Computers in Human Behavior Reports 22 (2026) 101025

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Development and validation of the TechTra Scales: A tool to assess perceptions of and needs for technology transparency in UX research

Ilka Hein [a,*], Daniel Ullrich [b], Sarah Diefenbach [a]

[a] Department of Psychology, LMU Munich, Leopoldstraße 13, 80802, Munich, Germany
[b] Department of Computer Science, LMU Munich, Frauenlobstraße 7a, 80337, Munich, Germany

ARTICLE INFO

Keywords:
Technology transparency
Explainability
Scale development
Human-computer interaction
User experience research

ABSTRACT

The users' subjective experience of a technology's transparency is a crucial factor in human-computer interaction, shaping trust, satisfaction, and use. Technology transparency refers to different types of information, such as what the technology currently does, how it gets its results, and why. Depending on the use context, more or less transparency can be helpful and desired from a user experience perspective, posing a design challenge and making it important to assess and contrast both users' transparency perceptions and needs. In three studies (N_1 = 191, N_2 = 23, N_3 = 73), covering different technologies, user groups, and methodological approaches, we developed the *TechTra Scales* — including *TechTra-N* for measuring users' transparency need and *TechTra-P* for users' transparency perception. Results support the relevance of differentiating between needs and perceptions, the scales' applicability and sensitivity to different technology interactions, and their validity. Moreover, we showed favorable scale properties and a high use intention of practitioners, underlining TechTra's significance for research and industry.

1. Introduction

Within human-computer interaction (HCI) and user experience (UX) research, transparency becomes an ever more central quality of technology (Deters et al., 2025; Kohl et al., 2019). Along with the rise of artificial intelligence (AI) in private applications such as smart home systems and search engines, it is increasingly important for users to understand what the technology is doing and how it arrives at particular outputs to use the technology in a helpful and responsible way. For example, if a smart oven automatically shuts off while cooking, users should understand whether it did so due to a detected fire risk, a sensor error, or a misreading of routine usage. Technology transparency refers to users having the information they want about the technology (DIN Deutsches Institut für Normung e.V., 2023), especially to what it currently does, how it reaches its outcomes, and why (Hein & Diefenbach, 2025; Hein et al., 2023; Miller, 2019). As a working definition, we may understand actual transparency as an adequate mental model of a technology's state and functioning. From a UX and design perspective, transparency is a particularly interesting and challenging construct, since a simple "the more the better" approach is not applicable (Felzmann et al., 2020). On the one hand, more transparency can help users to use the technology effectively and efficiently, resolve errors, and draw appropriate conclusions (e.g., Endsley, 2017; Hein & Diefenbach, 2025; Vasconcelos et al., 2024). Furthermore, it can increase their sense of control, trust, and positive affect, leading to a higher willingness to use the technology (e.g., Calvaresi et al., 2025; Papenmeier et al., 2022; Procter et al., 2023; Shin et al., 2020). On the other hand, too much transparency can be overwhelming for users, exceeding their cognitive capacity and deteriorating their experience (Bitzer et al., 2023; Diefenbach et al., 2022). This may also lead to an inappropriate use or reduced willingness to use the technology (Fox & Rey, 2024; Hein & Diefenbach, 2025; Hu et al., 2024; Springer & Whittaker, 2020). Correspondingly, the ongoing trend of minimalistic design (i.e., simple interfaces with as few components as possible (Gumber, 2023)) rather supports less transparency. To complicate things further, designing for transparency based on objective standards may not necessarily create transparency from a subjective user perspective, and vice versa. Users may not feel a lack of transparency if they actually do not understand what a technology is doing. Even if more transparency would be "good" for users (to support more adequate, informed technology use), they may not indicate this as a requirement in user tests. On the contrary, many attempts to increase transparency (e.g., through manuals,

* Corresponding author.
E-mail addresses: ilka.hein@psy.lmu.de (I. Hein), daniel.ullrich@ifi.lmu.de (D. Ullrich), sarah.diefenbach@psy.lmu.de (S. Diefenbach).

https://doi.org/10.1016/j.chbr.2026.101025
Received 6 November 2025; Received in revised form 10 March 2026; Accepted 18 March 2026
Available online 19 March 2026

tutorials) may be perceived as too complicated and not necessarily appreciated by users when evaluating a technology. Thus, transparency forms a moving target in UX design, for which simple design recommendations cannot be made.

It is therefore even more important for UX research to have tools to study users' transparency needs and perceptions in various situations, and, in a second step, use such insights as a basis for design recommendations. The present paper addresses this need and introduces the *TechTra Scales* to evaluate technology transparency from a UX perspective, separately assessing users' needs for and perceptions of transparency. Importantly, the scales are not intended to determine whether a technology is actually transparent (i.e., whether it conveys an accurate mental model of the technology) nor to serve as an evaluation of a technology in the sense that a high score indicates a superior product. For this purpose, additional research into the users' actual understanding of the system is essential (e.g., knowledge-related questions). However, measuring the subjective experience of users is crucial for UX research. For instance, it shows how different approaches to creating transparency lead to different perceptions, how much transparency users need depending on contextual factors, and how actual transparency and the experience and need for transparency can differ, which contributes to understanding perceptions of technology users and helps to create well-adapted transparency designs.

By now, since transparency is viewed as a more human-centered, subjective construct and less as an objective property of a technology (e.g., DIN Deutsches Institut für Normung e.V., 2023; Zednik, 2021), approaches for measuring transparency perceptions and needs have been developed. However, such measures are still lacking (Hein et al., 2023; Naveed et al., 2024) and have shortcomings. Especially, they (1) have limited applicability for different products, (2) emphasize explanations as means of creating technology transparency, (3) focus on specific facets of technology transparency, (4) have a lengthy and unwieldy structure, or (5) do not allow for target-performance comparison of technology transparency (see section 2.3 for a more detailed discussion).

Therefore, the present research contributes to the current literature by examining the meaningfulness of a distinction between users' transparency needs and perceptions and by creating a widely applicable and comprehensive UX tool for assessing users' transparency perceptions and needs that allows for target-performance comparisons. More precisely, the TechTra Scales serve the following purposes:

- Providing insights into users' transparency perceptions and needs for diverse technological products and means to create technology transparency from a UX perspective
- Obtaining initial indications for the reasons of users' assessments and possible design approaches by integrating a broad range of transparency facets
- Forming a low-threshold user survey with fewer items that can be easily used by researchers and practitioners
- Enabling target-performance comparisons during product development and for existing technological products in research and industry

To reach these goals, we conducted three studies. After specifying important transparency facets and generating items, we selected the items and gained initial insights into the scales' validity through the first study. In the second study, we evaluated the scales with UX and usability professionals, while again evaluating the properties of both scales. Lastly, the third study tested the scales in a real technology interaction with a varied technology design and provided a closer look at the construct validity. The paper first describes important related work, focusing on the significance of technology transparency, its understanding in existing research, and current approaches to measuring transparency as a UX variable. We then present the three consecutive studies on the development and validation of the TechTra Scales. Finally, we discuss all results and the TechTra Scales in terms of the relevance of target-performance comparisons in transparency design, the scales' validity, their theoretical contributions, their practical application, and future research directions.

2. Related work

2.1. Relevance of technology transparency

A balanced transparency design of technologies has emerged as a crucial aim in research and industry because of negative effects of maladjusted transparency and legal requirements. Transparency was shown to influence both the experience and behavior of users (Bitzer et al., 2023; Bujold et al., 2022; Hein & Diefenbach, 2025; Shin et al., 2020). Thus, it not only impacts the users' affect and satisfaction, but can also have far-reaching and safety-critical consequences through a change in users' behavior. For example, users who have a better mental model of the technology and understand how it works are more likely to recognize errors, correctly assess the accuracy of outputs, and act accordingly (Alonso & de la Puente, 2018; Gunning et al., 2021; Hein & Diefenbach, 2025). This can reduce the likelihood of technology-related accidents (e.g., with machines in the workplace) and poor technology-based decisions (e.g., through overreliance on economic or medical recommendations (Calvaresi et al., 2025)). Moreover, manufacturers may also face consequences if they do not appropriately consider transparency when designing their products. On the one hand, an impaired UX and therefore a reduced willingness to use the product may result in a loss of sales (Hein & Diefenbach, 2025; Shin et al., 2020; Vorm & Combs, 2022). On the other hand, manufacturers have to fear legal action if they do not comply with regulatory requirements. This risk is increasing as political bodies are passing more and more laws that stipulate transparency (e.g., European AI Act, Digital Services Act, General Data Protection Regulation (European Commission, 2025; Goodman & Flaxman, 2017; Kaushal et al., 2024)). At the same time, it has to be recognized that different stakeholders may have different perspectives and desires in terms of transparency. For instance, companies may also prioritize profit over user satisfaction and autonomy, and thus, focus neither on improving users' subjective perception of transparency nor on creating a transparent technology design that empowers users in their technology use. Yet, in times of a highly competitive market, users' experiences are, in addition to legal requirements, ever more important for profitability (e.g., Gaborov and Ivetić, 2022) and should therefore not be neglected by stakeholders.

Given the increasing relevance of transparency, it is important to advance research in this area as findings can, for instance, inform political decisions and help companies to design user-centered technologies. However, a good methodological basis is crucial, enabling the transparency perceptions and needs of users to be measured reliably and validly. The present paper contributes to such sound methodology and to facilitate the investigation of when a specific transparency level and design is appropriate from a UX point of view. To this end, we first reflect on current research on technology transparency in the next sections before we present our advanced tool.

