Understanding Stakeholders' Perceptions and Needs Across the LLM Supply Chain

Agathe Balayn
ServiceNow, Delft University of
Technology, Trento University
a.m.a.balayn@tudelft.nl

Lorenzo Corti
Delft University of Technology
l.corti@tudelft.nl

Fanny Rancourt
ServiceNow
fanny.rancourt@servicenow.com

Fabio Casati
ServiceNow
fabio.casati@servicenow.com

Ujwal Gadiraju
Delft University of Technology
u.k.gadiraju@tudelft.nl

Abstract

Explainability and transparency of AI systems are unde-
niably important, leading to several research studies and
tools addressing them. Existing works fall short of account-
ing for the diverse stakeholders of the AI supply chain who
may differ in their needs and consideration of the facets of
explainability and transparency. In this paper, we argue for
the need to revisit the inquiries of these vital constructs in
the context of LLMs. To this end, we report on a qualitative
study with 71 different stakeholders, where we explore the
prevalent perceptions and needs around these concepts.
This study not only confirms the importance of exploring the
"who" in XAI and transparency for LLMs, but also reflects
on best practices to do so while surfacing the often forgot-
ten stakeholders and their information needs. Our insights
suggest that researchers and practitioners should simulta-
neously clarify the "who" in considerations of explainability
and transparency, the "what" in the information needs, and
"why" they are needed to ensure responsible design and
development across the LLM supply chain.

A Case for XAI Across the Supply Chain

Research addressing explainability and transparency of AI
systems seldom clarifies the purpose and the target stake-
holders, often assuming the information needs of stakehold-
ers with little or no validation [17, 12, 23]. There has been
an increasing recognition that the development of AI sys-

S1 – Study participants.
We cover a plurality of domains in the supply chain, including engineering (AI developers, AI researchers), user experience (user research, content design), business (product managers, business analysts, customer service), and governance (legal and risk teams, government relations). The practical tasks of our participants involve, e.g., implementing, testing, and overseeing AI systems, or strategizing around the requirements of the system and broader organization (e.g., functional requirements, definition of risks, etc.). These participants' responsibilities span executives, managers, less senior employees, etc. Finally, we also interviewed end-users of the AI systems developed by our participants –they are internal to the companies developing the AI systems or independent thereof, and none are AI experts.

tems spans a long supply chain of stakeholders [26, 7, 20]. However, only AI developers [3, 8] and end-users [14, 10] have received considerable attention thus far. Other stakeholders such as legal teams or product managers remain largely absent in existing explorations. Inspired by prior work [9, 11, 13, 27], we argue for the necessity to broaden our understanding of AI stakeholders, their engagement and needs pertaining to explanations and transparency.

Particularly, in an era in which AI systems are increasingly closed-source and concealed behind API access (e.g., LLM-based chatbots), stakeholders may elicit *new* needs for explainability and transparency. Following is an example of such information needs across stakeholders. An *applied researcher* within organization A might be looking for an open-source foundation model of any organization and might select the model of *organization B* that is pre-trained on the dataset containing the most relevant data samples to the application they are developing. Then, the *legal team* of organization A might have to enquire about the *data collection practices* of organization B, e.g., to make sure that they respect licensing agreements [19] and labor rights of data annotators [1]. Finally, the *quality team* of organization C deploying the AI system they would buy from organization A would need to inspect explanations about the outputs of the system to make sure that the system performs well-enough for their application.

It is complex and challenging, yet vital to investigate and surface such underexplored aspects in the AI supply chain and relevant practices across organizations. Aligned with the notion of interpretative flexibility [16], stakeholders' perceptions of AI systems directly affect their needs and collaborative practices across the AI supply chain. It can therefore be challenging to appropriately investigate known and unknown explainability and transparency needs.

In this work, we take a formative step towards characterising "who" the stakeholders of explainability and transparency are across the AI supply chain and "what" their information needs are. We report early results from a qualitative inquiry about the perceptions and needs for explainability and transparency of 71 AI stakeholders across different organizations. Our results underscore the importance of grounding studies in the *who*, *what*, and *why* of explainability and transparency, reflect on methodological best practices, and bear implications for the HCXAI community in establishing and shaping relevant policies.

