Mind The Gap: Designers and Standards on Algorithmic System Transparency for Users

Bianca G. S. Schor
Compliant & Accountable Systems Group
University of Cambridge
United Kingdom
bgss2@cam.ac.uk

Chris Norval
School of Natural and Computing Sciences
University of Aberdeen
United Kingdom
chris.norval@abdn.ac.uk

Ellen Charlesworth
English Studies
Durham University
United Kingdom
ellen.m.charlesworth@durham.ac.uk

Jatinder Singh
Compliant & Accountable Systems Group
University of Cambridge
United Kingdom
jatinder.singh@cl.cam.ac.uk

ABSTRACT

Many call for algorithmic systems to be more transparent, yet it is often unclear for designers how to do so in practice. Standards are emerging that aim to support designers in building transparent systems, e.g by setting testable transparency levels, but their efficacy in this regard is not yet understood. In this paper, we use the 'Standard for Transparency of Autonomous Systems' (IEEE 7001) to explore designers' understanding of algorithmic system transparency, and the degree to which their perspectives align with the standard's recommendations. Our mixed-method study reveals participants consider transparency important, difficult to implement, and welcome support. However, despite IEEE 7001's potential, many did not find its recommendations particularly appropriate. Given the importance and increased attention on transparency, and because standards like this purport to guide system design, our findings reveal the need for 'bridging the gap', through (i) raising designers' awareness about the importance of algorithmic system transparency, alongside (ii) better engagement between stakeholders (i.e. standards bodies, designers, users). We further identify opportunities towards developing transparency best practices, as means to help drive more responsible systems going forward.

CCS CONCEPTS

• **Social and professional topics** → **Computing profession**; • **Human-centered computing** → **HCI theory, concepts and models**; • **Computing methodologies** → **Philosophical/theoretical foundations of artificial intelligence**;

KEYWORDS

design practice, standards, transparency, algorithmic systems, human-AI interaction, design guidelines, artificial intelligence

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642531

ACM Reference Format:
Bianca G. S. 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 (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3613904.3642531

1 INTRODUCTION

As algorithmic systems can have a demonstrable impact on individuals and entire communities [2, 58], there is an increased demand for more transparency over these technologies [15, 40]. Improving algorithmic system transparency—i.e. better transferring information from the system or its designers to a stakeholder about its action, operation, outcome, and potential impact—can support accountability [13, 76] and contribute to re-balancing information and power asymmetries between technical and non-expert stakeholders [75]. Laws mandating more transparent algorithmic systems are being drafted across the world [14, 56, 72], however designers sometimes struggle to realise and accord with transparency and other legal requirements [54]. One reason is that it is often unclear *how* they can do so in practice [54].

To effectively promote algorithmic system transparency, it is important to understand the perspective of designers in charge of building the systems, the challenges they are facing, and what would help them improve transparency. Recent HCI and Responsible AI (RAI) literature has identified a 'principles-to-practices' gap and called for more research on designers' experiences in the wake of the growing calls for RAI [49, 66, 74]. Addressing such a gap is important to assist with building more responsible algorithmic systems going forward. As well as with other resources, standards can contribute to guiding the system development process [4]. Standards can be used by lawyers, procurement specialists, and auditors to set specifications, e.g. to increase safety and interoperability. The role of standards is acknowledged, and their recommendations widely adopted in fields such as finance and computer-aided design [4, 73]. New standards are emerging in order to set measurable levels of transparency for algorithmic systems (§2): published in March 2022, the Institute of Electrical and Electronics Engineers 'Standard for Transparency of Autonomous Systems' (IEEE 7001) [33] is among the first specifically focusing on system transparency. This standard explicitly targets designers among other audiences [33, 78] and by setting guidelines it can thus influence their design. While such standards may have the potential to promote algorithmic system transparency, we hypothesise that there is a gap between the guidance that they provide and designers' understanding of what is helpful to increase the algorithmic transparency of their systems.

We therefore use IEEE 7001 to explore designers' understanding of algorithmic system transparency, and the degree to which their perspectives align with the standard's recommendations. Indeed, designers are cited among the target audience of this transparency-focused standard [78]. Like in IEEE 7001, we define designers as: "designers, developers, builders, maintainers, and operators, as well as decision-makers and procurers in organizations using and deploying autonomous systems" [33]. In line with recent HCI and RAI research [15, 49, 53, 55, 66, 74, 80], we specifically investigate designers' perspectives in relation to new transparency standards by exploring *(i)* designers' understanding of, and experiences building, transparent algorithmic systems; *(ii)* the degree to which their perspectives align with the standard's recommendations; and *(iii)* some key challenges and mechanisms that can support designers alongside standards in making more transparent systems.

For this, we undertake two studies with a mixed-method approach: first, we conduct an online questionnaire to survey designers' experiences and understanding of system transparency (§5.1). We then interview designers about various recommendations, scenarios, and definitions from IEEE 7001 (§3.2) and assess whether they find them understandable and appropriate to build transparent systems (§5.1). We use IEEE 7001 as an exemplar standard to explore designers' perspectives because *(i)* it directly describes designers as part of its audience and its "measurable [and] testable" approach to transparency makes it a practical example to explore designers' current practice (§3.2); *(ii)* according to its definition of designers, designers are among others directly in charge of promoting transparency in algorithmic systems [33, 78]; *(iii)* it is one of the first of few standards published on system transparency; *(iv)* it is intended as applicable across application domains (e.g. healthcare, education, transportation) and types of algorithmic systems; and *(v)* the IEEE is a prominent international professional organisation [30].

Note that transparency by itself is not a panacea to the issues associated with algorithmic systems [2, 13, 45, 69]; however, it is important as **transparency provides a basis for understanding, critiquing, and scrutinising algorithmic systems** [13, 76]. For transparency to be effective, it must be contextually appropriate, for a specific stakeholder in the particular situation [13, 52, 75, 76]. As different definitions of transparency cater for different people's needs [18, 75], we use IEEE 7001 definitions with a focus on non-expert end-users (see glossary A.1): *transparency* meaning "a transfer of information from a system or its designers to a stakeholder, which is truthful, contains information relevant to the causes of some action, decision or behaviour and is presented at a level of abstraction and in a form meaningful to the stakeholder" [32]; and *users* defined as: "persons who have only a brief interaction or who interact every day with an autonomous system" [33].

Note also that this paper is not intended an assessment of the IEEE 7001 standard per se, nor do we imply such standards are to be seen as design manuals for designers. Instead, we explore current design practices and IEEE 7001 transparency recommendations for users to in order to assess and expose any gap between the two. In doing so, we contribute towards promoting *algorithmic system transparency design* by : *(i)* better understanding designers' perspectives and experiences making transparent systems, *(ii)* describing the degree to which their perspectives align with the standard's recommendations, *(iii)* identifying some key challenges and mechanisms that can support designers alongside standards in making more transparent algorithmic systems, and *(iv)* identifying opportunities towards developing transparency best practices.

2 BACKGROUND

To give context, we first explore the literature on algorithmic system transparency before discussing the role of designers and design tools for implementing it. We then provide an overview of recent standards in human-centred design.

2.1 Algorithmic system transparency: legal requirements, definitions, and limitations

There are increasing calls for transparency in algorithmic systems. Transparency is among the most cited values in ethical AI guidelines [40], described by the OECD as supporting democracy and trust [26]. Transparency about algorithmic systems can assist accountability, by allowing scrutiny, contestation, and response to individual decisions and actions of the system and associated processes [13, 76]. In this way, transparency forms a basis for understanding and critique. Ehsan et al. [18] show how social transparency may help calibrate users' trust in algorithmic systems and improve decision-making. Moreover, laws are emerging worldwide that mandate algorithmic systems to be more transparent for various stakeholders. These include the proposed EU 'AI Act' [14], and the US 'Platform Accountability and Consumer Transparency Act' [56]. However, their practical implications for algorithmic system design is unclear [7], thus confirming the need for further practical guidance for designers.

Defining what constitutes transparency is complex. The literature on algorithmic system transparency is growing and spans technical [8, 79], legal [2, 17], ethical and policy oriented [40, 44], and interdisciplinary work [5, 46]. Even within computer science, transparency about algorithmic systems is a broad umbrella term that includes different interpretations [6, 46, 52]. However, it is generally acknowledged that transparency should be appropriate for a specific stakeholder group [52, 75]. Towards this, Ehsan et al. [18] argue for "social transparency," i.e. an explanation of AI-mediated decision-making that incorporates the socio-technical context. Alongside [13, 18, 46], we take a human-centered and systemic approach to transparency and argue the algorithmic system or its designers should provide, according to the stakeholder's needs, information on data, goals, outcomes, compliance, influence, usage as well as the algorithms employed [44]. The IEEE 7001 definition provides a starting point for our investigation (see §1). In this definition, designers are key actors in implementing transparency, which further calls for the need for them to take action.

Again, transparency is not an end in itself, and can even be detrimental in some instances [2, 13, 15, 45]. Transparency can clash with other central values of ethical algorithmic systems, such as privacy, or sometimes lead to other issues such as state surveillance [2, 23, 47, 75]. In order to avoid the "transparency fallacy" [7] or the illusion that transparency can resolve all the issues associated with algorithmic systems such as biases and discrimination, algorithmic system transparency must be considered as a means towards other ends. For accountability [77], for instance, this requires that transparency be *contextually appropriate*: a) relevant to the kinds of accountability needed, b) correct, complete, and representative, c) proportionate to the level of information each stakeholder needs, and d) comprehensible by them [13, 52]. Indeed, Kizilcec [43] shows that providing too much information can erode trust in algorithmic interfaces, a phenomenon Stohl et al. [69] call the "transparency paradox:" high availability of information can produce opacity. It follows that there is a clear need for research and guidance on how to best and meaningfully implement algorithmic system transparency in practice.