2.2. Understanding and definitions of technology transparency in current research

In current research, researchers referring to the transparency of a technology use diverse terms such as interpretability, explainability, intelligibility, understandability, and transparency, partly applying them as synonyms. This inconsistent terminology and varying definitions result in a confusing research landscape for scientists and practitioners (Brennen, 2020; Hein & Diefenbach, 2025). For instance, Barredo Arrieta et al. (2020) and Mohseni et al. (2021) defined understandability and intelligibility as having the same meaning. Likewise, transparency is deemed as being intertwined with understandability

(Barredo Arrieta et al., 2020), and transparency and explainability are seen as synonymous by some researchers and as distinct by others (Brennen, 2020; Mohseni et al., 2021). For AI systems, additional adjectives such as accountable, inspectable, reliable, or responsible emerged and further complicated the research area (Brennen, 2020; Capel & Brereton, 2023). The different understandings also show the partly objective views that are not focused on user perceptions, and the partly subjective, user-centered views on this research topic. For example, Mohseni et al. (2021) described transparent AI objectively as "an AI-based system that provides information about its decision-making process", while Shin and Park (2019) stated that "[t]ransparency […] cannot be measured as feature[.] of algorithms […] because they involve subjective dimensions that can be experienced and perceived by users.". Correspondingly, Shin (2021) highlighted "ability" in the term explainability, referring to the central role of users in making sense of systems' explanations.

For the present research and development of the TechTra Scales, we used technology transparency as an umbrella term for other existing terms (e.g., explainability, interpretability, understandability) because of several reasons. First, we thereby aligned with other researchers who organized the terminology and defined transparency as an overarching concept (Clinciu & Hastie, 2019; Mohseni et al., 2021; Shin & Park, 2019; Turri et al., 2024), not further fragmenting the research landscape. Second, transparency is a comprehensive construct (Eyert & Lopez, 2023; Turri et al., 2024), neither focusing on a specific outcome of transparent design (e.g., responsibility as suggested in "responsible AI") nor referring to single approaches for creating transparency (e.g., explanations as suggested in "explainability"). Correspondingly, explainability is often described as a means to achieve transparency (Deters et al., 2023; Naveed et al., 2024). Third, transparency can be considered with all technologies and is not so closely connected to AI products as, for instance, explainability – since not only AI but also tangible, low-tech systems can be perceived as more or less transparent and create a need for transparent design (Colin & Martin, 2023; Hein & Diefenbach, 2025). Fourth, transparency as a term should be easier to understand for users compared to, for example, the less tangible concept of interpretability – especially when illustrated through different facets of transparency (e.g., how a technology works, what it does). Yet, the understandability was also tested in study 1 to ensure the suitability of the term for the TechTra Scales. Thus, based on previous works (DIN Deutsches Institut für Normung e.V., 2023; Hein et al., 2023; Naveed et al., 2024), in this paper, we broadly define technology transparency as users having their desired information about the technology, especially regarding what it currently does, how it reaches its outcomes, and why. This underlines the subjective nature of the construct by emphasizing the importance of user perceptions (Suh et al., 2025). From a user perspective, an ideal level of transparency is reached when the user gets all the "desired" information (rather than all the objectively required or possibly available information, which would be overwhelming), and users' transparency need emerges as a motivational state when this desire is not fulfilled. This broad UX-based definition was further developed by highlighting key components of transparency (see sections 4 and 5.5). Crucially, we and the developed scales therefore do not focus on the users' actual understanding of a technology and accuracy of their mental model but their perceived transparency as an independent, important variable that shapes user reactions to technological systems. Before describing our studies in detail, we first review the current technology transparency measures that provided a basis for our work.

2.3. Existing measures of technology transparency

In line with the lacking consistency of terms and definitions, different measures emerged in the field of technology transparency with various names and contents. According to the objective and subjective views on transparency, *objective* and *human-centered evaluations* (Vilone & Longo, 2021) can be distinguished (see Naveed et al. (2024) and Lopes et al. (2022) for a similar differentiation). While the former include formal, mathematical metrics (e.g., percentage of correct rules), the latter comprise qualitative and quantitative studies with humans. Yet, evaluations and especially human-centered ones are increasingly but not always included in current research (Hein et al., 2023; Naveed et al., 2024; Suh et al., 2025), risking users' unfavorable reactions and far-reaching negative effects of maladapted transparency designs (see section 2.1). Focusing on human-centered measures, we present in the following an overview of currently used self-report scales while noting their shortcomings. Since our goal was to create low-threshold scales that can be analyzed with less effort and in a standardized manner, we do not include qualitative measures (e.g., interviews, think-aloud protocols (Turri et al., 2024)). Thus, the following paragraphs outline found categories of quantitative scales with pertinent examples.

Expert Assessment Scales as Proxies for Users' Transparency Evaluations. Expert surveys offer valuable ways to systematically assess transparency experiences based on specialist opinions. As an example, the *Explanation Goodness Checklist* (Hoffman et al., 2018) allows researchers or domain experts to evaluate features that render explanations good (e.g., understandability, completeness, trustworthiness). Likewise, Deters et al. (2025) developed heuristics that can be used by software engineers to rate a system's explanation. While such expert surveys are useful for an initial investigation, the users' perspective remains critical for product design and can deviate from expert assessments (Hein et al., 2023; Werz et al., 2024). Furthermore, the checklist only allows yes or no ratings, which do not enable nuanced comparisons between different transparency designs and may be difficult to answer, as transparency rather involves a gradual perception (Felzmann et al., 2020; Vorm & Combs, 2022).

Scales for Users' Satisfaction with Explanations. To better capture users' perspectives, Hoffman et al. (2018) developed the *Explanation Satisfaction Scale*, building on the Explanation Goodness Checklist with a more nuanced 5-point Likert scale. However, differences in item numbers and focus areas (e.g., usefulness to goals is not integrated in the checklist) impair the scales' direct comparability. The *System Causability Scale* (Holzinger et al., 2020) can also be applied to examine the perceived explanation quality, concentrating on the users' causal understanding. In addition, the *xAI for Human-Agent Interaction Survey* (Silva et al., 2022) with 30 items is available, capturing the transparency, usability, and simulatability of agent explanations and suggestions. The authors also proposed a 14-item version, which still needed to be validated. Overall, the scales focus on explanations as an important means of transparency but are therefore limited to this way of creating transparency and to developed products with fixed explanations. Moreover, only single facets of transparency are captured (e.g., accuracy, trustworthiness), while not all facets may be desired by users. For example, users do not always seek complete explanations (Miller, 2019; Mueller et al., 2019) as the Explanation Satisfaction Scale asks for. Thus, assessing the users' needs is essential to tailor transparency designs to users. Lastly, the scales partly target selected technologies and use contexts (e.g., agent assistance in the xAI for Human-Agent Interaction Survey).

Compiled Transparency Scales for Specific Studies. Several studies, such as those by Cramer et al. (2008), Wanner et al. (2022), Schott et al. (2024), and Shin (2021) have compiled items to assess transparency in specific contexts like art recommendations and online shopping assistants by mixing and modifying existing measures as well as adding self-generated items. Commonly, authors also created a single item to capture transparency (e.g., Buçinca et al., 2020; Cai et al., 2019; Ooge et al., 2022). This approach is efficient but simultaneously has weaknesses. On the one hand, the measures are designed for the specific study setting and technology and, therefore, not universally applicable. On the other hand, they are not thoroughly evaluated, which may question the validity of findings. Accordingly, Naveed et al. (2024) described such ad hoc questionnaires as a common pitfall that should be replaced by standardized questionnaires.

Transparency-Related Subscales of Measures for Related Constructs.

Researchers also used items of broader measures, such as the perceived understandability factor of the *Human-Computer Trust Scale* (Madsen & Gregor, 2000) and the transparency factor of the *Multidimensional Trust Questionnaire* (Roesler et al., 2022). Although such subscales are properly validated and can offer valuable insights into transparency, they were meant to contribute to another overarching construct and therefore likely focus on specific facets of transparency (e.g., how a system works, the predictability of its actions), neglecting other important facets such as what (personal) data is used. Also, some scales are only usable for single technologies, such as the Human-Computer Trust Scale for decision support systems.

Scales for Associated Constructs. Transparency is frequently mixed with constructs such as trust, mental model accuracy, or usefulness (Lopes et al., 2022; Naveed et al., 2024; Nunes & Jannach, 2017). This resulted in long questionnaires with numerous constructs, such as the one by Deters et al. (2025), which contains 69 items and, for example, puts trustability and transparency on one level. Yet, trustability is rather a target outcome of transparency (Speith & Langer, 2023) and should be considered downstream of transparency. While being comprehensive, such measures may blur important distinctions between constructs. Also, they can reduce the comparability of results due to their different composition and result in lengthy, not user-friendly inquiries.

Scales for Users' Transparency Needs. In contrast to perceived transparency, only a few measures focus on user needs. For instance, Schoonderwoerd et al. (2021) integrated three transparency need-related items without considering transparency need as an independent construct, and in Ooge et al. (2022), participants just marked reasons with a cross, why they wanted an explanation. The *Demand for Explainability Scale* (Weber et al., 2021) stands out as a validated questionnaire comprising seven items rated by users. Yet, it does not allow for a direct target-performance comparison as no compatible performance (i.e., transparency perception) scale exists. Also, the scale focuses on several facets of transparency (especially how a system works) and on explanations as one means of transparency. It additionally mixes items referring to a concrete technology design (e.g., "The system needs to be explained more."), more general items (e.g., "Understanding how the system works is important to me."), and not self-related items (e.g., "It is important that people understand how this system works."). Thereby, it may dilute the measurement of the users' needs. For example, users might have a low transparency need (e.g., due to high technical knowledge) but could strongly agree with the statement that people should understand the systems' functioning.