Study Methodology
We interviewed 71 participants, across 10 private organizations, and spanning a wide variety of roles (see **S1**). Participants were recruited via our professional network, with snowball and convenience sampling, to recruit a diversity of stakeholders. All participants develop or use LLMs. In the interviews (on average 45 mins.), we first asked our participants to describe their day-to-day activities, and to discuss their primary concerns, challenges, and needs concerning the harmful impact of AI systems. This allowed us to identify information needs. We then probed their understanding of explainability and transparency: what these concepts evoke, how important they are, and what the entailing challenges might be. We first analysed the interviews through deductive coding looking for the "what", "who", and challenges of the informational needs across stakeholders. Then, we performed a round of inductive coding to identify other recurrent themes.

Results
1) The "*what*": Explainability and Transparency
Generally, our participants viewed transparency as encompassing information beyond AI models (whereas XAI was viewed as being concerned with these models) such as

| S2 – Information items discussed by our AI stakeholders.	
Information need	%
"LLM performance":	
fairness-metric related	2
limits of LLM	22
"LLM decisions":	
confidence for each LLM output	2
fact that output is LLM-generated	18
log of LLM inputs & outputs	4
traceability of errors & outputs	9
Explainability:	
explanations of LLM output via LLM input & addit. info used by LLM	20
LLM system rationales leading to an LLM output	47
"LLM internals":	
LLM design decisions	53
LLM evaluation decisions	18
customer data handling	20
"Meta info":	
basic explanations about LLM	24
guidelines for good use of LLM by customer	9
who are the creators	2
open source	2
what other teams do about LLM or for a specific LLM	4

training data, code, and reports about other supply chain stakeholders. Some information needs related to the outputs of the models, be it their trustworthiness (accuracy or fairness), surrounding information (e.g., confidence scores), or explanations (as understood by the algorithmic XAI community). Other information items related to the internal workings of the LLM (e.g., details about its training dataset) or its 'meta-properties' (e.g., organization developing it).

Our study points toward several information needs that have been discussed in prior literature [17, 12, 2, 5]. Interestingly enough, more recent artefacts such as the Foundation Model Transparency Index [5], cover elements that our participants perceived as general concerns rather than specific to explainability and transparency (see **S2**). These include data labor, environmental impact, and broader ethical concerns. Regardless, participants focused on elements that are often considered incidental for XAI and transparency (e.g., confidence values for individual predictions) or "less-technical" (e.g., developers' discretionary choices, discussed by a few prior works [10, 20]). Some participants perceived XAI and transparency as information primarily useful for end-users and AI system deployers. Disclosure on LLM-generated samples (18% of participants) and basic explanations pertaining to LLMs (24% of participants) were of particular interest. Other participants discussed their own information needs.

2) The "why": Connecting "what" and "who"
Given the broad range of AI stakeholders interviewed, we note several trends concerning the purposes for explainability and transparency. Unsurprisingly, such trends are shaped by a stakeholder's role, prior knowledge, and network of collaborators (see Figure 1). Developers, algorithmic researchers, and quality assessors frequently mentioned technical information (e.g., prediction confidence and

design decisions) necessary to select which AI model to fine-tune, debug the overall system, or identify the potential algorithmic harms. On the other hand, UX designers and researchers provide guidelines to customers for responsible use of the AI system. Together with product managers, they emphasized disclosing essential information to customers or end-users (e.g., to gain buy-in) or to ethics review boards to attest to Responsible AI endeavors (e.g., appropriate handling of customer data). Such a purpose appears to be met by end-users' need for transparency from the makers of AI-powered products (e.g., to gauge the trustworthiness of a company). Finally, we note that XAI and transparency were attributed cross-functional purposes around fostering Responsible AI practices and accountability internally and externally across the AI supply chain.
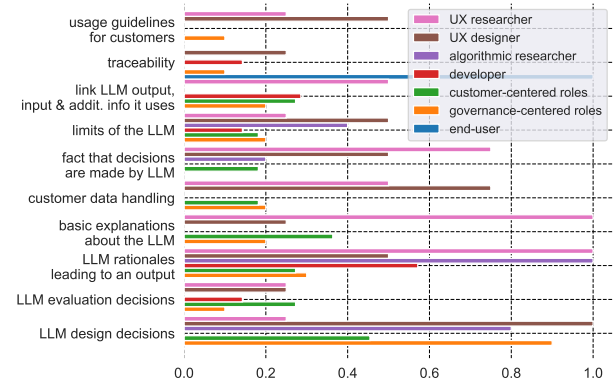
Figure 1: Information needs per category of stakeholders (fraction of the participants who mentioned the need). We excluded information needs mentioned by a single category of stakeholders: developers - LLM confidence (14%); customer-centered roles - log of LLM inputs/outputs (18%), fairness performance (9%); UX researcher - open-source (25%), work of other teams (50%); end-user - LLM creators (100%).