2.2 Implementing transparency for users in practice: the role of designers and design tools

Mere calls for more transparency have been shown to be inefficient [22, 82]. Designers thus have a key role in ensuring transparency is implemented in a contextually appropriate manner. Corbett and Denton [15] call for 'more research that moves "beyond abstractions and formalisms to drill down into the specific legal, institutional, historical, political, and cultural contexts. ..." where transparency is being applied.' Indeed, transparency mechanisms in algorithmic systems can be ineffective or underused in practice [18, 48, 61, 83]. To close the gap between explainable AI and users' transparency needs, Liao et al. [48] take a designer-centered approach and interview UX and design practitioners working on AI systems to understand current practices. Yildirim et al. [80] take a similar approach to understand practitioners' perceptions around human-AI guidelines, in particular Google's 'People + AI Guidebook,' which includes a chapter on explainability [60]. Likewise, Wang et al. [74] stress that UX design practice has evolved in relation to growing calls for Responsible AI, and that such emerging practices and associated challenges deserve further academic attention. Our paper builds on research that focuses on designers' perspectives [34, 65], by exploring designers' understanding of building transparent algorithmic systems for users in relation to a new standard in the field.

Various design principles, tools, and frameworks have been suggested to support designers in making algorithmic systems more transparent: "transparency enhancing tools" [39, 42]. Prominent mechanisms include datasheets for datasets [24], model cards [50] and fact sheets [67], which provide information about the models employed. Large technology companies such as Google and IBM also provide guidelines for AI development, including on explainability [29, 60]. The principle of "transparency by design," i.e. transparency built within the system from an early stage, aims to reconcile the "tension between transparency as a normative ideal and its translation to practical application" [22]. Though different in scope and methodology, these attempts aim to facilitate specific types of transparency in algorithmic systems, and generally tend not to target (end) users. More guidance is thus needed [40], especially as designers still struggle to implement transparency and other legal requirements into systems [53, 54]. This gap between available guidance and design practicalities motivates our investigation of designers' perspective related to standards promoting algorithmic system transparency for users, such as IEEE 7001.

2.3 Standards for human-centred design and transparent algorithmic systems

Human-centered AI has been increasingly studied by HCI and AI researchers [1, 11, 20, 64]. Ehsan et al. emphasise the need to operationalise human-centered perspective in XAI "to produce actionable frameworks, transferable evaluation methods, concrete design guidelines, and articulate a coordinated research agenda for XAI" [19, 21]. In this context, standards are emerging that provide overarching guidance for building more transparent systems, and thus deserve attention. The role of standards is primarily to guide stakeholders in adopting what is considered best practices, usually with general directions and requirements [12, 73]. Standards are often derived from or respond to innovative technology, for example in manufacturing and computer hardware, but can also influence the development of technology [4]. Guidelines and directives are published by a number of official standards bodies, such as the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), and the British Standards Institute (BSI) [35, 70, 71], Among them, IEEE has a global academic and technical community [30].

Standards are emerging in relation to human-centred design, and increasingly encompass systems' socio-technical contexts, as well as ethical considerations [31, 37]. In November 2022, the AI Standards Hub [28] listed nine published and five pre-draft standards in relation to human-computer interaction and human-centred design in its database. ISO/IEC 30150-1 'Information technology. Affective computing user interface (AUI) - Model' for example provides "a systematically defined model for affective computing user interfaces (AUI) and topics for AUI standardization" [36]. The IEEE also launched a series of 'human standards' focusing on the ethics of autonomous systems [12].

Some recent standards explicitly mention designers' role in building transparent systems, in particular IEEE 7001 [78], but few have been finalised and published yet. The UK's key body upholding information rights (the Information Commissioner's Office) and the UK's national institute for Data Science and AI (the Alan Turing Institute) have published guidance, aiming to provide "practical advice to help explain the processes, services, and decisions delivered or assisted by AI, to the individuals affected by them" [57]. The pre-draft ISO/IEC AWI 12792 'Information technology—Artificial intelligence—Transparency taxonomy of AI systems' likewise provides a taxonomy of information elements to help identify transparency needs in AI systems [38]. Published in November 2021 by the Central Digital and Data Office in the UK, the 'Algorithmic Transparency Standard' is one of the first published standards promoting transparency specifically, alongside IEEE 7001 [16]. IEEE 7001 is particularly relevant as one of the first international and umbrella standards that spans all algorithmic systems, whilst outlining specific and measurable levels of transparency for users of such systems. Moreover, IEEE 7001 explicitly mentions designers as part of its audience [33, 78]. We therefore use it to probe designers about their understanding of how to design more transparent systems.

3 IEEE 7001 'STANDARD FOR TRANSPARENCY OF AUTONOMOUS SYSTEMS'

We now describe IEEE 7001 and its transparency recommendations for users before using them to explore designers' understanding of transparency in the following sections.

3.1 IEEE 7001 as an exemplar standard on system transparency

IEEE 7001 is a prime example of standards on algorithmic system transparency, and one of the first to directly cite designers as responsible for developing transparent algorithmic systems. IEEE 7001 is therefore clearly relevant for the design community, and might also influence future standards, as it is indeed conceived to allow more domain-specific standards to complement it, e.g. in clinical AI [78]. In this paper, we focus on the designers' perspective given they are explicitly targeted by IEEE 7001, though note that they constitute but one group among many that could directly use and benefit from the standard [33, 78]. Because designers are cited as part of its audience, IEEE 7001 enables us to probe their understanding of algorithmic system transparency in relation to its recommendations. IEEE 7001 provides overarching guidelines for virtually all algorithmic systems, across application domains, and also has a potential resonance with the global professional IEEE community, i.e one of the largest technical organisation for engineering, computing, and technology information worldwide [30]. Further, IEEE 7001 enables the evaluation of transparency for (end) users, its recommendations categorised by stakeholder group. While the standard was published in March 2022, we used the August 2021 P7001 draft, as available on the IEEE website in September 2021, our studies taking place between Sept-Dec 2021. There were no significant changes between the draft and the published standard relevant for this work [32, 33].

3.2 IEEE 7001's "measurable, testable levels of transparency" for users (TLs)

One of IEEE 7001's key contributions is the introduction of six transparency 'levels' for users of autonomous systems, from 0 to 5, to standardise and audit system transparency at scale. Each level "*is a requirement, expressed as a qualitative property of the system which must be met,*" level 5 representing supposedly the greatest level of system transparency of all six levels. Since "*[all] levels are judged to be technically feasible while each successive level is typically more challenging,*" this paper tests them in relation to designers' experiences and practice, as they are core recommendations set out by IEEE 7001. We probe designers about the **'transparency levels for users' (TLs)** as described in Table 1. To discuss designers' understanding of the TLs, we use the IEEE 7001 definitions of 'transparency,' 'non-expert users' (users), and 'autonomous systems' (Table 2). We next detail our mixed-method approach.

Table 1: Transparency levels for users (TLs)

TLs	Definition
TL0	"barely or no transparency"
TL1	"accessible information including scenarios and general principles of operation"
TL2	"interactive training material"
TL3	"a functionality to get a brief and immediate explanation of the system's most recent activity"
TL4	"a functionality to get a brief and immediate explanation of the system's activity in a given situation"
TL5	"a continuous explanation of the behaviour, which adapts to the user's needs and context"

Table 2: Definitions of non-expert users (users), transparency, and autonomous systems based on IEEE 7001

Key term	Definition
Non-Expert Users	"persons who have only a brief interaction or who interact every day with an autonomous system."
Transparency	"a transfer of information from an autonomous system or its designers to a stakeholder that is truthful; contains information relevant to the causes of some action, decision, or behavior; and is presented at a level of abstraction and in a form meaningful to the stakeholder."
Autonomous System	"a system that has the capacity to make decisions itself in response to some input data or stimulus with a varying degree of human oversight or intervention depending on the system's level of autonomy."

4 METHODS: EXPLORING STANDARDS & DESIGNERS' APPROACHES TO SYSTEM TRANSPARENCY

As discussed, the question of how to effectively improve transparency for users of algorithmic systems remains a challenge. Here we describe our mixed-methods approach to identify support mechanisms for designers to develop more transparent algorithmic systems for users.

4.1 Mixed-method study design

We explore the potential gap between design practice and standards promoting transparency by investigating *(i)* designers' understanding and experiences building transparent systems (Study 1), *(ii)* to what degree their perspectives align with the standard's recommendations (Study 2), and *(iii)* some key challenges and mechanisms that can support designers in improving system transparency alongside standards (Study 1 and 2). Our first study (Study 1) surveys designers' approaches to design for system transparency with an online questionnaire, while our follow-up interviews (Study 2) test their understanding and opinions of IEEE 7001 transparency recommendations for users. In doing so, we do not assess how standards ultimately impact users – i.e. those using the algorithmic system. Rather, we explore designers' perspective on designing transparency for users in relation to these guidelines. Although IEEE 7001 does not provide granular design guidance nor advice on how to apply its recommendations, we use two scenarios from its appendices as examples of when the standard can be applied for system transparency specification [32]. These enable us to better contextualise the discussions with interviewees around specific systems and transparency issues.

4.2 Participant recruitment: designers as one group among IEEE 7001's target audience

We focus here on designers as one stakeholder group explicitly cited as the target audience of IEEE 7001: "The target audience of this standard are those designers, developers, builders, maintainers, and operators, as well as decision-makers and procurers in organizations using and deploying autonomous systems *(collectively, "designers")* of autonomous systems who either wish to or are required to engineer systems that have a certain degree of transparency" (emph added) [33]. We recruited survey and interview participants in line with this IEEE 7001 broad definition of 'designers,' by advertising our studies through entrepreneurial, design, academic, non-profit, and professional networks, via LinkedIn, previous colleagues, and direct invitations, using the following criteria, as self-described by our participants: *(i)* the industry, size, and type of organisation, *(ii)* their main role in designing algorithmic systems, as they describe it, *(iii)* their gender and experience in this role, *(iv)* their main country of professional activity. Moreover, our participants were involved in designing systems that fall under the standard's definition of "autonomous systems" (see A.1), where *(i)* the standard is meant to encapsulate a wide range of autonomous systems, including systems with relatively little autonomy [33, 78] and *(ii)* our participants self-referred to themselves as building autonomous systems (see examples below).