In conclusion, various scales for measuring technology transparency are currently available, but they have several shortcomings. More precisely, they …

- … often *focus on specific technologies* (e.g., recommender systems, AI agents) and are not universally applicable across technologies. A universal tool would be easier to use without the need to search for a context-specific one and could serve as a clear guideline for practitioners, encouraging user inquiries in design processes. Moreover, a broad applicability would enable comparisons of user evaluations across studies and technologies and should reduce the practice of scale adjustments that can compromise the validity and reliability of findings.
- … *concentrate primarily on evaluating explanations* provided by the technology. On the one hand, such measures typically require a developed product with concrete explanations, neglecting that transparency experiences should also be assessed early in the development process to prevent costly design errors. On the other hand, these tools lack applicability to more subtle transparency designs that increase understanding but are not clearly perceived as explanations or disclosures by users (e.g., icons, auditory or visual cues (Hein et al., 2024; van Berlo and Breves, 2025)). This creates the need for a tool that can be applied for various means to create transparency.
- … mostly assess a *limited set of transparency facets* (e.g., how a technology works), neglecting others such as its performance or use of personal data. A more comprehensive approach is needed to capture a broader range of transparency facets, making it possible to identify those that are most important to users for a specific product.
- … are partly *lengthy and impractical*. A more concise, user-friendly tool would encourage adoption in both academic and industry settings.
- … do *not compare desired and actual transparency designs*, making it difficult to identify gaps between the two. Since transparency needs vary across contexts, with too much transparency sometimes being counterproductive from a UX perspective, a refined approach should ensure that transparency designs align with user needs.

Nevertheless, the measures are fruitful approaches for creating an improved UX tool, as they provide inspiration for relevant transparency facets and item formulations that were already evaluated. Thus, our review formed the basis for the studies, which we explain in the next sections.

3. Study approach

We performed three consecutive studies, focusing on the item selection (study 1), adjustments of the scales (study 2), and testing of the scales in a real technology interaction (study 3). All studies were approved by the university's ethics committee (vote numbers: EK-MIS-2024-280, EK-MIS-2024-304, EK-MIS-2025-0335) and collected written informed consent of the participants to the data privacy. Moreover, we preregistered the studies and uploaded their materials, data sets, and analysis scripts to the OSF repository (see https://osf.io/w8g3k/overview?view_only=b0f496f48cf44818990bb3f58fad147d).
Regarding tools, we implemented all questionnaires via Unipark and performed all quantitative analyses with R and RStudio. Participant recruitment was conducted through various channels, such as social media sites, direct requests, university platforms, snowball sampling, and HCI conferences. As a prerequisite for all three studies, participants had to be of legal age and proficient in English and, for study 3, as a laboratory study including interaction with a robot via voice commands, also in German.

4. Initial item generation

As a foundation for the first study, we generated a set of items to measure users' transparency perception and need. Based on a comprehensive literature review on definitions and existing questionnaires and an integration of previous research findings (e.g., Corbett & Dove, 2024; Hein & Diefenbach, 2025; Vorm & Miller, 2020), we identified core facets of transparency as listed in Table 1.

Subsequently, we generated two items per facet (see Table 2) based on the facet's meanings, existing questionnaires, and rules for item formulation (e.g., no double-barreled items, no complex sentence structures (Lenzner & Menold, 2016; Moosbrugger & Kelava, 2020)). This allowed us to examine different item formulations and sub-facets. We did not develop reverse-coded items because they would make the scales less easy to use and may be an additional source of variance (e.g., Suárez-Álvarez et al., 2018; Weijters & Baumgartner, 2012). To make valid target-performance comparisons possible and the scales easier to apply, we used the same facets and item wordings for both the transparency perception and need. Response options ranged from 1 (*strongly disagree*) to 7 (*strongly agree*) to get differentiated ratings and offer a neutral option if users do not have a strong opinion on a facet.

In addition to the items for each facet, we included three general items to assess transparency more broadly (e.g., general understandability) and to see if they yield comparable results and correlate with the facets. Thereby, we explored whether broader items are necessary to include, as focusing on the facets provides more precise information on

Table 1
The transparency facets with exemplary references.

Facet	Exemplary References
How	"How" and "Conceptual Tree Model Visualization" as intelligibility types (Lim & Dey, 2009) "How does it work?" as trigger for explanations (Mueller et al., 2019) and explanation need (Hoffman, Mueller, et al., 2023) "How" as explanation type (Mohseni et al., 2021), transparency type (Hein & Diefenbach, 2025), category of explainability needs (Liao et al., 2020), and explanatory question (Miller, 2019)
Why	"Why" and "Why not" as intelligibility types (Lim & Dey, 2009), explanation types (Mohseni et al., 2021), categories of explainability needs (Liao et al., 2020), and intelligibility queries (Wang et al., 2019) "Why didn't it do z?" as trigger for explanations (Mueller et al., 2019) and explanation need (Hoffman, Mueller, et al., 2023) "Why" as explanatory question (Miller, 2019) and transparency type (Hein & Diefenbach, 2025)
How to use	"How do I use it?" as trigger for explanations (Mueller et al., 2019) and explanation need (Hoffman, Mueller, et al., 2023) "Control" as intelligibility type (Lim & Dey, 2009) "How to" as intelligibility query (Wang et al., 2019) "How to use" as transparency type (Hein & Diefenbach, 2025)
Performance	"Certainty" as intelligibility type (Lim & Dey, 2009) and intelligibility query (Wang et al., 2019) "What can't it do? What are its limitations?" as trigger for explanation need (Hoffman, Miller, et al., 2023) "Performance" as category of explainability needs (Liao et al., 2020) and transparency type (Hein & Diefenbach, 2025) "Accuracy" as explanation attribute (Hoffman, Mueller, et al., 2023)
Data	"Input[s]" as intelligibility type (Lim & Dey, 2009) and intelligibility query (Wang et al., 2019) "Data" as category of explainability needs (Liao et al., 2020) and transparency type (Hein & Diefenbach, 2025)
What	"What did it just do?" and "What will it do next?" as triggers for explanations (Mueller et al., 2019) and explanation need (Hoffman, Mueller, et al., 2023) "Output[s]" as intelligibility type (Lim & Dey, 2009), intelligibility query (Wang et al., 2019), and category of explainability needs (Liao et al., 2020) "What" as explanatory question (Miller, 2019) and transparency type (Hein & Diefenbach, 2025)
Error details	"What do I do if it gets it wrong?" and "How do I avoid the failure modes?" as triggers for explanations (Mueller et al., 2019) "What do I do if it gets it wrong?", "How do I avoid the failure modes?", and "How can I repair the tool or help it work better?" as triggers for explanation need (Hoffman, Miller, et al., 2023) "Error details" as transparency type (Hein & Diefenbach, 2025)
Usefulness	"How will it help me to do a better job?", "What does it achieve?", and "How much effort will this take?" as triggers for explanation (Mueller et al., 2019) and explanation need (Hoffman, Miller, et al., 2023) "Usefulness" as explanation attribute (Hoffman, Mueller, et al., 2023)
Trustworthiness	"What do I do if I mistrust or distrust the tool?" as trigger for explanation need (Hoffman, Miller, et al., 2023) "Trustworthiness" as explanation attribute (Hoffman, Mueller, et al., 2023)

the reasons for users' transparency experience and possible design solutions. Moreover, it would increase the ease of use if item beginnings are identical across scales. Nevertheless, the broader items served as a measure of validity. All initially generated items are listed in Table 2.

5. Study 1: Item selection and initial evaluation

In the first study, we selected items based on user evaluations of a broad range of technologies (Hein & Diefenbach, 2025). More precisely, we studied transparency perceptions and needs based on users' experiences with

- their washing machine as a tangible, less complex, and more established technology,
- the social media platform YouTube as a digital and frequently used website, and
- an AI-based chatbot (e.g., ChatGPT) as a new digital product using recent technological developments.

In addition, the study's goal was to gain insights into the validity of

Table 2
Initially generated items for both the perception of and need for transparency with references.

Facet	Items for Perception of/ Need for Technology Transparency	References for Inspiration
	It seems transparent to me …/ It should be transparent to me …	
How	… how the technological product works.	(Hoffman et al., 2018; Mueller et al., 2019; Weber et al., 2021)
	… what the logic of the technological product is.	(Hein & Diefenbach, 2025; Liao et al., 2020; Schmude et al., 2025)
Why	… why the technological product behaves the way it does.	(Cramer et al., 2008; Wanner et al., 2022; Weber et al., 2021)
	… why the technological product shows a certain behavior in a specific situation.	(Hein & Diefenbach, 2025; Silva et al., 2022)
How to use	… how to use the technological product.	(Hoffman et al., 2018; Madsen & Gregor, 2000; Mueller et al., 2019)
	… how to operate the technological product to achieve what I want.	(Hein & Diefenbach, 2025; Hoffman et al., 2018; Wanner et al., 2022)
Performance	… how accurately the technological product works.	(Hein & Diefenbach, 2025; Hoffman et al., 2018; Liao et al., 2020)
	… what the limitations of the technological product are.	(Hoffman, Miller, et al., 2023; Liao et al., 2020)
Data	… what the behavior of the technological product is based on.	(Cramer et al., 2008; Deters et al., 2025; Wanner et al., 2022)
	… what personal data the technological product uses.	(Hein & Diefenbach, 2025; Liao et al., 2020)
What	… what the technological product is doing at the moment.	(Hein & Diefenbach, 2025; Hoffman et al., 2018; Mueller et al., 2019; Wanner et al., 2022)
	… what the technological product will do next.	(Hein & Diefenbach, 2025; Hoffman et al., 2018; Mueller et al., 2019; Ooge et al., 2022)
Error details	… what I have to do in case of an error.	(Hein & Diefenbach, 2025; Hoffman et al., 2018; Mueller et al., 2019)
	… what error messages or warnings of the technological product mean.	(Hein & Diefenbach, 2025; Liao et al., 2020)
Usefulness	… how the technological product assists me in reaching my goals.	(Hoffman et al., 2018; Madsen & Gregor, 2000; Wanner et al., 2022)
	… how the technological product helps me to receive what I need.	(Hoffman et al., 2018; Hoffman, Miller, et al., 2023)
Trustworthiness	… how much I can trust the technological product.	(Hoffman et al., 2018)
	… how reliable the technological product is.	(Hoffman et al., 2018; Liao et al., 2020; Wijekoon et al., 2024)
General	The technological product seems/should be easily understandable.	(Deters et al., 2025; Shin, 2021; Wijekoon et al., 2024)
	The technological product provides me/should provide me with enough information to understand it.	(Deters et al., 2025; Hoffman et al., 2018; Silva et al., 2022) (Hoffman et al., 2018; Wijekoon et al., 2024)
	The transparency of the technological product is/should be satisfying.	

both scales after item reduction. In the next sections, we present the sample characteristics, the procedure and materials used, and the data analysis strategy. Then, we report the key results and item selection procedure. Lastly, we describe findings of the validity analyses, also examining the separate assessment of users' transparency need and perception.

