, and Peter Eckersley. 2020. Explainable machine learning in deployment. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 648–657.

[12] IBM THINK Blog. 2017. Transparency and Trust in the Cognitive Era. https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/

[13] Rishi Bommasani, Kevin Klyman, Shayne Longpre, Sayash Kapoor, Nestor Maslej, Betty Xiong, Daniel Zhang, and Percy Liang. 2023. The foundation model transparency index. *arXiv preprint arXiv:2310.12941* (2023).

[14] Rishi Bommasani, Daniel Zhang, Tony Lee, and Percy Liang. 2023. Improving transparency in ai language models: A holistic evaluation. *HAI Policy & Society* (2023).

[15] Karen L Boyd. 2021. Datasheets for datasets help ML engineers notice and understand ethical issues in training data. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–27.

[16] Andrea Brennen. 2020. What Do People Really Want When They Say They Want" Explainable AI?" We Asked 60 Stakeholders.. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–7.

[17] Approved by Executive Committee. 2018. AI Principles of Telefonica. https://www.telefonica.com/en/wp-content/uploads/sites/7/2021/11/principios-ai-eng-2018.pdf

[18] Michael Chromik and Martin Schuessler. 2020. A Taxonomy for Human Subject Evaluation of Black-Box Explanations in XAI. *Exss-atec@ iui* 1 (2020), 1–7.

[19] Douglas Cirqueira, Dietmar Nedbal, Markus Helfert, and Marija Bezbradica. 2020. Scenario-based requirements elicitation for user-centric explainable AI: A case in fraud detection. In *International cross-domain conference for machine learning and knowledge extraction*. Springer, 321–341.

[20] Lorenzo Corti, Rembrandt Oltmans, Jiwon Jung, Agathe Balayn, Marlies Wijsenbeek, and Jie Yang. 2024. "It Is a Moving Process": Understanding the Evolution of Explainability Needs of Clinicians in Pulmonary Medicine. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.

[21] Karina Cortiñas-Lorenzo, Siân Lindley, Ida Larsen-Ledet, and Bhaskar Mitra. 2024. Through the Looking-Glass: Transparency Implications and Challenges in Enterprise AI Knowledge Systems. *arXiv preprint arXiv:2401.09410* (2024).

[22] Anamaria Crisan, Margaret Drouhard, Jesse Vig, and Nazneen Rajani. 2022. Interactive model cards: A human-centered approach to model documentation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 427–439.

[23] Roxana Daneshjou, Mary P Smith, Mary D Sun, Veronica Rotemberg, and James Zou. 2021. Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review. *JAMA dermatology* 157, 11 (2021), 1362–1369.

[24] Karl de Fine Licht and Jenny de Fine Licht. 2020. Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & society* 35 (2020), 917–926.

[25] Shipi Dhanorkar, Christine T Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. 1591–1602.

[26] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–19.

[27] Upol Ehsan, Samir Passi, Q Vera Liao, Larry Chan, I Lee, Michael Muller, Mark O Riedl, et al. 2021. The who in explainable ai: How ai background shapes perceptions of ai explanations. *arXiv preprint arXiv:2107.13509* (2021).

[28] Upol Ehsan, Koustuv Saha, Munmun De Choudhury, and Mark O Riedl. 2023. Charting the sociotechnical gap in explainable ai: A framework to address the gap in xai. *Proceedings of the ACM on human-computer interaction* 7, CSCW1 (2023), 1–32.

[29] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Martina Mara, Marc Streit, Sandra Wachter, Andreas Riener, and Mark O Riedl. 2021. Operationalizing human-centered perspectives in explainable AI. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 1–6.

[30] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Elizabeth Anne Watkins, Carina Manger, Hal Daumé III, Andreas Riener, and Mark O Riedl. 2022. Human-Centered Explainable AI (HCXAI): beyond opening the black-box of AI. In *CHI conference on human factors in computing systems extended abstracts*. 1–7.

[31] Heike Felzmann, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society* 6, 1 (2019), 2053951719860542.

[32] Verena Fulde. 2018. Guidelines for Artificial Intelligence. https://www.telekom.com/en/company/digital-responsibility/details/artificial-intelligence-ai-guideline-524366

[33] Christian Garbin and Oge Marques. 2022. Assessing methods and tools to improve reporting, increase transparency, and reduce failures in machine learning applications in health care. *Radiology: Artificial Intelligence* 4, 2 (2022), e210127.

[34] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.

[35] Google. [n. d.]. AI Principles: Objectives for building beneficial AI. https://ai.google/static/documents/EN-AI-Principles.pdf

[36] Sony Group AI Ethics Guidelines. 2018. AI Engagement within Sony Group. https://www.sony.com/en/SonyInfo/csr_report/humanrights/AI_Engagement_within_Sony_Group.pdf

[37] Benjamin Haibe-Kains, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Massive Analysis Quality Control (MAQC) Society Board of Directors Shraddha Thakkar 35 Kusko Rebecca 36 Sansone Susanna-Assunta 37 Tong Weida 35 Wolfinger Russ D. 38 Mason Christopher E. 39 Jones Wendell 40 Dopazo Joaquin 41 Furlanello Cesare 42, Levi Waldron, Bo Wang, Chris McIntosh, Anna Goldenberg, Anshul Kundaje, et al. 2020. Transparency and reproducibility in artificial intelligence. *Nature* 586, 7829 (2020), E14–E16.