We collected 50 questionnaire responses (10 of which were discarded as they only answered demographic questions) and interviewed 22 participants among these (see Table 3). Out of n=40 surveyed participants, 8 identify as developers, 8 product/business managers, 7 designers, 6 engineers/technical managers, 5 academics and 5 entrepreneurs (n=1 did not answer). These categories coincide broadly with IEEE 7001 definition of 'designers'. Most work in technology companies (n=20), mainly in Europe (n=26) or the UK (n=8), and declare their main responsibilities are to design (n=25) or develop/build (n=17) systems (see Fig. 3 in A.3). Out of n=22 interviewees, 13 work in tech startups and small businesses, 3 in public sector or academia, 3 in tech consultancy, and 3 in finance. 10 are engineers or data scientists, 9 are UX/UI or product designers, and 3 are business entrepreneurs. Examples of systems developed by surveyed participants include: glucose prediction based on patient (glucometer), automatic soil analysis interpretation, and content moderation systems, and our interviewees also mention having experience developing among others: mobile gaming apps, IoT devices including audio or visual recognition, and ML-based hiring and education services. Such systems fall under the IEEE 7001 definition of autonomous systems (see A.1). Note that we do not claim to represent the perspective of all possible users of IEEE 7001 with this sample, but rather to illustrate some of the themes and challenges raised by designers in charge of various aspects of algorithmic systems. Participants were offered into a randomised draw for two online vouchers as compensation, one for each study.

4.3 Study one (online survey): designers' understanding and experiences of system transparency

We first use an online questionnaire to explore designers' understanding and experiences building transparent systems for users, as well as some key challenges they might be facing and mechanisms to support them. The 48-question questionnaire (see Supplementary Material) probes participants about: *(i)* their understanding of system transparency for users, *(ii)* their experiences, approaches, and the challenges they have met to design system transparency for users, and lastly, *(iii)* their understanding of the transparency levels (TLs), how to design them in practice, and the TLs' potential impact in system design. We randomise the TLs being considered whenever appropriate. Open text box questions also include contextual information on each TL, copied from IEEE 7001. All questions are optional and some are conditional, i.e. only appear to respondents who have selected specific answers. While all participants are asked about all TLs, each participant is only asked, at random, about how to design: *(i)* either TL1 and 3 (respectively "accessible information including scenarios and general principles of operation" and "a functionality to get a brief and immediate explanation of the system's most recent activity") or *(ii)* TL2, 4, and 5 (i.e. "interactive training material," "a functionality to get a brief and immediate explanation of the system's activity in a given situation," and "a continuous explanation of the behavior, which adapts to the user's needs and context"). Thus, no participant was shown the total 48 questions. This is to enable a wider exploration of issues, while limiting the length and cognitive load for participants. A survey response is taken into account in our analysis when at least one non-demographic question has been answered.

4.4 Study two (interviews): exploring designers' opinions on transparency levels for users in context

To further investigate the extent to which designers' perspectives align with IEEE 7001 recommendations, as well as some key challenges and mechanisms to increase transparency for users, we conducted follow-up individual interviews of ~30-45 mins on Zoom video conferencing software [84] with all 22 survey participants who agreed to be contacted (see Table 3). The interviews started with demographic questions and a discussion of the terms 'transparency' and 'standards,' as well as the potential challenges faced by interviewees when designing system transparency for users. We then suggested an exercise with two scenarios extracted from IEEE 7001 and tested designers' understanding and opinions about TL1, 2 and 3 in these contexts. The scenarios correspond with minimal alteration to the standard's appendixes titled 'Content Moderation for AI' (*Moderation*) and 'Credit Scoring System' (*Credit*) (see A.2). We use the Moderation scenario to ask participants how they would improve the transparency of a video hosting website about its content moderation for content creators; the Credit scenario is about "communicating more transparently the decision-making processes to loan applicants" regarding a credit scoring system used by a loans company. We chose these scenarios for the potential risks on users they exemplify, for being quickly understandable, and for

Table 3: Interviewees' technology areas, self-described roles, gender, years of experience, organisation sizes and countries of activity. Interviewees are referred as I[number] in the following sections.

ID	Technology Area	Self-Described Role	Gender	Years of experience	Company size	Country
I1	Banking	Business Entrepreneur	Male	4.5	<50	France
I2	Technology Company	Software Engineer	Male	30	>100,000	USA
I3	Finance	Software Engineer	Male	15	<50	France
I4	Recruiting & HR	Software Engineer	Male	5	<50	France
I5	Academia	UX Designer	Female	7	N/A	Poland
I6	Technology Company	Business Entrepreneur	Male	2	<50	France
I7	Design	Software Engineer	Male	7	<1,000	UK
I8	Recruiting & HR	Product Manager	Male	7	<50	Spain
I9	Healthcare	Software Engineer	Male	10	<50	UK
I10	Design	UX/UI Designer	Male	3	<50	France
I11	Online Games	Software Engineer	Female	1.5	<1,000	France
I12	Technology Company	Computer Scientist	Male	3	<50	Spain
I13	Education	Business Entrepreneur	Female	2	<50	France
I14	Finance	Product Manager	Male	3	<100	France
I15	Finance	UX Designer	Female	1	<50	Israel
I16	Consultancy	UX Designer	Female	4.5	<100	France
I17	Consultancy	Product Manager	Male	6	<50	France
I18	Technology Company	Product Manager	Male	9	<50	France
I19	Public Sector	UX Designer	Male	4.5	N/A	France
I20	Academia	Data Scientist	Male	2	N/A	UK
I21	Technology Company	Hardware Engineer	Male	4	<50	France
I22	Technology Company	Data Scientist	Male	7	<10,000	Germany

being significantly distinct. We did not test TL0, 4 and 5 in the interviews as IEEE 7001 do not recommend these TLs for users in these two scenarios. Eight interviews were conducted in English, fourteen in French – to minimise the language barrier.

Each participant was presented a single scenario and a single TL. Twelve interviewees were presented with the scenario then the TL, and the remaining ten interviewees heard about the TL before the scenario. This enabled us to test whether individual TLs were deemed more appropriate in the context of the two scenarios when presented *before* the interviewee could reflect on potential interventions in that context. Each interview consisted of three parts, with a slide deck to introduce the TLs and scenarios. The latter provide context for participants to identify interventions they think could improve system transparency, and to test their opinion regarding the relevance of specific TLs towards this. We then asked interviewees whether the scenario's description on the slide was clear to them and could help them improve transparency for users. This allowed us to test the relevance of the TLs with no major misunderstandings on the scenarios themselves. We finally introduced one TL (TL1, 2, or 3), and asked participants whether this was clear and/or appropriate to promote more transparent systems in this scenario.

4.5 Data analysis

We analysed the survey open text box answers (§5.1) and the interviews separately with Braun and Clarke's thematic analysis, a six phase method to identify core themes in the designers' written comments and speech [10]. The six phases (familiarising with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, producing the report) were completed over four steps and two iterations. We first extracted codes (phases 1 and 2) from the interview transcripts and notes. We then watched the interview video recordings, and reviewed the survey answers to group the codes into main themes (phase 3). We finally compared and merged the themes from both studies (phase 4), before repeating the process. These are presented in Table 4 and our result section include numerical accounts of the key points raised by participants as well as detailed quotes illustrating these (phases 5 and 6). For clarity, we have grouped these into 'overarching categories' in Table 4. We counted as a theme any topic raised by more than three participants in either studies. We used Qualtrics [62] to deliver and analyse the questionnaire, and Otter.ai [59] for interview transcripts. This research was approved by our institutional research ethics committee.

5 RESULTS

To promote transparency in practice, we explore designers' understanding of algorithmic system transparency, and the degree to which their perspectives align with IEEE 7001 recommendations. Towards this, we survey participants' experiences and interview them to probe their opinions on how appropriate these recommendations are in two scenarios provided by the standard. We sampled participants in line with IEEE 7001 definition of designers and our mixed-method study provided consistent results overall: the gap between IEEE 7001 and designers' perspective is important. We now describe this gap and some key challenges and mechanisms to promote transparency in practice.

5.1 Highlights from our findings

Study 1 explores designers' understanding and experiences making transparent systems for users. Table 4 summarises the key themes raised by our 40 surveyed designers and 22 interviewees, according to our thematic analysis of open text box answers (survey) and speech (interviews). We elaborate on these themes in the following subsections. The survey shows n=30 out of 36 agree that designing transparency for users is important (n=4 say they neither agree nor disagree and n=2 somewhat disagree with the statement). Further, n=23 out of 36 think it is difficult to do so, which suggests a need for supporting designers towards this.

Table 4: Summary of the key themes raised by designers in the survey (Study 1) and interviews (Study 2)

Overarching category	Themes
What are standards?	provide guidelines; set expectations; rarely used; GDPR
Which stakeholders?	providers; end-users
What systems & risks?	autonomous; opacity; privacy risks
What is effective system transparency?	what is "good" transparency?; which granularity?; good UX; how to define it?; what for? rarely think of it; data protection; context-specific; difficult to implement
What are the key challenges?	clash with other priorities; lack of resources; design challenges; technical challenges; defining the terms; lack of awareness; lack of incentives
What design interventions can help?	communication; legal solutions; design solutions; organisational solutions; technical solutions; empathy

Study 2 then investigates the extent to which designers' perspectives align with IEEE 7001 guidelines. The 22 follow-up interviews reveal the gap between IEEE 7001 and designers' approaches to system transparency is important: despite designers understanding the scenarios overall, they do not find the individual TLs recommended in IEEE 7001 appropriate in such contexts. Fig. 1 represents the results from each interview as a pair of one circle and one square, with a green-to-red colour scale. The large prevalence of orange and red squares highlights that only four designers found TL1-3 appropriate (which entail respectively "accessible information including scenarios and general principles of operation," "interactive training material," and "a functionality to get a brief and immediate explanation of the system's most recent activity"), whereas 50% found them appropriate only under certain conditions, and 32% did not. This suggests that interviewees are unlikely to implement these IEEE 7001 recommendations in similar contexts; and again, that more support seems needed to bridge the standard and designers' approaches in practice. Moreover, the interviews confirm participants face multiple challenges in implementing algorithmic system transparency for users, and yet they are able to suggest design principles towards it.