5.1. Participants

218 individuals completed the questionnaire. We excluded 27 participants who did not rate all items on one of the two developed scales and answered the attention check item incorrectly. This resulted in a final sample size for the analyses of $N = 191$. The participants' ages ranged from 18 to 63 years, with an average of 27.36 years ($SD = 7.40$). 61.67% of the sample identified as female and 38.33% as male. As the highest academic degree, 36.70% indicated having a university entrance qualification, 35.00% a completed vocational training, and 24.44% a university degree. Moreover, most participants were employees (65.00%, including public officials) or students (29.44%).

5.2. Procedure and materials

In an online survey, participants first rated their perception of the three technological products in random order using the developed 21 items (see Table 2), which were also presented in random order. Subsequently, they could indicate what could be improved about the items and filled out the short version of the *Affinity for Technology Interaction (ATI-S) Scale* (Wessel et al., 2019) (Cronbach's $\alpha = .81$), of the *Need for Cognition (NfC-S) Scale* (Beißert et al., 2014) (Cronbach's $\alpha = .59$), and of the *Need for Closure (NFC-S) Scale* (Roets & van Hiel, 2011) (Cronbach's $\alpha = .85$). Then, they specified their transparency need for each technological product via the *Demand for Explainability Scale* (Weber et al., 2021) (Cronbach's $\alpha = .89$) and the 21 need for transparency items (see Table 2). Again, the participants could comment on aspects they noticed about the items. Next, we asked for experience with and attitude towards the three technological products (one item each rated from 1 (*very negatively*) to 7 (*very positively*)), their understanding of the concept of technology transparency (one item rated from 1 (*strongly disagree*) to 7 (*strongly agree*)), and their general need for technology transparency (two items rated from 1 (*strongly disagree*) to 7 (*strongly agree*), Cronbach's $\alpha = .64$). Lastly, participants indicated their demographics and could choose between course credit or a raffle of four 25€ vouchers as compensation. All materials can be found in the OSF repository (see section 3).

5.3. Data analysis

First, we calculated a one-sample *t*-test to explore whether the term technology transparency was deemed understandable by users, that is, whether the understandability rating significantly deviated from 4 as a neutral midpoint of the scale. Then, we analyzed the developed items for both scales using Cronbach's alpha and McDonald's omega, indicators for the response distribution (i.e., mean, median, standard deviation, range, skewness, kurtosis), corrected item-total correlations, and principal component analyses (PCAs). For the latter, we previously checked the Kaiser-Meyer-Olkin (KMO) index for sampling adequacy and Bartlett's test for sphericity. To determine the number of components, we ran parallel analyses. However, all PCAs were computed with two components to see the magnitude of loadings on the second component. We performed all analyses aggregated for the three technologies (reported in section 5.4) and separately for each technology (available in Supplementary material A). Additionally, we summarized the qualitative comments to support the item selection.

After item selection, we correlated the final scales with associated constructs to obtain insights into the scales' validity. As constructs, we examined the Demand for Explainability Scale, the three general items of the transparency perception and need scales, the general need for transparency, ATI, need for cognition, need for closure, as well as participants' experience with and attitude towards the technology. Lastly, we compared the perception and need ratings between the three technologies by one-way repeated measures ANOVAs with the technology as the independent variable and the final scales as the dependent variables, and subsequently, pairwise *t*-tests as post-hoc analyses.

5.4. Results

Participants indicated a high understandability of the term technology transparency ($M = 6.11$, $SD = 1.00$) that was significantly above the midpoint of the scale ($t(180) = 28.41$, $p < .001$). Moreover, the developed items showed a high reliability through Cronbach's alpha ($\alpha_{\text{Perception of transparency}} = .94$, $\alpha_{\text{Need for transparency}} = .97$) and McDonald's omega ($\omega_{\text{Perception of transparency}} = .94$, $\omega_{\text{Need for transparency}} = .97$). The reliability changed little when one item was dropped (i.e., $\alpha_{\text{Need for transparency}}$ only changed to .96 for some items). Furthermore, the descriptive values for the response distributions and the corrected item-total correlations can be seen in Table 3. Skewness values affirmed symmetric response distributions and kurtosis values mesokurtic response distributions – with the second *how to use* item showing the greatest deviation from a normal distribution for perception of transparency. Also, item-total correlations were reasonably strong, supporting that the items measured the same construct. For the transparency need, both *error detail* items and *data 2* showed the lowest correlations, and for the transparency perception, additionally both *how to use* items yielded lower correlations. The *how to use* items also stood out slightly because of their high means, lower standard deviations and ranges.

Table 4 presents the results of the PCAs for both constructs. Consistent with the previous results, we found the lowest first factor loadings for transparency perception for items *how to use 2*, *data 2*, *error details 1* and *2*. For transparency need, this was the case for *data 2*, *error details 1* and *2*.

5.5. Item selection

Based on the quantitative results, qualitative statements, and theoretical considerations, we selected the items for the final scales (see Table 5). In general, we strived for one item per facet to cover the different transparency facets. Thus, we chose *how 1* because of its slightly higher factor loading and item-total correlation, and because the wording "logic of the technology" of *how 2* was seen as unclear by participants. *Why 1* and *performance 1* were preferred due to similar quantitative findings and comments on the understandability of the facet's other item ("The meaning of why the technological product shows a certain behavior is not quite clear […]", "Not every question make[s] sense for every technology (e.g., limitations of a washing machine)"). The same applied for *trustworthiness*, where we kept the first item for clarity reasons ("Unsure about how to rate the reliability"). Since the findings for *how to use* were ambiguous, the first item was selected to increase the distinctiveness to the *usefulness* facet. For the latter, we retained the first item to enhance the fit to the facet, as the results also did not clearly favor one item.

However, for the *data* and *what* facet, we decided to integrate both items because they cover different aspects of the facets, and because the *personal data* facet was especially important to users in previous studies (Hein & Diefenbach, 2025; Wang et al., 2023; Weitz et al., 2022). Contrarily, we removed the *error details* items from the final scales as they showed less favorable results both quantitatively and qualitatively (e.g., "'Error' could have different meanings for different technologies") and may be difficult to answer when no error occurred during the technology interaction. The *general* items were also not retained because they yielded ratings comparable to those of the items for the facets, but do not allow for concrete design recommendations and make the scales less easy to use when not all items have the same beginning.

Table 3
Descriptive values and corrected item-total correlation of each item across the three technologies.

Item [a]	*M*	*SD*	Median	Range	Skewness	Kurtosis	Corrected Item-Total Correlation
Perception of technology transparency							
How 1	4.79	1.14	4.67	[2.00; 7.00]	−0.27	−0.27	.72
How 2	4.92	1.17	5.00	[2.00; 7.00]	−0.33	−0.38	.72
Why 1	4.57	1.08	4.67	[1.67; 7.00]	−0.16	−0.38	.77
Why 2	4.40	1.14	4.33	[1.67; 7.00]	−0.18	−0.52	.74
How to use 1	6.00	0.73	6.00	[3.67; 7.00]	−0.67	0.02	.54
How to use 2	5.70	0.77	5.67	[2.33; 7.00]	−0.76	1.16	.50
Performance 1	4.75	1.07	5.00	[1.33; 7.00]	−0.34	−0.20	.77
Performance 2	4.58	1.11	4.67	[1.67; 7.00]	−0.24	−0.06	.61
Data 1	4.44	1.19	4.33	[1.67; 7.00]	−0.11	−0.51	.71
Data 2	3.42	1.20	3.33	[1.00; 7.00]	0.23	0.04	.51
What 1	4.63	1.17	4.67	[1.00; 7.00]	−0.05	−0.32	.62
What 2	4.59	1.11	4.67	[1.33; 7.00]	−0.14	−0.12	.60
Error details 1	3.83	1.16	4.00	[1.00; 7.00]	0.22	−0.37	.56
Error details 2	3.99	1.18	4.00	[1.00; 7.00]	−0.02	−0.48	.54
Usefulness 1	5.48	0.88	5.67	[2.00; 7.00]	−0.63	0.81	.66
Usefulness 2	5.51	0.91	5.67	[2.67; 7.00]	−0.54	0.05	.65
Trustworthiness 1	4.31	1.03	4.33	[1.33; 7.00]	0.01	−0.20	.69
Trustworthiness 2	4.66	0.94	4.67	[2.00; 7.00]	−0.17	−0.04	.68
General 1	5.49	0.91	5.67	[2.67; 7.00]	−0.50	0.06	.62
General 2	5.09	1.02	5.00	[1.33; 7.00]	−0.56	0.44	.59
General 3	4.32	1.08	4.33	[1.33; 7.00]	0.00	−0.12	.64
Need for technology transparency							
How 1	4.54	1.20	4.59	[1.00; 7.00]	−0.34	0.22	.81
How 2	4.39	1.21	4.43	[1.00; 7.00]	−0.25	−0.01	.72
Why 1	4.60	1.10	4.66	[1.67; 7.00]	−0.43	0.00	.82
Why 2	4.78	1.16	4.83	[1.67; 7.00]	−0.36	−0.20	.79
How to use 1	4.68	1.29	4.72	[1.00; 7.00]	−0.30	−0.17	.74
How to use 2	4.88	1.17	4.92	[2.00; 7.00]	−0.27	−0.38	.79
Performance 1	4.96	1.02	4.97	[2.67; 7.00]	−0.12	−0.43	.79
Performance 2	4.97	1.06	4.99	[2.33; 7.00]	−0.19	−0.33	.73
Data 1	4.60	1.14	4.61	[1.67; 7.00]	−0.11	−0.12	.78
Data 2	5.04	1.02	5.06	[1.33; 7.00]	−0.30	0.60	.60
What 1	4.37	1.27	4.38	[1.00; 7.00]	−0.16	−0.04	.77
What 2	4.46	1.20	4.51	[1.00; 7.00]	−0.42	0.23	.78
Error details 1	5.12	1.16	5.16	[2.00; 7.00]	−0.37	−0.29	.60
Error details 2	5.08	1.17	5.14	[2.00; 7.00]	−0.38	−0.28	.61
Usefulness 1	4.66	1.19	4.69	[1.33; 7.00]	−0.23	0.01	.79
Usefulness 2	4.77	1.22	4.80	[1.00; 7.00]	−0.25	−0.08	.81
Trustworthiness 1	5.03	1.12	5.09	[1.33; 7.00]	−0.47	0.14	.74
Trustworthiness 2	5.06	1.09	5.08	[2.00; 7.00]	−0.16	−0.50	.80
General 1	4.33	1.30	4.34	[1.00; 7.00]	−0.11	0.15	.72
General 2	4.69	1.16	4.69	[1.00; 7.00]	−0.06	−0.13	.82
General 3	4.73	1.09	4.72	[1.33; 7.00]	−0.05	−0.08	.77