[38] Kashyap Haresamudram, Stefan Larsson, and Fredrik Heintz. 2023. Three levels of AI transparency. *Computer* 56, 2 (2023), 93–100.

[39] Xin He, Yeyi Hong, Xi Zheng, and Yong Zhang. 2023. What are the users' needs? Design of a user-centered explainable artificial intelligence diagnostic system. *International Journal of Human–Computer Interaction* 39, 7 (2023), 1519–1542.

[40] Lukas-Valentin Herm, Kai Heinrich, Jonas Wanner, and Christian Janiesch. 2023. Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. *International Journal of Information Management* 69 (2023), 102538.

[41] Steven Hicks, Debesh Jha, Vajira Thambawita, Pål Halvorsen, Bjørn-Jostein Singstad, Sachin Gaur, Klas Pettersen, Morten Goodwin, Sravanthi Parasa, Thomas de Lange, et al. 2021. Medai: Transparency in medical image segmentation. *Nordic Machine Intelligence* 1, 1 (2021), 1–4.

[42] Michael Hind, Stephanie Houde, Jacquelyn Martino, Aleksandra Mojsilovic, David Piorkowski, John Richards, and Kush R Varshney. 2020. Experiences with improving the transparency of AI models and services. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[43] Robert R Hoffman, Shane T Mueller, Gary Klein, Mohammadreza Jalaeian, and Connor Tate. 2023. Explainable AI: roles and stakeholders, desirements and challenges. *Frontiers in Computer Science* 5 (2023), 1117848.

[44] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M Drucker. 2019. Gamut: A design probe to understand how data scientists understand machine learning models. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[45] Marie Hornberger, Arne Bewersdorff, and Claudia Nerdel. 2023. What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence* 5 (2023), 100165.

[46] Jinglu Jiang, Surinder Kahai, and Ming Yang. 2022. Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies* 165 (2022), 102839.

[47] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.

[48] Sunnie SY Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. "Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[49] René F Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2390–2395.

[50] Siu-Cheung Kong, Man-Yin William Cheung, and Olson Tsang. 2024. Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. *Computers and Education: Artificial Intelligence* 6 (2024), 100214.

[51] Siu-Cheung Kong, William Man-Yin Cheung, and Guo Zhang. 2021. Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers and Education: Artificial Intelligence* 2 (2021), 100026.

[52] Siu-Cheung Kong, William Man-Yin Cheung, and Guo Zhang. 2023. Evaluating an artificial intelligence literacy programme for developing university students' conceptual understanding, literacy, empowerment and ethical awareness. *Educational Technology & Society* 26, 1 (2023), 16–30.

[53] Patrick Lambe. 2014. *Organising knowledge: taxonomies, knowledge and organisational effectiveness*. Elsevier.

[54] Matthias Carl Laupichler, Alexandra Aster, and Tobias Raupach. 2023. Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy. *Computers and Education: Artificial Intelligence* 4 (2023), 100126.

[55] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User experience design professionals' perceptions of generative artificial intelligence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[56] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–15.

[57] Q Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly understanding: Information needs for model transparency to support design ideation for AI-powered user experience. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–21.

[58] Q Vera Liao and Jennifer Wortman Vaughan. [n. d.]. Ai transparency in the age of llms: A human-centered research roadmap. ([n. d.]).

[59] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.

[60] Ana Lucic, Madhulika Srikumar, Umang Bhatt, Alice Xiang, Ankur Taly, Q Vera Liao, and Maarten de Rijke. 2021. A multistakeholder approach towards evaluating AI transparency mechanisms. *arXiv preprint arXiv:2103.14976* (2021).

[61] Tambiama Madiega. 2021. Artificial intelligence act. *European Parliament: European Parliamentary Research Service* (2021).

[62] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

[63] Sina Mohseni, Jeremy E Block, and Eric Ragan. 2021. Quantitative evaluation of machine learning explanations: A human-grounded benchmark. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 22–31.

[64] Sina Mohseni, Niloofar Zarei, and Eric D Ragan. 2021. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11, 3-4 (2021), 1–45.

[65] Shane T Mueller, Elizabeth S Veinott, Robert R Hoffman, Gary Klein, Lamia Alam, Tauseef Mamun, and William J Clancey. 2021. Principles of explanation in human-AI systems. *arXiv preprint arXiv:2102.04972* (2021).

[66] José Luiz Nunes, Gabriel DJ Barbosa, Clarisse Sieckenius de Souza, Helio Lopes, and Simone DJ Barbosa. 2022. Using model cards for ethical reflection: a qualitative exploration. In *Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems*. 1–11.