5.2 Few participants know of or use standards to design user-facing algorithmic systems

In addition to investigating designers' understanding of IEEE 7001, we probe them on their general understanding and use of standards. Almost none of our 22 interviewees knows of or uses standards for the user-facing aspects of algorithmic system design. Eighteen do not use standards to design user-facing algorithmic systems and ten do not seem to know what an official standard is ("I don't know what standards are" (I12), "What do you mean?" (translated, I6), "I don't know any standards" (translated, I14)) or cite the EU General Data Protection Regulation (GDPR) as an example of a 'standard' which is actually a regulation (though it does encourage the use of standards and best practices). None cite standards specifically on algorithmic system transparency, though we note that these were still under development and thus not widely accessible when we interviewed them in late 2021. While this is not surprising—given not all standards target designers directly—this indicates the challenge for IEEE 7001 to impact real-world design practices. Indeed, many participants cite other resources they use to design user interfaces, such as internal guidelines and processes, or external ones such as the Nielsen Norman group's UX best practices (3 interviewees) [25]. These seem more widespread among designers, and could represent a pathway towards improving system transparency. Standards bodies could thus better engage with designers to raise awareness of standards and shape their specifics.

5.3 IEEE 7001 transparency recommendations are critiqued by participants

While our participants do not currently use standards to design user-facing algorithmic systems, they also question the relevance of IEEE 7001 transparency recommendations for users.

5.3.1 Few participants say they are likely to implement the TLs. Only 10 out of 40 surveyed designers declare being likely to provide TL2 ("interactive training material"), TL3 ("a functionality to get a brief and immediate explanation of the system's most recent activity"), and TL5 ("a continuous explanation of the behaviour, which adapts to the user's needs and context"), and n=14 to provide TL4 ("a functionality to get a brief and immediate explanation of the system's activity in a given situation") going forward. On probing whether the TLs are currently implemented, n=17 say TL1 ("accessible information including scenarios and general principles of operation") is provided half of the time, whilst the majority (n=21 to 29 think TL1-5 are rarely implemented). This shows IEEE 7001's impact on systems design could be limited in practice. The following subsections highlight why participants question how appropriate IEEE 7001 recommendations are in a given context.

5.3.2 IEEE 7001's scenarios and TLs are overall clear but not precise enough to enable actual implementation. Both the Moderation and Credit scenarios provided by IEEE 7001 seem overall clear for interviewees, though not for all (Fig. 1). As discussed in §4, these provide examples of when IEEE 7001 recommends TL1-3 for users as system specification. We use the scenarios as contexts to gather designers' opinions on how appropriate TLs' are for users, and thus need to confirm interviewees understand the scenarios and TLs. Two participants find the Moderation scenario rather unclear: "It

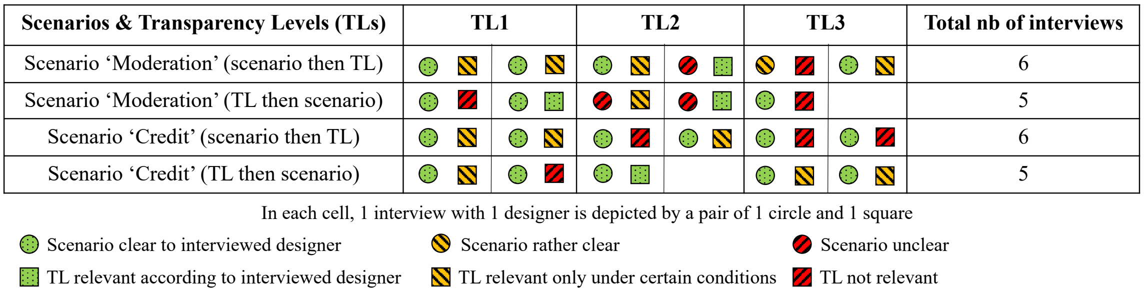

Figure 1: A Summary of the interview results for the Moderation and Credit scenarios and TL1-3. Two combinations (TL2 then Credit scenario; TL3 then Moderation scenario) were tested once due to time constraints.

feels very heavy, vague" (I15), "how do I intervene in this?" (translated, I10). Another explicitly asks us to explain it to them. Likewise, one participant describes the Credit scenario by saying: "based on my experience writing requirements, I would say it's badly written, go back and try again" (I2), which suggests the wording could be clarified. Similarly, TL1-3 are considered clear, though three interviewees find TL2 unclear ("I don't understand what that means. Can you give me an example?" (I15)), and two say TL3 is too abstract: "this is basically saying the solution to transparency in the system is to explain to the user in simple language why a decision has been made in the system, which makes sense, but that's incredibly abstract, (...) what does that system look like? What is the interface? What the actions are that the user does and at what point? (...) what this looks like visually is still not clear." (I7).

Indeed, while participants seem to understand IEEE 7001 scenarios and TLs overall, over half of them think the three TLs do not provide enough information to enable designers to implement them in practice. When asked about TL1, one interviewee points out how unclear it can be to evaluate whether it has been effectively implemented: "there is no information here about what kind of users they are, what things they might need to know, what things they might want to know, how they would best receive information. Some people like to see videos, some people like to read texts. Some people like to be told. Some people like to be shown. Is the user disabled in anyway? There is not enough specificity here to provide ... Have I provided accessible information [i.e TL1], yes or no? Nope, can't answer that question" (I2). Similarly, another participant wonders about TL3: "but which granularity do you go for?" (translated, I21). We acknowledge that many standards are not meant to provide granular design guidance, but their target audience (here designers among others) should still be able to understand what mean in practice the "measurable, testable levels of transparency" IEEE 7001 recommends they reach in their systems, and in this context our findings suggest otherwise. We do not test TL0, 4 and 5 with interviewees because they are not recommended by IEEE 7001 for these scenarios, but our survey results suggest similar findings to the interviewees. In all, IEEE 7001's lack of precision for interviewees does not undermine their ability to discuss the TLs in the two scenarios, but is part of what they find problematic, and could hinder the standard's adoption. Those involved in building standards should take note.

5.3.3 Participants seem to find TLs difficult to implement and not always appropriate in the IEEE 7001 scenarios. Most TLs seem difficult to implement, according to surveyed designers. As a reminder, IEEE 7001 considers TL1 the easiest to implement and TL5 the more challenging. Our survey reflects this only for TL1 and 5 (see Fig. 2), e.g. it shows TL2 is deemed the most difficult to implement after TL5. Only TL1 seems 'easy' for the majority (n=17). Whilst IEEE 7001 provides the Moderation and Credit scenarios as examples where TL1-3 should be implemented, we observe that out of the 11 designers interviewed with the Moderation scenario, the majority is not convinced that TL1-3 are appropriate here (Fig. 1): three say the TLs could be potentially appropriate, but only if designed in a specific way. For instance, one says about TL2: "the important thing is to not make it too heavy because users don't actually care about that, like they just want to get through it, they're not in the mood to read or do a heavy on-boarding, so yeah keep it short and actually useful" (I15). Likewise, two participants say the TLs are not the best options, and another explains TL3 is not appropriate in the Moderation scenario, as it can be "frustrating" for users: "I think having the information afterwards would often just be frustrating. I think it would be like a step up from were we got to but, often like, to get a decision, you have to put in a lot of time, and it's time you could save like if you had more information about what it was actually looking for in the first place. Maybe like insurance, it would cause more people to game the system, which isn't good, like to lie, but the content creation is like, you could literally decide to create a different video." (I20).

The same TLs are judged even less appropriate in the Credit scenario: all interviewees find they are either incomplete, unrealistic, or not adapted to this socio-technical context. One comments about TL2: "I don't really see the point of designing a simulator [for users]" (translated, I16). Another describes TL1 as likely "frustrating" for users: "there is clearly a need to explain to the users why a loan has been denied (...) to me, what's missing here is what to do, the next steps, how to improve my credit scoring, can I do something to change my situation? Because only presenting information can create frustration (...) it's essential to be able to act upon that information, otherwise it's just a justification from the provider, from the company who is backing itself" (translated, I14). Another comments "I think [TL1] is clear, I would question its usefulness," before highlighting the risk it could lead to for users: "assume that you are writing this for a malicious software engineer, who will take this

Perceived Difficulty of Transparency Levels	Total	Extremely difficult	Somewhat difficult	Neither easy nor difficult	Somewhat easy	Extremely easy	N/A
TL1: accessible information including scenarios and general principles of operation	32	0 0%	6 19%	7 22%	11 34%	6 19%	2 6%
TL2: interactive training material	32	2 6%	14 44%	6 19%	5 16%	2 6%	3 9%
TL3: a functionality to get a brief and immediate explanation of the system's most recent activity	31	4 13%	7 23%	7 23%	10 32%	1 3%	2 6%
TL4: a functionality to get a brief and immediate explanation of the system's activity in a given situation	31	4 13%	8 26%	11 35%	8 26%	0 0%	0 0%
TL5: a continuous explanation of the behaviour, which adapts to the user's needs and context	32	12 37%	14 44%	5 16%	0 0%	0 0%	1 3%

Figure 2: Numbers of survey participants show most transparency levels (TLs) seem difficult to provide.

specification and implement it to the letter in the most unpleasant way (...) Suppose I had three or four reasons for explaining why I'm going to say no to a loan, and I choose to pick the one that is going to cause the most offence" (I2). Finally, finds TL3 "pointless" but TL2 more appropriate, as it provides information to the user before they start using the system: "if the goal is transparency [on the use of personal data], I don't understand the point of telling the user after they have given access to their personal data. For me, it should happen beforehand" (translated, I18). These results were consistent regardless of the order in which the scenario and the TL were presented. Thus they outline the challenges of creating and implementing transparency standards, given TLs are central to IEEE 7001 guidelines and TL1-3 are recommended in these scenarios.