[a] See Table 2 for wordings of all items.

5.6. Analysis of the final TechTra Scales

The final scales, including the concrete instructions and items, are available in Supplementary material B. For insights into the validity of the final scales, correlations with associated variables are available in Table 6. As assumed, the Demand for Explainability Scale, overall transparency need, and general items of the transparency need (see Table 2) correlated significantly positively with TechTra-N and general items of the transparency perception with TechTra-P. From the individual differences variables (i.e., ATI, need for cognition, need for closure), only the latter had a significant positive relation with TechTra-P. Yet, the experience with and attitude towards the technology as technology-dependent variables both correlated positively with TechTra, showing that more technology experience and a more positive attitude towards it were associated with higher perceived but, for technology experience, also with needed transparency.

Furthermore, the differences between the three technologies on the scales matched our expectations based on previous research (Bitzer et al., 2023; Hein & Diefenbach, 2025; Wang & Yin, 2021) (see Fig. 1): While the AI chatbot was perceived as the least transparent, the need for transparency was highest for this technology. In contrast, participants rated the washing machine as the most transparent and their transparency need for this technology as the highest. However, TechTra-P and TechTra-N did not correlate significantly ($r(181) = -.09$, $p = .337$), which underlines the relevance of investigating both.

6. Study 2: Practitioner evaluation

After obtaining the first version of the TechTra Scales via the first study, the goal of the second study was to gain quantitative and qualitative insights into professionals' evaluation of the scales. We therefore investigated

- how both transparency scales were evaluated by UX/usability professionals and
- what can be done to improve the scales for them.

Additionally, we reexamined the reliability and relation between the perception and need scale. The following sections provide information on the study's participants, procedure and materials, and data analysis. Subsequently, we describe the quantitative and qualitative results and the adjustments made to the scales.

Table 4
Principal component analyses for both constructs across the three technologies.

Item [a]	Perception of Technology Transparency			Need for Technology Transparency		
	Component 1 loadings	Component 2 loadings	Communality [b]	Component 1 loadings	Component 2 loadings	Communality [b]
How 1	.76	−.14	.60	.84	−.17	.73
How 2	.76	−.21	.63	.75	−.15	.59
Why 1	.80	.07	.64	.84	.02	.70
Why 2	.77	.15	.62	.82	.07	.67
How to use 1	.60	−.49	.60	.77	−.34	.71
How to use 2	.56	−.58	.65	.81	−.19	.70
Performance 1	.81	.08	.66	.82	.10	.68
Performance 2	.65	.10	.43	.76	.23	.63
Data 1	.75	.03	.57	.81	.10	.66
Data 2	.54	.55	.59	.63	.37	.53
What 1	.67	−.04	.44	.80	−.07	.64
What 2	.64	.07	.41	.80	−.10	.66
Error details 1	.58	.43	.52	.62	.55	.69
Error details 2	.56	.37	.45	.64	.57	.73
Usefulness 1	.71	−.33	.61	.82	−.18	.70
Usefulness 2	.71	−.38	.64	.83	−.19	.73
Trustworthiness 1	.72	.26	.59	.77	.26	.66
Trustworthiness 2	.72	.15	.54	.82	.23	.73
General 1	.67	−.29	.54	.75	−.35	.68
General 2	.63	−.04	.40	.84	−.24	.76
General 3	.67	.32	.56	.80	−.21	.68

[a] See Table 2 for wordings of all items.
[b] Communality describes the proportion of item variance that is explained by this two-factor solution.

Table 5
Items of the final TechTra Scales.

Facet	Items of TechTra-P/ TechTra-N
	It seems transparent to me …/ It should be transparent to me …
How	… how the technological product works.
Why	… why the technological product behaves the way it does.
How to use	… how to use the technological product.
Performance	… how accurately the technological product works.
Data (data basis)	… what the behavior of the technological product is based on.
Data (personal data)	… what personal data the technological product uses.
What (observability)	… what the technological product is doing at the moment.
What (predictability)	… what the technological product will do next.
Usefulness	… how the technological product assists me in reaching my goals.
Trustworthiness	… how much I can trust the technological product.

6.1. Participants

All participants were practitioners in the field of UX/usability, including various jobs (e.g., UX/UI Designer, UX Engineer, UX Manager, Product Owner, User Researcher, UX Specialist, UX Writer, Usability Tester, UX Architect). A total of $N = 23$ participants completed the questionnaire. They were 24 to 58 years old with an average of 34.87 years ($SD = 9.66$). As for their gender, 60.87% indicated female and 39.13% male. Most participants had a university degree (73.91%) or a PhD (17.39%) in diverse backgrounds, such as psychology, computer science/HCI, and usability/UX. 82.61% of the participants were currently employed and 8.70% each self-employed or working students, covering jobs from UX researchers and designers to UX consultants and engineers, for instance. Most professionals (78.26%) were usually involved in the development of technological products, surveyed users or experts as part of product developments ($M = 4.57$, $SD = 1.08$), and used questionnaires in their professional life ($M = 4.78$, $SD = 1.20$; both scales from 1 (*never*) to 7 (*every day*)).

6.2. Procedure and materials

We implemented an online study in which participants were first screened for whether and what type of professional they are (i.e., whether they are usually involved in technology development or not). Then, they read information on the term of technology transparency and the goals of TechTra and could test both scales by filling them out for a

Table 6
Pearson correlations of the final scales with related variables.

Scale	Demand for Explainability Scale	Overall Transparency Need	Transparency Need (General Items)	Transparency Perception (General Items)	ATI [a] Scale	Need for Cognition Scale	Need for Closure Scale	Experience with Technology	Attitude Towards Technology
TechTra-P	.12	.17*	−.10	.71***	.03	−.13	.24**	.26**	.40***
TechTra-N	.63***	.47***	.80***	−.10	.07	.08	.02	.25**	.07

*p < .05, **p < .01, ***p < .001 (adjusted according to the Benjamini–Hochberg procedure).
[a] ATI = Affinity for Technology Interaction.

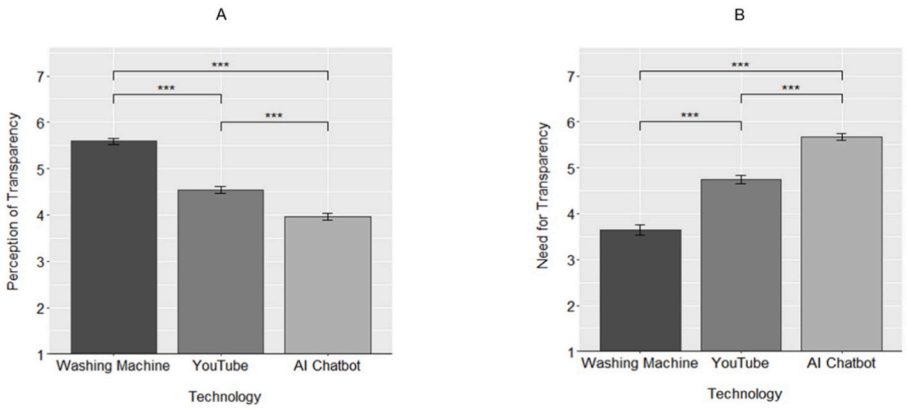

Fig. 1. Differences in transparency perception and need between technologies: Means and standard errors of (A) TechTra-P ratings of transparency perception and (B) TechTra-N ratings of transparency need for the three technologies. *p < .05, **p < .01, ***p < .001 (adjusted according to the Benjamini–Hochberg procedure).

chatbot as exemplary technology. Next, they evaluated TechTra using Likert scales and open-text responses. Precisely, we asked for the applicability, possibility of insights into reasons for responses, ease of use, suggestions for improvement, potentially lacking facets and unsuitable products, as well as the willingness to use. Finally, they voluntarily indicated further remarks, their demographic data, and experiences with surveys and questionnaires. As an incentive, they could participate in a raffle for two 50€ vouchers. The OSF repository shows all materials used (see section 3).