3) Challenges faced by AI stakeholders

Our study highlighted several challenges AI stakeholders face concerning explainability and transparency both at an individual and collaborative level.

At an individual level, participants expressed feasibility tensions in implementing explainability and transparency, e.g., between XAI and customers' data confidentiality. For instance, the quality assessment team discussed challenges with *information access*. They reported on their need to get to know customers and end-users of the LLMs, to appropriately assess these LLMs (e.g., considering social biases). Yet, they emphasized that accessing any information about such external stakeholders is impossible from their position in the AI supply chain. Additionally, technical implementation details are often difficult to accurately report due to (1) the fast-paced development of LLMs, (2) potential exposure to adversarial attacks, and (3) business secrets.

Participants in our study reported mixed needs for disclosure or concealment of information. Broadly, at a collaborative level, many stakeholders appear to *lack knowledge* about the functioning of AI and explainability and transparency. For instance, several deployers of AI systems recognized not being aware of potential harms caused by AI systems and hence not being able to ask for relevant information in that regard. Other participants, such as some algorithmic researchers, were only cognizant of algorithmic XAI concepts and debates around AI open-source, which shows a *narrow perspective* on XAI. This prevents them from identifying and understanding information relevant to either their or others' work.

As a consequence, we identified *misconceptions* (e.g., a participant argued that making an AI system transparent necessarily leads to ensuring the fairness of its outputs) and *terminological confusion* around XAI and transparency (e.g., most participants use "transparency", "explainability" or "accountability" assuming a certain meaning that they disentangle only when prompted), which, more broadly, can cause miscommunication and mishandling of AI systems. Finally, participants' surroundings and time constraints (owing to the ongoing LLMs race) affect their views on XAI and transparency, hampering appropriate reflections.

Discussion & Implications

Our results emphasise the importance of adopting a broad view of the AI supply chain when discussing XAI and transparency, illustrating an existing lack of understanding of explainability and transparency across it. As prior works (e.g., [4, 2, 13, 18, 21, 15]) had not identified as many stakeholders and their purposes [25, 24, 6], nor all information items, we argue that there is a strong need for future work to be more comprehensive. The supply chain view also enabled us to surface concrete factors that impact information needs, and diverse, complex challenges to be tackled in the future. Our preliminary insights provide a range of methodological lessons for future work.

- The complex relations between the "what", "who", and "why" echo the need to investigate XAI and transparency in context, without solely focusing on the information items (i.e., the "what"). This reaffirms a call for future work on XAI and transparency to be grounded within the "who" and "why", as proposed, e.g., by the AI FactSheet methodology [22, 20]. While many XAI and transparency tools have been proposed (e.g., [12, 17, 5, 2]), they focus solely on the "what" and require revisiting in light of our stakeholders' needs and concomitant challenges.

- Our study emphasizes the need for methodological triangulation to develop a holistic understanding of stakeholders' perceptions and needs across the AI supply

chain. XAI and transparency are complex constructs with a nuanced and overlapping scope. They are perceived differently by those in different positions within the supply chain, present great terminological ambiguity (not to be mistaken with diversity), and evoke different information items, stakeholders, and purposes. In a sense, they are boundary objects to discuss information exchanges and needs across stakeholders. In our study, we used these constructs as starting points for discussions. We observed that AI stakeholders referred to different information needs for broader goals while discussing their current work and concerns. E.g., knowledge of data access rights per customer was discussed by AI developers for training or evaluating the AI system but not in the context of XAI and transparency. Hence, we believe that future studies would benefit from adopting complementary objects of study: studying not only perceptions of XAI, transparency and the needs thereof, but also in-depth information needs of the different stakeholders while refraining from referring to the terms XAI and transparency.