5.4 Key challenges for improving system transparency

5.4.1 What 'transparency' actually means is unclear to most participants. Most participants expressed that they find the term 'transparency' unclear. Two designers differentiate transparency between hidden and visible features, whilst three consider 'transparency' good UX and clear communication. For example, one participant stresses the challenge of explaining complex machine learning algorithms to users intelligibly and faithfully: "the first thing I think about is the algorithm, the second thing is the educational message. Transparency and simplification enable to communicate a complex message to non-expert users (...) but you have to take shortcuts for it to be impactful and digestible and so, as a result, you lose on transparency, paradoxically" (translated, I14). Another acknowledges there are different types of transparency, and cites the UX of an order process vs. the system's ethical/environmental impact: "it depends what transparency we're talking about, like is Amazon transparent on the ethical and environmental implications of your interaction with the system? No, definitely not, no e-commerce website or very few would be. Is it transparent on the behaviour of the system? For example, once you've interacted with it to place an order, what the status of that order is? What the journey of that order is? How you can interact with that order? Yes, in that sense it is very transparent" (I7). After being given the IEEE 7001 definition of 'transparency' (§2) in the survey, one interviewee expressed that they found this definition "incomprehensible" (I2), and several others asked what system transparency means, which confirms the general confusion regarding the meaning and significance of term in this context. Beyond the quality of IEEE 7001 definition, this suggests a wider community understanding of what transparency means in practice is needed.

5.4.2 Participants' awareness about the risks posed by algorithmic systems seems primarily focused on privacy. Most surveyed designers think they are aware of and identify risks posed by their algorithmic systems: n=33 identify their systems as directly impacting users in high-stakes situations, e.g. in employment, healthcare, credit risk etc. Among the risks posed by their systems, the privacy risk for users and their personal data is identified the most often (n=19), followed interpretability risks (n=13), i.e. complex model that may not be fully understood by users. There is a general awareness of the risk of gathering personal data, perhaps due to GDPR's presence in Europe: when asked what information they provide to users to mitigate such risks, almost half of our respondents say they communicate GDPR-compliant information or privacy terms (see Table 4). Our study thus indicates a similar push for change on other issues linked to algorithmic system transparency is still needed.

5.4.3 Participants identify several challenges to design transparency: clash of priorities, lack of resources, and clear definitions. Moreover, designers identified obstacles for system transparency (Table 4, Table 5). The key challenges from surveyed respondents were potential clashes with other priorities (including commercial ones)

Table 5: Non-exhaustive list of additional challenges in designing for transparency, as raised by interviewees.

Other challenges	Examples cited
user or product-specific	fear of using technology
difficult choice of wording and format	complex concepts are difficult to explain
users' limited attention span	many users skip onboarding
lack of specification to improve transparency	misunderstandings on what transparent interfaces mean
lack of empathy/awareness among internal stakeholders	users' needs are not always prioritised
quantifying the added value of designing for transparency	lack of clear metrics to convince hierarchy

(n=6), followed by the lack of financial means (n=3). Some mention the technical challenge of providing explanation for automated decisions, and the design challenges of "understanding the end-users difficulties" and "getting the right visual design." When asked about TLs specifically (Table 6), designers raise the lack of time, a clash with other priorities, the lack of designing tools and methods, and the difficulty of defining terms, e.g. 'transparency'. The interviews confirm this: almost all designers raise at least two challenges to implement transparency in user-facing algorithmic systems. Many match those identified in the survey (Table 6), others are summarised in Table 5.

Table 6: Most surveyed designers raise challenges to implement IEEE 7001 Transparency Levels (TLs).

Challenges to implementing TLs	Nb of respondents
Lack of time	24 (60%)
Clash with other priorities	18 (45%)
Lack of designing tools and methods	13 (32%)
Defining the terms (e.g. transparency)	12 (30%)
Lack of awareness in teams	9 (22%)
Lack of financial means	9 (22%)
Lack of awareness in management	8 (20%)
Other	7 (17%)

5.5 Support mechanisms for designers: raising awareness, providing tools, and outlining best practices

These challenges represent opportunities as to ways forward on supporting meaningful transparency for users; indeed, participants then went on to identify support mechanisms and design principles to promote transparency in practice, as detailed below.

5.5.1 Raising designers' awareness is needed. Our study reveals the need and potential for raising awareness about transparency among designers. Though seven designers say they had never thought about transparency before, many show further interest in transparency after participating in our study: "it makes me want to go and have a look at a couple of websites to see if there is this notion of transparency there and how it materialises in practice, I don't feel that it is, to put it simply, so I am going to have a look, actually, straight after this" (translated, I21). Half of the participants say they feel curious or more aware about transparency issues in the interviews ("it awakens me to this topic that I do not see from this angle" (translated, I18), "it's made me think about what we mean by transparency" (I7)), and two even shared transparency-related news and resources via email after the interview (including a a screenshot of YouTube's change of general conditions, where a paragraph about 'transparency' had been circled by the designer). Three participants also raise doubts about whether making systems more transparent for users is part of their role, or more broadly raise the question of who is responsible for it, whereas others state that all the stakeholders should contribute to implementing system transparency for users, ("that would be the job of the UX designer (...) maybe in tandem with the product manager?" (I7)) and that transparency considerations are relevant for all aspects of system design: "it's always interesting because then you can take these principles, you can apply them in different aspects, whether this is designing of a web page, designing of a, I don't know, larger system, designing of some back end code" (I9). Thus, further research toward raising designers' awareness of the importance of designing for transparency, like during our interview, as well as incentivising them, and better defining the key stakeholders responsible for developing, maintaining, auditing or regulating it is crucial for improving system transparency. Similarly, raising awareness about new transparency-focused standards could make such standards more effective in practice given these are still recent and many participants declare not knowing or using standards to design user-facing algorithmic systems (§5.2).

5.5.2 Several designers identify help that could support them in designing for transparency. To overcome the challenges raised above, we asked participants to describe what resources could be useful for them to improve system transparency for users. When asked to advise others trying to design for transparency, only seven surveyed designers recommend using standards; they primarily recommend asking more experienced colleagues (n=14) or finding other sources of help (n=10) (see examples below). Some also comment on the need to "work with users to determine what they want to know and what is a suitable format to present that information." We discuss calls for co-creation below. Others emphasise the need to "understand requirements and constraints in the business context," "use empathy" or "community resources," better communicate or "hire an analyst to interpret the model." Moreover, surveyed designers explicitly comment on what could help them to improve system interfaces in practice: they need "examples and best practices," as well as "explanations on how and when to implement them," but also "targets on what is necessary and when; what the end user is in need of; what the company I work for actually wants to provide." Similarly, "a clear description of why we do it, a clear demand from end user" would help. A few respondents also wish for more resources: either more "understanding from [their] management" and of the user's needs, "more time," or "a person dedicated to solving issues [related to algorithmic system transparency]." Lastly, better tools were often cited as required, e.g. frameworks for "quick and easy implementation" to improve system transparency. These requirements confirm the challenges raised above, but also pave the way for standards and research to better support designers in designing for transparency.

5.5.3 Participants have suggestions for tools that they think can assist in making systems more transparent. Participants are able to identify tools that they think could help increase system transparency for users. Table 7 summarises the tools mentioned in the survey. Our interviews confirm these, as all participants suggested interventions they thought might facilitate transparency in algorithmic systems. Given participants' initial uncertainties regarding what transparency means in this context, not all may be relevant, which confirms the need to raise designers' aware about transparency. Yet, our findings indicate that, when given an appropriate understanding of transparency, designers are likely to suggest interventions they think may increase transparency. This is encouraging

Table 7: Tooling suggested by surveyed designers to improve system transparency

Category of tools	Examples cited
user-centric tools	user tests, shadowing, workshops, FAQ, communication material, tutorials, videos, user guides, explanations
internal tools	guidelines, training for employees, writing tools to simplify language, examples of successful approaches
technical features	functional specifications, technical release notes, and user guides
legal tools	audit records, online dispute resolution, contracts, memos on GDPR, GDPR guidelines

as it shows the potential for productive collaborations with designers such as our participants and those producing standards and guidance. Thus supporting them in identifying best practices can offer an opportunity to improve algorithmic systems (§6.1).

5.6 Principles for 'good' transparency in algorithmic systems for users

We further asked surveyed designers as to what transparent user-facing systems might consist of. Table 8 lists the main principles designers identified in the survey's open text boxes. Whilst non-exhaustive nor unanimous, this list summarises themes raised at least twice. Though further assessment of these principles is needed, it can be helpful for identifying best practices and examples of more transparent systems. Note that these recommendations (Table 8) do not match those provided by IEEE 7001 as the standard is not intended to provide granular design guidance. As a reminder, IEEE 7001's main contribution is the transparency levels: TL1 "accessible information including scenarios and general principles of operation;" TL2 "interactive training material;" TL3 "a functionality to get a brief and immediate explanation of the system's most recent activity;" TL4 "a functionality to get a brief and immediate explanation of the system's activity in a given situation;" and TL5 "a continuous explanation of the behaviour, which adapts to the user's needs and context"). This further indicates a gap between what designers and standards consider appropriate to promote transparency. Such 'bottom-up' information, as provided (and desired) by practitioners at the 'coal-face', is useful for those—be they practitioners, standards bodies, academics, and policy-makers—that strive towards increasing algorithmic system transparency in practice.