6.3. Data analysis

To reinvestigate the properties of the scales, we first calculated the Cronbach's alpha of both scales and their Pearson correlation. For the evaluation variables (e.g., applicability, ease of use), we then conducted one-sample *t*-tests to analyze whether the professionals' rating differed significantly from the midpoint of the scale as a neutral response. The qualitative data were inductively summarized based on content analysis according to Kuckartz (2014). We enhanced the credibility of this analysis through various strategies, such as letting two experienced researchers with different backgrounds (i.e., psychology, HCI) review the analysis processes and results, as well as building consensus on the implications of the results.

6.4. Results

We found a high reliability for both scales ($\alpha_{\text{TechTra-P}} = .87$, $\alpha_{\text{TechTra-N}} = .93$) and, as in study 1, no significant correlation between the scales ($r(23) = .14$, $p = .537$). Moreover, Table 7 shows the professionals' evaluations, indicating that they saw both scales as applicable and easy to use. They could also gain first insights into the users' reasons for their assessments by looking at the items separately and were willing to use TechTra. These results changed little when we excluded the five participants who were not often involved in technology developments.

For a deeper understanding of the professionals' view, Table 8 shows the qualitative findings, divided into positive and critical remarks.

6.5. Scale adjustments

Overall, the professionals evaluated TechTra favorably. Yet, they mentioned critical points, which we thoughtfully considered. By integrating researchers from different disciplines and consensus building, we decided on two adjustments of TechTra: First, we added an open response field for TechTra-P, asking for experiences that were decisive for the ratings, and two open response fields for TechTra-N, requesting what exactly users want to have transparent and what not. Thereby, we addressed the remark on the depth of insights and implemented a mixed-methods approach, which enables a better understanding of the underlying aspects of users' experiences (Naveed et al., 2024). Moreover, such qualitative responses can provide possible alternative reasons for ratings that the professionals also commented on. Yet, as the remark "importance of other constructs" suggests, experiences can be driven by contextual and individual factors. Such influences can additionally be ruled out during the analysis (e.g., by measuring AI literacy and integrating it as a covariate). Second, we altered the instructions for TechTra-P and TechTra-N (in parentheses) to "Please think of how transparent you experienced (wish) the technological product to be and rate your perceptions (needs) in terms of the statements below.", addressing the comment on the scales' presentation. However, we refrained from highlighting keywords as this is not possible in every software for implementing questionnaires.

We did not directly implement the other remarks for various reasons. Concerning the relevance for specific products/contexts, we already excluded items that might be difficult to answer across use cases in study

Table 7
Means, standard deviations, and *t*-test results for professionals' evaluation of TechTra.

Evaluation Variable [a]	*M*	*SD*	*t*(22) [b]	*p*	Cohen's *d*
Applicability to investigate user transparency perception	5.57	1.08	6.95	<.001	1.45
Applicability to investigate user transparency need	5.48	1.24	5.72	<.001	1.19
Insight into user's reasons for transparency perception	5.43	1.08	6.37	<.001	1.33
Insight into user's reasons for transparency need	5.17	1.30	4.32	<.001	0.90
Ease of use	6.04	1.43	6.86	<.001	1.43
Willingness to use	5.26	1.54	3.92	<.001	0.82

[a] Evaluation variables were rated by $N = 23$ professionals on a scale from 1 to 7.
[b] The *t*-test results indicate whether professionals' evaluation of TechTra differed significantly from the neutral value of 4.

Table 8
Qualitative results of the professionals' evaluation of TechTra.

Category	Description	Example Quotes	Frequency [a]
Positive remarks			
Usefulness	TechTra was perceived as useful for research and inquiry into user needs, especially due to its standardization and the possibility of using it iteratively during product development.	"A standardized questionnaire for transparency would benefit my research in analyzing users' standpoints and perceptions […]." "I can well imagine asking about expectations at the start of a new product […] and then tracking […] transparency […] during iterative development."	5 (21.7%)
Easiness	TechTra was perceived as clear, easy to use, and easy to answer.	"The questions are very clear and straightforward." "Questions are […] easy to answer"	5 (21.7%)
Variety of aspects	TechTra encompasses various important aspects of transparency.	"I think the questionnaire asks a broad range of questions directed to transparency." "I have the feeling that important (partial) aspects of transparency are addressed."	4 (17.4%)
Overview	TechTra offers an overview and first impression of user transparency needs and perceptions.	"Since it is about overall perception or overall need, I think it is good that the items are formulated relatively generically." "The questions provide an initial and general overview."	3 (13.0%)
Wording	The language was perceived as easily understandable.	"Language used is not too technical, that is good." "The items are comprehensible, and it is easy to understand which aspect of transparency they are aimed at."	3 (13.0%)
Presentation	The 7-point scale and item order led to a positive evaluation of how the items were presented.	"I think it's good that a 7-point scale is used."	3 (13.0%)
Target-performance comparison	Combining the TechTra Scales makes a direct target-performance comparison possible, also due to the similar wording.	"I […] think it's very good that the same […] wordings are used in […] the two questionnaires. Comparability of expectations and actual implementations is […] easily possible."	2 (8.7%)
Applicability to contexts/ products	TechTra can be used in various contexts or for various products.	"Quick inquiry possible in different settings."	2 (8.7%)
Critical remarks			
Depth of insights	TechTra may provide more insights into reasons for users' answers, problematic product features, and design solutions. Open questions about reasons and wishes were desired.	„the questions are too closed to get a more specific insight of a user's reason for his or her answers" "I think open formats would also be necessary in any case to have a better understanding of why [users] have a lower need"	9 (39.1%)
Relevance for specific products/contexts	The relevance of items could depend on the technology and use context. Answers for complex products might vary for contexts, whereas for simple products, some items might be too detailed.	"I can […] imagine that certain items are less relevant for one product or another." "For a simple website […] the questions about 'understanding the current activity of the product' are too in-depth."	7 (30.4%)
Presentation	TechTra's presentation could be improved by highlighting keywords and shortening the introduction while putting more emphasis on the scale's goal.	"the keywords in each line could be in bold so that it is immediately obvious what it is about." "The introduction to the questionnaire could be more concise but more structured."	5 (21.7%)
Wording	Due to the similar wording, items may be interrelated and hard to differentiate. Also, the understandability could be improved by more practically oriented wording.	"1-2 questions are formulated very similarly, perhaps formulate in a more selective way." "Might be beneficial to make questions more tangible, some are hard to understand."	4 (17.4%)
Length of scales	Fewer items may be beneficial for practical use where transparency is just one aspect.	"A short-scale would be nice, maybe single-item or 3 items only." "Generally, the questions should be shorter."	4 (17.4%)
Limitations of need assessment	TechTra-N items might not measure true need due to ceiling effects or users' insufficient insight into their needs. Asking about the need might influence the perceived need.	"The second questionnaire could show ceiling effects." "The participant may not be aware of their needs."	4 (17.4%)
Alternative reasons for ratings	Items could be interpreted differently, as transparency might be influenced by other factors (e.g., previous technology experience, trust in the company, dependence on the product).	"Some parts […] might be clear for the user from the amount of interaction they already had […] and not due to the displayed transparency." "[My need] depends on how much I trust the providing company, how is the overall experience […], how much I need this product and whether I have time pressure."	3 (13.0%)
Importance of other constructs	TechTra may integrate further interesting constructs.	"No item that relates to 'AI literacy'." "Maybe it would help to measure different constructs […]. In many cases, there are different variables which are interesting."	3 (13.0%)
Unsuitability of trustworthiness aspect	Trust was perceived as a consequence of transparency rather than a facet on a similar level.	"To me, it can be transparent how a product works, what it does, which data it uses, etc., and the answers to that then build trust."	1 (4.3%)

[a] Amount and percentage of the 23 participants who commented on the respective category.

1 (e.g., *error details*). Also, participants can always choose the scale midpoint as a neutral response. Nevertheless, transparency perceptions, needs, and the items' relevance should depend on the use case, and this dependence is a crucial reason for target-performance comparisons. Since twice as many participants commented positively on the wording and easiness as negatively, we did not alter the formulations but will be mindful of future remarks of participants. Future studies could also develop an even shorter version of the scales for a "quick and dirty" evaluation, but this would not capture all transparency facets and therefore reduce the insights gained. The mentioned limitations of the need assessment could be ruled out via study 1, as we showed significant differences between technologies. Nevertheless, further investigations should more deeply examine transparency needs in diverse use contexts. Lastly, we agreed with the critical remark that *trustworthiness* is more a consequence of transparency, but TechTra measures *trustworthiness* as a facet rather than an outcome of transparency (i.e., is it transparent to users whether they can trust the technology, rather than if they actually trust it).

7. Study 3: Application in real technology interaction and evaluation of construct validity

After refining the TechTra Scales through the second study, we applied TechTra to a real technology interaction and explored whether it is sensitive to changes in the technology's design and what relations exist to other constructs in terms of construct validity. Therefore, we posed the following research questions (RQs):

- RQ1: How can the developed scale for measuring transparency perception (TechTra-P) reflect differences in the design of a technology in a real interaction?
- RQ2: What effects does the design have on the developed scale for need for transparency (TechTra-N)?
- RQ3: How do TechTra-P and TechTra-N ratings relate to associated constructs?

More precisely, we assumed this hypothesis for RQ1:

- Hypothesis: Users' perception of transparency is higher with an explanation of the technology's functioning than without such an explanation.