- We recommend bearing in mind stakeholders' different interpretations of AI "explainability" and "transparency" –often confused– and ensuring conceptual clarity in discussions involving these terms. This can be done by asking participants to clarify their use of the terms or providing clear definitions depending on the goal of the studies.

- Our findings can aid the development of a conceptual framework around AI information needs, that could be used to carefully craft further studies. Such a framework could start with the identified who, what, and why of AI information needs, and also encompass personal obstacles, organizational challenges, tensions and preferences.

Policy, Explainability, and Transparency.

Our results finally point out the impossibility of solely relying on stakeholders' perceived information needs to define XAI and transparency obligations. Indeed, this topic was revealed to be highly dependent on stakeholders' knowledge and on their subjective judgment of the various tensions highlighted – many information items found in prior work were not mentioned by any participant, with or without a technical role or background, despite all of them being decision makers across the supply chain and about explainability and transparency considerations. The identified informational challenges hint at the potential undesirability to answer any information need. Hence, "who" should be responsible for making such decisions remains to be explored both within organizations and on a regulatory level.

REFERENCES

[1] Moritz Altenried. 2020. The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class* 44, 2 (2020), 145–158.

[2] 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.

[3] 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.

[4] 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.

[5] 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).

[6] 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.

[7] Jennifer Cobbe, Michael Veale, and Jatinder Singh. 2023. Understanding accountability in algorithmic supply chains. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency.* 1186–1197.

[8] Julien Delaunay, Christine Largouët, Luis Galárraga, and Niels Van Berkel. 2023. Adaptation of AI Explanations to Users' Roles. In *Workshop on Human-Centered Explainable AI (@ CHI 2023).*

[9] 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 *Designing Interactive Systems Conference 2021.* 1591–1602.

[10] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021a. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–19.

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

[12] 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.

[13] Amy K Heger, Liz B Marquis, Mihaela Vorvoreanu, Hanna Wallach, and Jennifer Wortman Vaughan. 2022. Understanding Machine Learning Practitioners' Data Documentation Perceptions, Needs, Challenges, and Desiderata. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–29.

[14] 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.

[15] Markus Langer, Daniel Oster, Timo Speith, Holger Hermanns, Lena Kästner, Eva Schmidt, Andreas Sesing, and Kevin Baum. 2021. What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence* 296 (2021), 103473.

[16] Uli Meyer and Ingo Schulz-Schaeffer. 2006. Three forms of interpretative flexibility. (2006).

[17] 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.

[18] Milda Norkute. 2021. AI explainability: Why one explanation cannot fit all. In *ACM CHI Workshop on Operationalizing Human-Centered Perspectives in Explainable AI (HCXAI)*.

[19] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M Bender, Emily Denton, and Alex Hanna. 2021. Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns* 2, 11 (2021).

[20] David Piorkowski, John Richards, and Michael Hind. 2022. Evaluating a methodology for increasing AI transparency: A case study. *arXiv preprint arXiv:2201.13224* (2022).

[21] Alun Preece, Dan Harborne, Dave Braines, Richard Tomsett, and Supriyo Chakraborty. 2018. Stakeholders in explainable AI. *arXiv preprint arXiv:1810.00184* (2018).

[22] John Richards, David Piorkowski, Michael Hind, Stephanie Houde, and Aleksandra Mojsilović. 2020. A methodology for creating AI FactSheets. *arXiv preprint arXiv:2006.13796* (2020).

[23] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. 2017. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825* (2017).

[24] 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.

[25] 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.

[26] David Gray Widder and Dawn Nafus. 2023. Dislocated accountabilities in the "AI supply chain": Modularity and developers' notions of responsibility. *Big Data & Society* 10, 1 (2023), 20539517231177620.

[27] Mireia Yurrita, Agathe Balayn, and Ujwal Gadiraju. 2023. Generating Process-Centric Explanations to Enable Contestability in Algorithmic Decision-Making: Challenges and Opportunities. *arXiv preprint arXiv:2305.00739* (2023).