6 DISCUSSION

Mere calls for increased system transparency can be ineffective in practice [22, 82]. However, transparency does have an important role to play – it can be helpful in supporting accountability, fairness, scrutiny, contestability and other aims [2, 13, 76, 81], which work towards managing the risks of algorithmic systems. By exploring designers' understandings and experiences regarding algorithmic system transparency as well as their opinions and perspectives on the specific transparency recommendations from IEEE 7001, we have highlighted some of the key challenges and principles for improving system transparency alongside standard guidelines.

Table 8: Principles for 'good' transparency over algorithmic interfaces according to surveyed designers

For designers, 'good' systems provide information that…
* is clear and easy to understand
* is complete and not hidden on the user interface
* is accountable or certified
* gives confidence and trust to users
* is accessible and gives the correct level of information
* is simple, intuitive, playful, and empathetic
* enables to configure the system, override it, and identify who is responsible for its impact
* is compliant with personal data protection
* enables informed decision-making and interaction
* is given at the right moment in the user journey

In doing so, we help draw further attention to the 'principles-to-practices' gap, showing that it extends to transparency standards, and indicate the need and opportunities for addressing this gap to help ensure that ongoing efforts (e.g. standards, RAI initiatives) are effective in practice. As one of the first international standards focusing on transparency, IEEE 7001 sets measurable levels of transparency in algorithmic systems according to specific stakeholder groups. Despite its potential for promoting transparency for algorithmic system users, our results show that participants do not find its recommendations relevant in the scenarios it provides as examples. We argue such standards should be part of a wider concerted effort to raise designers' awareness alongside more engagement with stakeholders (i.e. standards organisations, users, designers etc). We now suggest further mechanisms to better support designers before discussing our study's limitations.

6.1 Suggestions for supporting designers in making algorithmic systems more transparent for users

Based on the results, we now outline ways in which standards, raising transparency awareness, and closer collaborations with various stakeholders can better support designers in improving transparency for users.

6.1.1 Raising awareness about transparency among designers. Our study reveals the potential impact of transparency-focused standards, such as IEEE 7001, might be limited until there is more awareness amongst designers generally (i.e. beyond those already engaged in interested communities) of the need and importance of transparency, and of the role of such standards. For example, our qualitative insights highlight how our short online interviews (study 2) raised not only awareness but also interest and curiosity for system transparency among participants (see §5.5.1). We encourage more efforts in this direction, for example through co-design approaches. Co-design has already been tested to help close the 'principles-to-practices' gap, e.g. with an impact assessment framework for responsible AI values [66] and a checklist to operationalise fairness in AI [49]. Likewise, we argue co-design [51, 68] and stakeholder engagement ('participatory AI' approaches [9, 85]) approaches are means not only for improving transparency implementation (where users and other stakeholders are part of the problem-solving endeavour), but also for reaching agreements on what transparency means in practice (with closer connections between standards bodies and designers for example [80]). Indeed, even within the transparency research community, the meaning, value and implementation contexts for 'transparency' are still being debated [15]. In this context, the AI standards hub launched by the Alan Turing Institute, UK, organised two inter-disciplinary workshops in January 2023 for individuals to take part in developing two transparency-focused standards [3].[1] We argue such discussions are needed, including to highlight the limitations of transparency, and how it might clash with other core values of responsible AI, such as security, privacy and data protection [23, 41]. Likewise, standards should be part of a wider concerted effort to raise designers' awareness alongside more engagement with stakeholders, as promoted by the AI standards hub [27].

6.1.2 Further identifying examples for future standards and transparency metrics. Our study, along with a lack of literature in the space, suggests no clear consensus has emerged on what best practices might be so far. And yet one key suggestion from our interviewees is the need to provide them with clear examples of how to design for transparency. One summarises this as follows: "a couple of examples would have been great (...) because, in fact, since I am not used to thinking about it in theoretical terms, as a result the way [TLs] were phrased, I struggled to link it back to what I might have seen, because I'm immediately going to ask myself: what does it look like? I understood all the words, you see, it's just that I didn't manage to link it back to things that I do in practice in UX or in design" (translated, I19). Thus providing more real-life examples in standards might constitute an opportunity to further align them with designers' understanding of transparency in practice, and thus facilitate their implementation. As a result, showcasing examples of good transparency practices could become a lever to improve system designs overall. Our findings regarding transparency characteristics of 'good' systems (Table 8) could potentially inform, support, and perhaps become transparency standards. Such characteristicscould be used to adapt transparency recommendations and bring them closer to designers' perspective, thus helping to operationalise more responsible AI going forward [19, 21].

6.1.3 Supporting designers in experimenting with transparency interventions. As our results show, all interviewees were able to come up with design interventions they thought could increase transparency in a given system. We acknowledge their relevance might vary depending on designers' understanding of transparency and experience. However, the design interventions recommended by IEEE 7001 do not seem to be the most effective, according to our participants, in making systems more transparent. This aligns with the idea that transparency becomes *meaningful* only when contextually appropriate [13, 15]. Building upon the ideas raised in §5.5.3 and Table 8, we thus encourage more work on facilitating designers' experimenting with transparency interventions in context. For example, designers seem to report various tools to capture, manage, and collaborate on ideas [34, 65]. These could be used to promote a collective effort to improve transparency in practice. Such design efforts can coincide with Felzmann et al.'s transparency-by-design framework "that can act as a reflection tool for different stakeholders to integrate transparency considerations into their practice" [22].

6.1.4 Engaging various stakeholders more in standards development. One of the key findings from our study is the need for further engagement between various stakeholders, namely standards organisations, designers, and users on ways for making systems more transparent. We identify a gap between standards' recommendations and designers' approaches to improving algorithmic system transparency (see quote in §6.1.2). This confirms the 'principles-to-practices' gap identified in the responsible AI literature [15, 49, 66] extends to standards bodies promoting transparency. Those developing standards could therefore better support designers by engaging with them more closely, and vice-versa, such that the standards are more targeted, relevant, and fit for purpose. This collective approach to promote transparency has also been highlighted by Rakova et al., who show organisational structures have an impact on the implementation of responsible AI initiatives [63]. One starting point for such efforts could be to focus on addressing the main challenges faced by designers willing to build more transparent systems, such as the ones we identify in §5.4.3. Our study also indicates more engagement with users is key to move forward [48] (see, e.g., §5.5.2). Moreover, several designers have raised doubts about whether making systems more transparent for users is part of their role (see §5.5.1). Connecting practical interventions directly with someone responsible for designing, monitoring, and/or auditing transparency over algorithmic systems and its potential impact could therefore greatly facilitate a general (and practical) understanding of transparency and improve system designs. The recent launch of the AI Standards Hub (Oct 2022) is an example of initiatives and inter-disciplinary communities that aim to bridge the gap between standards bodies, academics and practitioners [27].

6.2 Study limitations

Finally, we identify three areas that can further complement and confirm our findings. First, our sample size is limited due to time constraints, and influenced by the main author's professional and online networks. For example, most of our participants are based in France/Europe: while 75% of surveyed designers come from France/Europe, 20% from the UK, 5% from the Middle East, Asia, North America, and Oceania, and none Africa and South America. Therefore similar studies could be conducted with a larger or different sample, for example with those from the Global South. We also acknowledge that designers are one among other stakeholders explicitly targeted by IEEE 7001. We focus on their approach and do not claim to represent the perspective of all possible users of IEEE 7001, though these deserve attention going forward. We also faced a language barrier in our study: despite speaking at least general English, most participants are not native English speakers (only 18% of our interviewees are) and several raised the language barrier as a factor hindering their understanding of the terms used in the survey. We thus translated orally into French the scenarios and

[1] Pre-drafts 'ISO/IEC AWI TS 6254 Information technology — Artificial intelligence — Objectives and approaches for explainability of ML models and AI systems,' and 'ISO/IEC AWI 12792 Information technology — Artificial intelligence — Transparency taxonomy of AI systems' [3, 38]

TLs for 5 interviewees, who reported they understood them overall. Given the global nature of transparency concerns, it is important that prominent standards are made available and accessible across a broad range of communities. Lastly, this paper is not an assessment of all standards, nor of IEEE 7001 in its entirety. Our aim was to better understand designers' perspective on algorithmic system transparency by probing their understanding of IEEE 7001 recommendations for users. We chose IEEE 7001 for this study, given its high profile, overarching scope, and level of development, as a representative exemplar. We use it to indicate potential real-world issues stemming from algorithmic transparency standards, though we acknowledge other standards have been published since. IEEE 7001 is significant as it is *(i)* one of the first published international standards that seeks to address system transparency specifically, *(ii)* from IEEE which represents a large international professional community, and *(iii)* spans virtually all algorithmic systems. Our study thus paves the way for a broader assessment of IEEE 7001, and studies on other standards to compare and expand our results.

7 CONCLUSION

This paper has explored the gap between standards' recommendations and design practices. It takes a step towards bridging this by better understanding designers' perspectives, suggesting mechanisms to support them in building more transparent systems, and discussing considerations for developing future transparency standards. Closing this gap is important for ensuring more responsible algorithmic systems going forward. Towards this, we have explored designers' experiences of transparency and understanding of IEEE 7001 guidelines ('Standard for Transparency of Autonomous Systems'), i.e. one of the first international standards published in the field. Despite this standard's potential, our mixed-method study shows that IEEE 7001 might not be effectively implemented by designers in practice, due to a mismatch between their respective approaches; e.g. interviewees do not seem to find key IEEE 7001 transparency recommendations relevant in the scenarios provided by the standard. Moreover, most participants seemed eager to learn more about the concept of transparency despite not always being familiar with it, so our study highlights the potential for raising awareness among designers. Whilst we acknowledge transparency is not a panacea, it plays an important role in facilitating understanding, critique, contestation and accountability, providing a means that can help in better aligning systems with the public interest. In light of our findings, we therefore call for collaborative efforts across communities that *(i)* raise designers' awareness of the importance of designing for transparency in algorithmic systems in combination with *(ii)* further engagement between stakeholders (i.e. standards organisations, designers, users etc.). We further contribute by highlighting that more research is needed to *(iii)* solidify definitions, enable a better understanding of transparency, and facilitate its implementation in algorithmic systems, and *(iv)* identify examples, best practices, and users' needs. This paper thus identifies important challenges, but also mechanisms to support designers in making algorithmic systems more transparent for users.