To investigate the RQs and hypothesis, we conducted a Wizard of Oz (WoZ) study with the humanoid Pepper robot and varied the technology design (no explanation vs. explanation) by the robot not explaining or explaining the functioning of an age estimation and a plant recognition (see Fig. 2 for Pepper's responses during the age estimation and the OSF repository mentioned in section 3 for the whole interaction scripts). WoZ means that participants believe they are interacting with an autonomous robot that is actually controlled by a human, the so-called wizard, sitting in an adjacent room (Kelley, 2018; Schecter et al., 2023). We chose the two robot functions (i.e., age estimation and plant recognition) because they allowed for a manipulation close to the central elements of the transparency definition, have similar characteristics to not dilute effects (e.g., both rely on visual cues), and enabled consistent responses of the robot across participants (e.g., same deviation of the age estimate from the true age for each participant). In the following, we describe the participants of the study, further information on the procedure and materials, and the data analysis methods. Then, we present the results separated by the three RQs and the scale analysis to again test the scale properties.

7.1. Participants

Participants were not considered if they did not rate all items of TechTra-P and answered the attention check item incorrectly ($n = 2$). Moreover, we excluded participants if they failed both manipulation checks, which asked whether the robot indicated how it estimated the age and identified the plant ($n = 6$). Thus, the final sample consisted of $N = 73$ participants with $n = 39$ in the no explanation and $n = 34$ in the explanation condition.

Participants' ages ranged from 18 to 74 years, with an average of 26.79 years ($SD = 12.18$). 60.27% identified themselves as female, 38.36% as male, and 1.37% as having a diverse gender. The highest academic degree of most participants was a university entrance qualification (57.53%) and a university degree (35.62%). As employment status, 75.34% indicated being a student, 10.96% an employee (including public officials), and 5.48% a pensioner.

7.2. Procedure and materials

In the first part, participants indicated their demographics and experience with and attitude towards robots (items based on Heerink et al., 2009; Koverola et al., 2022; Nomura, 2014). Additionally, an item tested if they correctly identified the three plants lying in front of them (i.e., rose, sunflower, and tulip). Then, participants had an audio-recorded conversation with the Pepper robot under the guidance of the investigator, who followed a script (see OSF repository mentioned in section 3). Comparability was also enhanced by using a wizard who played predefined robot behaviors during the interaction according to the randomly assigned condition with or without explanation. In the condition with explanation, Pepper provided a verbal explanation of how it estimated the participant's age and recognized a plant shown by the participant, whereas in the condition without explanation, such explanations were not given. Participants were consistently estimated to be one year older based on the demographic information indicated at the

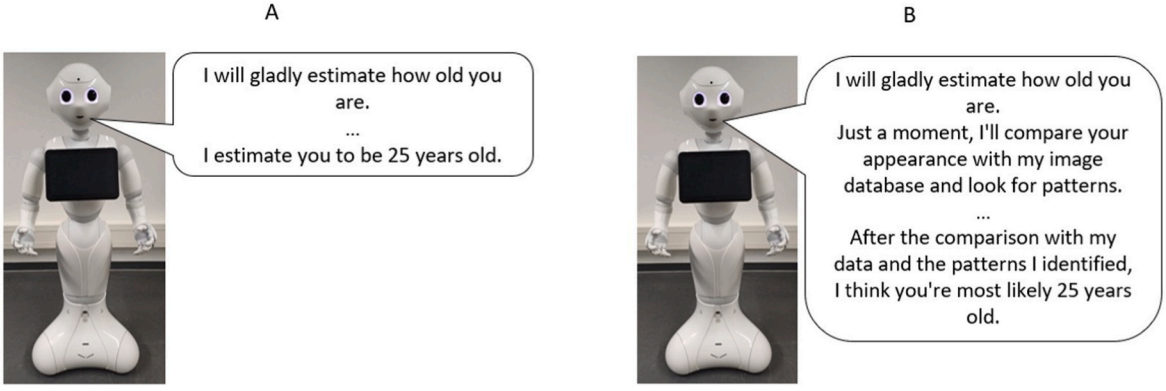

Fig. 2. Variation of the technology design: A) Age estimation without explanation, B) Age estimation with explanation. The age estimate was adjusted by the WoZ by adding one year to the participant's actual age.

beginning, and the plant recognition was always correct due to an audio and video connection to the wizard. Figs. 2 and 3 illustrate the study's setting. After the interaction, participants filled out TechTra-P and TechTra-N (see Table 5). Subsequently, they completed the *Transparency of Robots Scale* (Angelopoulos et al., 2024), *AttrakDiff* (Hassenzahl et al., 2003), an adapted version of the *ISONORM 9241/10-S* (Pataki et al., 2006), the *Positive and Negative Affect Schedule* (PANAS) (Watson et al., 1988), and *INTUI* (Ullrich & Diefenbach, 2010). Next, the manipulation checks (see section 7.1) followed and a question each about the reasons they assumed for the robot's age estimation and plant identification. Afterwards, their technology self-efficacy (Hughes, 2013) and preferred compensation (course credit or 8€) were assessed. Lastly, the investigator noted the plant selected by the participant and any technical errors or irregularities during the testing. The measures and Cronbach's alpha for all subscales are available in the OSF repository (see section 3).

7.3. Data analysis

To investigate the effect of the technology design on the TechTra-P ratings (RQ1), we computed a two-sample, two-sided *t*-test. Additionally, we checked the robustness of the effect using an ANCOVA with the attitude towards and experience with robots as covariates. For comparisons at the item level, a profile diagram was created. We also calculated a two-sample, two-sided *t*-test to compare the technology design conditions in terms of the TechTra-N ratings (RQ2). To further analyze the relation between the transparency perception, transparency need, and technology design conditions, we integrated a paired *t*-test comparing the TechTra-P and TechTra-N ratings, as well as a mixed ANOVA with the conditions as a between-subjects and the TechTra Scales as a within-subjects factor. Exploring the relations of the TechTra-P and TechTra-N ratings with associated constructs (RQ3), we looked at Pearson correlations of the TechTra Scales and the constructs studied (i.e., transparency of robots, interaction principles of the ISONORM 9241/10-S, positive and negative affect, intuitive use, technology self-efficacy).

Additionally, we performed a scale analysis as in the first study. Therefore, we calculated the Cronbach's alpha and a PCA for each scale. For each item, we checked the distribution of responses (i.e., mean, median, standard deviation, range, skewness, kurtosis) and the corrected item-total correlation.

7.4. Results

7.4.1. Effects of technology design on transparency perception

Supporting our hypothesis and the applicability of TechTra to a real technology interaction, TechTra-P ratings were significantly lower in the no explanation ($M = 3.96$, $SD = 1.35$) than the explanation condition ($M = 4.69$, $SD = 1.21$), $t(70.92) = -2.43$, $p = .018$, $d = 0.57$. The effect was even stronger when ruling out the influences of robot attitudes and experiences through the ANCOVA, $F(1,69) = 6.95$, $p = .010$, $\eta^2 = .08$. Specifically, participants with a positive attitude towards robots generally perceived the transparency to be higher, $F(1,69) = 13.96$, $p < .001$, $\eta^2 = .15$, while the experience with robots had no significant effect on TechTra-P ratings, $F(1,69) = 2.00$, $p = .162$. Furthermore, Fig. 4 shows that participants had a higher transparency perception on all items in the explanation than the no explanation condition. This was also reflected in the qualitative responses. In the explanation condition, statements occurred such as „When pepper tried to guess my age […], he precisely explained how he managed to do so.“, while in the no explanation condition, an exemplary statement was “Analyzing the face and flower, it is impossible to understand how the robot arrives at the results.”.

7.4.2. Effects of technology design on transparency need

Regarding TechTra-N, we did not find a significant difference between the conditions, $t(70.79) = -0.35$, $p = .730$, as assumed. At the item level, participants especially wanted to have transparency about how far they can trust the robot ($M = 6.45$, $SD = 1.17$) and what personal data it uses ($M = 6.22$, $SD = 1.08$). Less relevant to them was how the robot works ($M = 5.26$, $SD = 1.70$) and what it will do next ($M = 5.42$, $SD = 1.42$). Accordingly, participants indicated in the qualitative responses that they do not want precise information on the functioning (e.g., “I don't need to know exactly in detail how the technology works.”). Likewise, statements referred to the robots' next actions (e.g., “I don't think that it's important to have the robot explaining during the conversation what it will do next […]. This would make any conversation feel unnatural […].”).

The paired *t*-test underlined the participants' different ratings on TechTra-P and TechTra-N in that the transparency perception ($M = 4.30$, $SD = 1.33$) was lower than the transparency need ($M = 5.87$, $SD = 0.86$), $t(71) = 9.03$, $p < .001$, $d = 1.06$. In addition, the interaction effect of the condition and the TechTra scale in the mixed ANOVA was not significant, suggesting that the chosen manipulation did not significantly lead to a higher need fulfillment, $F(1,71) = 3.73$, $p = .057$.

7.4.3. Relation of transparency perception and need with associated constructs

Table 9 presents the correlations between the TechTra Scales and associated constructs. Like in the other two studies, transparency perception and need did not correlate significantly. Moreover, as expected, especially TechTra-P showed significant correlations with the scales measured.

7.4.4. Scale analysis

Cronbach's alpha indicated a good to high reliability of both scales ($\alpha_{\text{TechTra-P}} = .91$, $\alpha_{\text{TechTra-N}} = .79$) (Kline, 2016). Supplementary material C also presents the descriptive values, corrected item-total correlations, and results of the principal components analyses. As suggested by the high internal consistency of TechTra-P, results showed high item-total correlations, response distributions close to a normal distribution, and

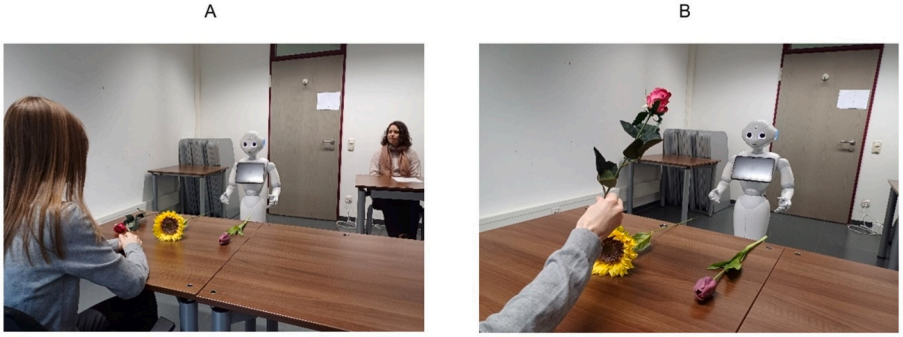

Fig. 3. Setting of the study: A) View of the participant on the Pepper robot and the investigator, B) Participant showing Pepper one of the plants to recognize.