[67] Government of Canada. 2023. The Artificial Intelligence and Data Act (AIDA). https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document
[68] White House Office of Science and Technology Policy. 2022. Blueprint for an AI Bill of Rights: making automated systems work for the American People.
[69] Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–9.
[70] Vasiliki Papadouli. 2022. Transparency in artificial intelligence: a legal perspective. *Journal of Ethics and Legal Technologies* 4 (2022), 25–40.
[71] JD Perchik, AD Smith, AA Elkassem, JM Park, SA Rothenberg, M Tanwar, PH Yi, A Sturdivant, S Tridandapani, and H Sotoudeh. 2023. Artificial intelligence literacy: developing a multi-institutional infrastructure for AI education. *Academic radiology* 30, 7 (2023), 1472–1480.
[72] David Piorkowski, John Richards, and Michael Hind. 2022. Evaluating a methodology for increasing AI transparency: A case study. *arXiv preprint arXiv:2201.13224* (2022).
[73] OECD policy observatory. 2024. OECD AI Principles overview. https://oecd.ai/en/ai-principles
[74] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–52.
[75] Australia's AI Ethics Principles. 2023. Australia's Artificial Intelligence Ethics Framework. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles
[76] Mahima Pushkarna, Andrew Zaldivar, and Oddur Kjartansson. 2022. Data cards: Purposeful and transparent dataset documentation for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1776–1826.
[77] PwC. [n. d.]. The responsible AI framework. https://www.pwc.co.uk/services/risk/insights/accelerating-innovation-through-responsible-ai/responsible-ai-framework.html
[78] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
[79] Yao Rong, Tobias Leemann, Thai-Trang Nguyen, Lisa Fiedler, Peizhu Qian, Vaibhav Unhelkar, Tina Seidel, Gjergji Kasneci, and Enkelejda Kasneci. 2023. Towards human-centered explainable ai: A survey of user studies for model explanations. *IEEE transactions on pattern analysis and machine intelligence* (2023).
[80] Philipp Schmidt, Felix Biessmann, and Timm Teubner. 2020. Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* 29, 4 (2020), 260–278.
[81] Bianca GS Schor, Chris Norval, Ellen Charlesworth, and Jatinder Singh. 2024. Mind The Gap: Designers and Standards on Algorithmic System Transparency for Users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
[82] Innovation Secretary of State for Science and Technology. 2023. A pro-innovation approach to AI regulation. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
[83] Aubrey A Shick, Christina M Webber, Nooshin Kiarashi, Jessica P Weinberg, Aneesh Deoras, Nicholas Petrick, Anindita Saha, and Matthew C Diamond. 2024. Transparency of artificial intelligence/machine learning-enabled medical devices. *NPJ Digital Medicine* 7, 1 (2024), 21.
[84] Kacper Sokol and Peter Flach. 2020. Explainability fact sheets: a framework for systematic assessment of explainable approaches. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 56–67.
[85] Jane Southworth, Kati Migliaccio, Joe Glover, David Reed, Christopher McCarty, Joel Brendemuhl, Aaron Thomas, et al. 2023. Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence* 4 (2023), 100127.
[86] Karin Stolpe and Jonas Hallström. 2024. Artificial intelligence literacy for technology education. *Computers and Education Open* 6 (2024), 100159.
[87] Lingyun Sun, Zhuoshu Li, Yuyang Zhang, Yanzhen Liu, Shanghua Lou, and Zhibin Zhou. 2021. Capturing the trends, applications, issues, and potential strategies of designing transparent AI agents. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–8.
[88] Harini Suresh, Steven R Gomez, Kevin K Nam, and Arvind Satyanarayan. 2021. Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.
[89] Tom Van Nuenen, Xavier Ferrer, Jose M Such, and Mark Coté. 2020. Transparency for whom? Assessing discriminatory artificial intelligence. *Computer* 53, 11 (2020), 36–44.
[90] Sebastian Vollmer, Bilal A Mateen, Gergo Bohner, Franz J Király, Rayid Ghani, Pall Jonsson, Sarah Cumbers, Adrian Jonas, Katherine SL McAllister, Puja Myles, et al. 2020. Machine learning and artificial intelligence research for patient

benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *bmj* 368 (2020).
[91] Michael Vössing, Niklas Kühl, Matteo Lind, and Gerhard Satzger. 2022. Designing transparency for effective human-AI collaboration. *Information Systems Frontiers* 24, 3 (2022), 877–895.
[92] Joel Walmsley. 2021. Artificial intelligence and the value of transparency. *AI & society* 36, 2 (2021), 585–595.
[93] Bingcheng Wang, Pei-Luen Patrick Rau, and Tianyi Yuan. 2023. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & information technology* 42, 9 (2023), 1324–1337.
[94] Yao Xie, Melody Chen, David Kao, Ge Gao, and Xiang'Anthony' Chen. 2020. CheXplain: enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[95] John Zerilli, Umang Bhatt, and Adrian Weller. 2022. How transparency modulates trust in artificial intelligence. *Patterns* 3, 4 (2022).
[96] Wencan Zhang and Brian Y Lim. 2022. Towards relatable explainable AI with the perceptual process. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–24.