ACKNOWLEDGMENTS

The Compliant & Accountable Systems Group and CN acknowledges the financial support of UK Research & Innovation (EP/P024394/1, EP/R033501/1, ES/T006315/1), The Alan Turing Institute, and Microsoft (via the Microsoft Cloud Computing Research Centre). EC is supported by Arts and Humanities Research Council, Grant number: AH/R012415/1.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 ACM Human and Computer Interaction (CHI) conference on human factors in computing systems (CHI'18)*. Association for Computing Machinery, New York, NY, USA, 1–18.

[2] Rachel Adams. 2020. *Transparency: new trajectories in law.* Routledge.

[3] The Alan Turing Institute AI Standards Hub. 2023. Towards transparent and explainable AI: Workshop to inform ISO/IEC standards development. https://aistandardshub.org/events/towards-transparent-and-explainable-ai-workshop-to-inform-iso-iec-standards-development/

[4] Robert H Allen and Ram D Sriram. 2000. The Role of standards in innovation. *Technological Forecasting and Social Change* 64, 2-3 (2000), 171–181.

[5] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* 20, 3 (2018), 973–989.

[6] Carolyn Ball. 2009. What is transparency? *Public Integrity* 11, 4 (2009), 293–308.

[7] Bettina Berendt. 2022. The AI act proposal: towards the next transparency fallacy? Why AI regulation should be based on principles based on how algorithmic discrimination works. https://people.cs.kuleuven.be/~bettina.berendt/Papers/berendt_2022c_last_author_version.pdf

[8] Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José M. F. Moura, and Peter Eckersley. 2020. Explainable machine learning in deployment. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 648–657. https://doi.org/10.1145/3351095.3375624

[9] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Arlington, VA, USA) *(EAAMO '22)*. Association for Computing Machinery, New York, NY, USA, 8 pages. https://doi.org/10.1145/3551624.3555290

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[11] Tara Capel and Margot Brereton. 2023. What is human-centered about human-centered AI? A map of the research landscape. In *Proceedings of the 2023 ACM Human and Computer Interaction (CHI) Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 23 pages. https://doi.org/10.1145/3544548.3580959

[12] Raja Chatila and John C Havens. 2019. The IEEE global initiative on ethics of autonomous and intelligent systems. In *Robotics and well-being*. Springer, 11–16.

[13] Jennifer Cobbe, Michelle Seng Ah Lee, and Jatinder Singh. 2021. Reviewable automated decision-making: a framework for accountable algorithmic systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 598–609. https://doi.org/10.1145/3442188.3445921

[14] European Commision. 2021. Proposal for a regulation of the European Parliament and the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[15] Eric Corbett and Emily Denton. 2023. Interrogating the T in FAccT. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1624–1634.

[16] Central Digital and Data Office. 2021. Algorithmic transparency standard. https://www.gov.uk/government/collections/algorithmic-transparency-standard

[17] Lilian Edwards and Michael Veale. 2018. Enslaving the algorithm: from a "aight to an explanation" to a "right to better decisions"? *IEEE Security & Privacy* 16, 3 (2018), 46–54.

[18] 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 ACM Human-Computer Interaction (CHI) Conference on Human Factors in Computing Systems (CHI'21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445188

[19] 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 ACM Human-Computer Interation (CHI) Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 6 pages. https://doi.org/10.1145/3411763.3441342

[20] 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.

[21] Upol Ehsan, Philipp Wintersberger, Elizabeth A Watkins, Carina Manger, Gonzalo Ramos, Justin D. Weisz, Hal Daumé Iii, Andreas Riener, and Mark O Riedl. 2023. Human-centered explainable AI (HCXAI): coming of age. In *Extended Abstracts of the 2023 ACM Human-Computer Interation (CHI) Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3573832

[22] Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2020. Towards transparency by design for artificial intelligence. *Science and Engineering Ethics* 26, 6 (2020), 3333–3361.

[23] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication* 2020-1 (2020).

[24] 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.

[25] Nielson Norman Group. 2021. Nielson Norman Group. https://www.nngroup.com/

[26] Angel Gurría and OECD Secretary-General. 2011. Openness and transparency - pillars for democracy, trust and progress. https://www.oecd.org/fr/corruption/opennessandtransparency-pillarsfordemocracytrustandprogress.htm

[27] AI Standards Hub. 2022. AI Standards Database Search. https://aistandardshub.org/

[28] AI Standards Hub. 30 November 2022. AI Standards Database Search. https://aistandardshub.org/ai-standards-search/?_sft_standards_topic=human-centred-design,human-computer-interaction

[29] IBM. 2019. Design for AI. Explainability. https://www.ibm.com/design/ai/ethics/explainability

[30] IEEE. 2021. IEEE. https://www.ieee.org/

[31] IEEE. 2021. IEEE 7000™-2021. Integrating ethical and functional requirements to mitigate risk and increase innovation in systems engineering design and development. https://engagestandards.ieee.org/ieee-7000-2021-for-systems-design-ethical-concerns.html

[32] IEEE. 2021. P7001/D3 Draft Standard for Transparency of Autonomous Systems. https://www.techstreet.com/ieee/standards/ieee-p7001?gateway_code=ieee&vendor_id=6929&product_id=2188891

[33] IEEE. 2022. 7001-2021. IEEE Standard for Transparency of Autonomous Systems. https://ieeexplore.ieee.org/document/9726144

[34] Nanna Inie and Peter Dalsgaard. 2017. How interaction designers use tools to capture, manage, and collaborate on ideas. In *Proceedings of the 2017 ACM Human-COomputer Interaction (CHI) Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2668–2675. https://doi.org/10.1145/3027063.3053210

[35] The British Standards Institute. 2022. bsigroup.com. https://www.bsigroup.com/en-GB/standards/

[36] The British Standards Institute. 2022. ISO/IEC 30150-1:2022 Information technology — Affective computing user interface (AUI) — Part 1: Model. https://knowledge.bsigroup.com/products/information-technology-affective-computing-user-interface-aui-model/standard/preview

[37] British Standards Institution. 2016. Robots and Robotic Devices: Guide to the Ethical Design and Application of Robots and Robotic Systems. https://shop.bsigroup.com/products/robots-and-robotic-devices-guide-to-the-ethical-design-and-application-of-robots-and-robotic-systems/standard

[38] ISO/IEC. 2022. ISO/IEC AWI 12792 Information technology — Artificial intelligence — Transparency taxonomy of AI systems. https://aistandardshub.org/ai-standards/information-technology-artificial-intelligence-transparency-taxonomy-of-ai-systems/

[39] Milena Janic, Jan Pieter Wijbenga, and Thijs Veugen. 2013. Transparency enhancing tools (TETs): an overview. In *2013 Third Workshop on Socio-Technical Aspects in Security and Trust*. IEEE, 18–25.

[40] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (2019), 389–399.

[41] Dimitra Kamarinou, Christopher Millard, and Jatinder Singh. 2017. Machine learning with personal data. In *Data protection and privacy: the age of intelligent machines*. Hart Publishing.

[42] 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 ACM Human-computer Interaction (CHI) conference on human factors in computing systems (CHI'2020)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[43] René F Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 ACM Human-Computer Interaction (CHI) conference on human factors in computing systems (CHI'2016)*. Association for Computing Machinery, New York, NY, USA, 2390–2395.

[44] Ansgar Koene, Chris Clifton, Yohko Hatada, Helena Webb, and Rashida Richardson. 2019. A governance framework for algorithmic accountability and transparency. (2019). https://doi.org/10.2861/59990

[45] Joshua Alexander Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. Accountable algorithms. *University of Pennsylvania Law Review* 165 (2017), 633–705. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/

[46] Stefan Larsson and Fredrik Heintz. 2020. Transparency in artificial intelligence. *Internet Policy Review* 9, 2 (2020).

[47] David Lehr and Paul Ohm. 2017. Playing with the data: what legal scholars should learn about machine learning. *UC Davis Law Review* 51 (2017).

[48] 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 ACM Human-Computer Interaction (CHI) Conference on Human Factors in Computing Systems (CHI'20)*. Association for Computing Machinery, New York, NY, USA, 1–15.

[49] Michael A Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In *Proceedings of the 2020 ACM Human-Computer Interaction (CHI) conference on human factors in computing systems (CHI'20)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[50] 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 2019 ACM conference on fairness, accountability, and transparency (FAccT'19)*. Association for Computing Machinery, New York, NY, USA, 220–229.

[51] Henrik Mucha, Dennis Mevißen, Sebastian Robert, Ricarda Jacobi, Kirsten Meyer, Winfried Heusler, and Daniel Arztmann. 2020. Co-Design futures for AI and space: a workbook sprint. In *Extended Abstracts of the 2020 ACM Human-Computer Interation (CHI) Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3375203

[52] Chris Norval, Jennifer Cobbe, and Jatinder Singh. 2021. Towards an accountable Internet of Things: a call for 'reviewability'. In *Privacy by Design for the Internet of Things: Building Accountability and Security*. The Institution of Engineering and Technology.

[53] Chris Norval, Kristin Cornelius, Jennifer Cobbe, and Jatinder Singh. 2022. Disclosure by design: designing information disclosures to support meaningful transparency and accountability. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 679–690. https://doi.org/10.1145/3531146.3533133

[54] Chris Norval, Heleen Janssen, Jennifer Cobbe, and Jatinder Singh. 2021. Data protection and tech startups: the need for attention, support, and scrutiny. *Policy & Internet* (2021).

[55] Chris Norval and Jatinder Singh. 2023. A Room With an Overview: Towards Meaningful Transparency for the Consumer Internet of Things. *IEEE Internet of Things Journal* (2023), 1–1. https://doi.org/10.1109/JIOT.2023.3318369

[56] State Internet Information Office of the People's Republic of China. 2020. S.4066 - A bill to require transparency, accountability, and protections for consumers online. http://www.cac.gov.cn/2021-08/27/c_1631652502874117.htm

[57] Information Commissioner's Office and The Alan Turing Institute. 2020. Explaining decisions made with AI. https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai

[58] Cathy O'neil. 2016. *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown.

[59] Otter.ai. 2021. Otter.ai. https://otter.ai/

[60] Google Pair. 2019. People + AI Guidebook. https://pair.withgoogle.com/guidebook/chapters

[61] 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 ACM Human-Computer Interaction (CHI) conference on human factors in computing systems (CHI'21)*. Association for Computing Machinery, New York, NY, USA, 1–52.

[62] Qualtrics. 2021. Qualtrics. https://www.qualtrics.com/fr/?rid=langMatch&prevsite=en&newsite=fr&geo=FR&geomatch=

[63] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where responsible AI meets reality: practitioner perspectives on enablers

for shifting organizational practices. *Proceedings of the 2021 ACM conference on Human-Computer Interaction* 5, CSCW'21, 1–23.
[64] Gonzalo Ramos, Jina Suh, Soroush Ghorashi, Christopher Meek, Richard Banks, Saleema Amershi, Rebecca Fiebrink, Alison Smith-Renner, and Gagan Bansal. 2019. Emerging perspectives in human-centered machine learning. In *Extended Abstracts of the 2019 ACM Human-Computer Interaction (CHI) Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299014
[65] Mary Beth Rosson, Susanne Maass, and Wendy A. Kellogg. 1986. Designing for designers: an analysis of design practice in the real world. *SIGCHI Bulletin* 17, SI (May 1986), 137–142. https://doi.org/10.1145/30851.30873
[66] Daniel Schiff, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. 2020. Principles to practices for responsible AI: closing the gap. *arXiv preprint arXiv:2006.04707* (2020).
[67] Kacper Sokol and Peter Flach. 2020. Explainability fact sheets: a framework for systematic assessment of explainable approaches. In *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAccT'20)*. Association for Computing Machinery, New York, NY, USA, 56–67.
[68] Marc Steen, Menno Manschot, and Nicole De Koning. 2011. Benefits of co-design in service design projects. *International Journal of Design* 5, 2 (2011).
[69] Cynthia Stohl, Michael Stohl, and Paul M Leonardi. 2016. Managing opacity: information visibility and the paradox of transparency in the digital age. *International Journal of Communication* 10 (2016). https://ijoc.org/index.php/ijoc/article/view/4466/1530
[70] the International Electrotechnical Commission. 2022. iec.ch. https://www.iec.ch/homepage
[71] the International Organization for Standardization. 2022. iso.org. https://www.iso.org/home.html
[72] European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* L119 (4 May 2016), 1–88.
[73] Andrew Walter. 2011. *Governing finance: East Asia's adoption of international standards*. Cornell University Press.
[74] Qiaosi Wang, Michael Madaio, Shaun Kane, Shivani Kapania, Michael Terry, and Lauren Wilcox. 2023. Designing responsible AI: adaptations of UX practice to meet responsible AI challenges. In *Proceedings of the 2023 ACM Human-Computer Interaction (CHI) Conference on Human Factors in Computing Systems (CHI'23)*. Association for Computing Machinery, New York, NY, USA, 1–16.
[75] Adrian Weller. 2019. Transparency: motivations and challenges. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer, 23–40.
[76] Maranke Wieringa. 2020. What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3351095.3372833
[77] Rebecca Williams, Richard Cloete, Jennifer Cobbe, Caitlin Cottrill, Peter Edwards, Milan Markovic, Iman Naja, Frances Ryan, Jatinder Singh, and Wei Pang. 2022. From transparency to accountability of intelligent systems: moving beyond aspirations. *Data & Policy* 4 (2022), e7. https://doi.org/10.1017/dap.2021.37
[78] Alan FT Winfield, Serena Booth, Louise A Dennis, Takashi Egawa, Helen Hastie, Naomi Jacobs, Roderick I Muttram, Joanna I Olszewska, Fahimeh Rajabiyazdi, Andreas Theodorou, et al. 2021. IEEE P7001: a proposed standard on transparency. *Frontiers in Robotics and AI* (2021), 225.
[79] Ning Xie, Gabrielle Ras, Marcel van Gerven, and Derek Doran. 2020. Explainable deep learning: A field guide for the uninitiated. *arXiv preprint arXiv:2004.14545* (2020).
[80] Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating Hhw practitioners use human-AI guidelines: a case study on the People+ AI guidebook. In *Proceedings of the 2023 ACM Human-Computer Interaction (CHI) Conference on Human Factors in Computing Systems (CHI'23)*. Association for Computing Machinery, New York, NY, USA, 1–13.
[81] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling fairness perceptions in algorithmic decision-making: The effects of explanations, human oversight, and contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 134, 21 pages. https://doi.org/10.1145/3544548.3581161
[82] Tal Z Zarsky. 2013. Transparent predictions. *U. Ill. L. Rev.* (2013), 1503.
[83] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAccT'20)*. Association for Computing Machinery, New York, NY, USA, 295–305.
[84] ZOOM.us. 2021. ZOOM. https://zoom.us/
[85] Douglas Zytko, Pamela J. Wisniewski, Shion Guha, Eric P. S. Baumer, and Min Kyung Lee. 2022. Participatory design of AI systems: opportunities and challenges across diverse users, relationships, and application domains. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 4 pages. https://doi.org/10.1145/3491101.3516506

A APPENDIX

A.1 Glossary

In this paper, we use the following definitions from IEEE 7001 [32]:

- **Autonomous system**: a system that has the capacity to make decisions itself in response to some input data or stimulus with a varying degree of human oversight or intervention depending on the system's level of autonomy.
- **Designers**: designers, developers, builders, maintainers, and operators, as well as decision-makers and procurers in organizations using and deploying autonomous systems (collectively, "designers") of autonomous systems who either wish to or are required to engineer systems that have a certain degree of transparency.
- **Non-expert users (users)**: persons who have only a brief interaction or who interact every day with an autonomous system.
- **Transparency**: a transfer of information from an autonomous system or its designers to a stakeholder that is truthful; contains information relevant to the causes of some action, decision, or behavior; and is presented at a level of abstraction and in a form meaningful to the stakeholder. Transparency should be mindful of the stakeholders' likely perception and comprehension, and should avoid disclosing information in a manner that, while technically true, is framed in a way that leads to misapprehension.
- **(IEEE 7001) Transparency level for users (TL)**: "a requirement, expressed as a qualitative property of the system which must be met" in terms of transparency for users
- **TL0**: "barely or no transparency"
- **TL1**: "accessible information including scenarios and general principles of operation"
- **TL2**: "interactive training material"
- **TL3**: "a functionality to get a brief and immediate explanation of the system's most recent activity"
- **TL4**: "a functionality to get a brief and immediate explanation of the system's activity in a given situation"
- **TL5**: "a continuous explanation of the behaviour, which adapts to the user's needs and context"

A.2 Scenarios

In Study 2 (interviews), participants were presented on slides one of the following scenarios, extracted from IEEE 7001 [32]. Note that we did not change the wording, including the American spelling, except for the word "characteristics," which refers here to the transparency levels TL1, 2 and 3, as defined in IEEE 7001, so as not to bias our study in case participants were already familiar with IEEE 7001. Due to lack of time, we presented each participant with only one TL and one scenario in total.

A.2.1 Moderation: Content Moderation for AI..

- We use this fictional scenario to explore how to design 3 characteristics to provide system transparency for non-expert end-users.
- A video hosting website has been accused by activists of using keywords to prevent monetization of potentially controversial content. To mitigate a potential scandal, the website decides to communicate more transparently the decision-making processes to content creators.
- The content creators are the non-expert end-users here. They require a medium level of understanding of how the system functions, including the ability to ask the system to explain its decisions, or to pre-emptively interrogate if something is likely to be deemed problematic.

A.2.2 *Scenario Credit: Credit Scoring System.*

- We use this fictional scenario to explore how to design 3 characteristics to provide system transparency for non-expert end-users.
- A loans company asks their credit scoring provider to apply 3 characteristics to their technology, to more transparently communicate the decision-making processes to loan applicants.
- The loan applicants are the non-expert end-users here. Transparency is very important to them as the assessment is of their own particulars, and they deserve a chance to understand why they have been assessed in a particular way, and to seek to redress in the event that information is incorrect or is assessed unfairly.

A.3 Summary of Demographics from our survey (Study 1)

Figure 3 summarises the demographic information provided by our n=40 surveyed designers (Study 1). The responses to the categories 'self-described profession,' 'experience,' 'sector,' and 'area' are independent from each other.

Self-described profession	Designer	Developer	Engineer/Technical Manager	Academic	Product/ Business Manager	Entrepreneur
Total respondents n=39	7 (18%)	8 (20.5%)	6 (15%)	5 (13%)	8 (20.5%)	5 (13%)
Experience	1 year	1-3 years	4-6 years	7-9 years	10+ years	
Total respondents n=39	1 (2.5%)	11 (28.2%)	11 (28.2%)	5 (13%)	11 (28.2%)	
Sector	Education / HR	Public/ non-profit	Banking/ Finance	Healthcare	Consulting	Technology
Total respondents n=40	7 (17.5%)	3 (7.5%)	5 (12.5%)	1 (2.5%)	4 (10%)	20 (50%)
Area	Europe	UK	Asia	Oceania	North America	Middle East
Total respondents n=40	26 (65%)	8 (20%)	2 (5%)	1 (2.5%)	2 (5%)	1 (2.5%)

Figure 3: A Summary of the demographics information of surveyed designers (Study 1).