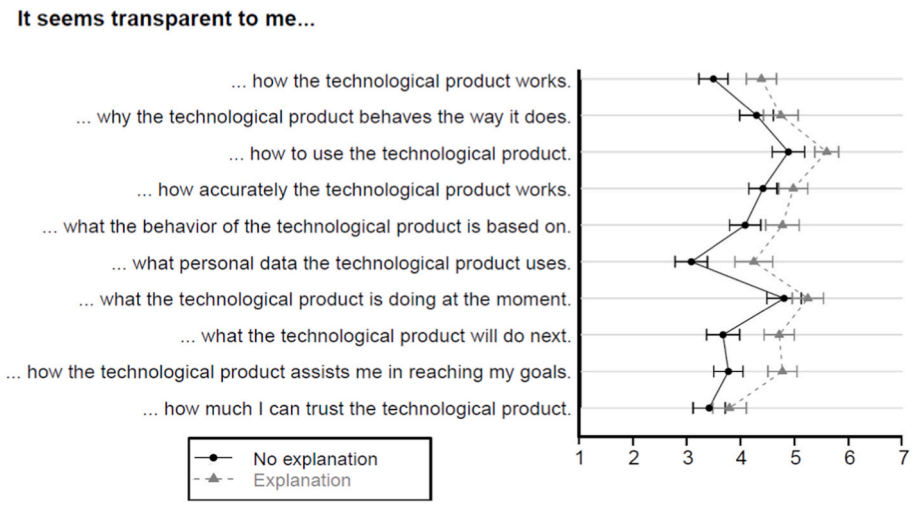

Fig. 4. Differences in transparency perception at the item level: Average ratings and standard errors of TechTra-P ratings for each item, separated by the technology design conditions (no explanation vs. explanation).

Table 9
Pearson correlations of TechTra-P and TechTra-N with related constructs.

	TechTra-P	TechTra-N
TechTra-P	—	.13
TechTra-N	.13	—
Transparency of Robots Scale		
– Illegibility	−.62***	−.30*
– Explainability	.59***	.15
– Predictability	.61***	.20
AttrakDiff		
– Attractiveness	.27*	.36**
– Pragmatic quality	.38**	.33*
– Hedonic quality	.10	.24
ISONORM 9241/10-S		
– Suitability for the user's tasks	.25	.17
– Self-descriptiveness	.42**	.02
– Controllability	.16	.01
– Conformity with user expectations	.35**	.18
– Use error robustness	.41**	.12
– User engagement	.29*	.20
– Learnability	.34**	.13
Positive and Negative Affect Schedule (PANAS)		
– Positive affect	.13	.21
– Negative affect	−.07	−.08
INTUI Questionnaire		
– Gut feeling	−.24	−.21
– Effortlessness	.28*	.15
– Magical experience	.15	.21
– Verbalizability	.52***	.12
Technology Self-Efficacy	.23	.23

*p < .05, **p < .01, ***p < .001 (adjusted according to the Benjamini–Hochberg procedure).

loadings on the first component of at least .66. For some items, TechTra-N ratings differed more from a normal distribution, had lower item-total correlations and loadings. The results are likely due to the concrete operationalization in the study, since especially the items with the lowest and highest ratings were affected. For instance, the leptokurtic (i.e., sharper) distribution of the *trustworthiness* item suggested that participants consistently had a high need for insights on the robot's trustworthiness. Likewise, the lower item-total correlation and loading on the first component of the *how* item represented that this item received the lowest need ratings. Nevertheless, parallel analyses of both scales suggested the existence of one component.

8. Discussion

Users' experience of a technology's transparency gained significance in HCI research and industry, but a broadly applicable and comprehensive UX tool for assessing and comparing users' transparency needs and perceptions was still missing. We closed this research gap by developing the *TechTra Scales* – consisting of *TechTra-N* for measuring transparency need and *TechTra-P* for assessing transparency perception – in three consecutive studies:

- In the *first study* ($N = 191$), we selected ten items for each scale based on users' experience with three different technologies. The item choice depended on quantitative and qualitative analyses and the coverage of key facets of transparency, while simultaneously aiming for an economic, concise measurement. Furthermore, we obtained promising results for the scales' validity, such as correlations with related measures but not between the scales, and expected differences between the technologies.
- The *second study* tested the scales with UX and usability professionals ($N = 23$). The experts evaluated the scales favorably, and both scales were highly reliable without correlating significantly. Thus, only small adjustments were made by adding in-depth qualitative questions and revising the introduction.
- In the *third study* ($N = 73$), we applied the scales to a real robot interaction, in which we varied the robots' design in a between-subjects design (no explanation vs. explanation). As assumed, TechTra-P ratings were higher in the explanation than the no explanation condition, while TechTra-N ratings did not differ between groups. Again, meaningful relations with associated measures, the independence of both scales, and expected scale properties were found.

The results of all studies are discussed in the following sections, focusing first on the importance of target-performance comparisons for technology transparency and the validity of the scales. Subsequently, we present the theoretical contributions and practical implications of our research and scales. To conclude, we describe limitations of the studies along with central areas for future research.

8.1. Relevance of target-performance comparison in technology transparency investigations

As stated in previous research on technology transparency (Barman et al., 2024; Liao et al., 2020; Liao & Varshney, 2022) and consistent

with its increasing importance in the HCI and UX field (Andrada et al., 2022; European Commission, 2019; Usmani et al., 2023), transparency was shown to be a need of users in the studies conducted. This was apparent, for instance, in study 3 where the participants provided on average a high TechTra-N rating (i.e., 5.87 out of 7 scale points). However, the need's relevance depended on the technological product. For example, users had a higher transparency need for AI-based chatbots than for YouTube and their washing machine (see study 1), suggesting that one cannot simply assume a consistently high need level and that need assessments are crucial to implement adequate transparency designs with regard to UX. Furthermore, transparency needs were also found to depend on the use context and user characteristics (Cech, 2021; Hein & Diefenbach, 2025; Huff Jr et al., 2021; You et al., 2023, see also qualitative responses in studies 2 and 3), which makes user needs even more of a moving construct and context-specific assessments essential. Likewise, the perceptions of transparency altered between technologies (study 1) and their designs (study 3). They also differ between target groups (Huff Jr et al., 2021; Liao & Varshney, 2022) and the technology's use cases (e.g., according to the users' experience and knowledge (Bitzer et al., 2023; DIN Deutsches Institut für Normung e.V., 2023; Hein & Diefenbach, 2025)), which renders the measurement of transparency perceptions non-trivial for UX research.

Moreover, transparency needs and perceptions were independent in all three studies, showing that it is not possible to draw conclusions from one to the other and that a separate assessment is useful – and thus the implementation of both TechTra-P and TechTra-N. A comparison of the scales can then determine whether a product is designed insufficiently or overly transparently to meet users' subjective needs. More precisely, the TechTra Scales can identify relevant transparency facets for a particular technology and potential gaps in their realization from a UX point of view. This enables design recommendations and a well-adjusted transparency design through an iterative use of the scales.

8.2. Supported validity of the TechTra Scales

The validity of the TechTra Scales was supported in the three studies conducted. We showed meaningful relations of the scales with associated constructs and validated measures, and corresponding to the non-significant correlation between TechTra-N and TechTra-P, different patterns for both scales. For instance, TechTra-N correlated more strongly with the Demand for Explainability Scale and users' general transparency need than TechTra-P, while TechTra-P was more closely linked to users' general transparency perception (see study 1). Similarly, all factors of the Transparency of Robots Scale correlated highly with TechTra-P but not with TechTra-N in study 3. The results indicate that while TechTra overlaps with related constructs in terms of construct validity, it also captures distinct aspects, supporting the need to study technology transparency separately from other UX constructs. For example, TechTra-P had positive, yet mostly moderate relationships with established design guidelines (e.g., self-descriptiveness, conformity with user expectations), the technology's attractiveness, and its intuitive use. The results further highlight that transparency is rather a pragmatic quality facilitating an effective and efficient technology use than a stimulating, hedonic quality. Correspondingly, transparency did not significantly correlate with affect and magical experience during technology use. Interestingly, participants with higher transparency perceptions indicated acting more effortlessly but at the same time more deliberately with an improved ability to verbalize operating steps, suggesting that transparency enhances cognitive processing and mental model development without impairing the users' experience.

Furthermore, participants' interindividual characteristics (i.e., affinity for technology interaction, need for cognition, need for closure, technology self-efficacy) mostly did not correlate significantly with the TechTra Scales. This again emphasizes that the transparency need is context-dependent and worth investigating for each technology. Nevertheless, users' technology-specific characteristics (i.e., their experience with and attitude towards the particular technology) may serve as first indications for users' transparency perceptions and needs since they positively influenced TechTra-P and TechTra-N in studies 1 and 3.

Additionally, differences in TechTra ratings between technologies and technology designs matched our assumptions and previous results (Bitzer et al., 2023; Hein & Diefenbach, 2025; Wang & Yin, 2021). Hence, the perceived transparency was highest for the washing machine, significantly lower for YouTube, and lowest for AI-based chatbots, whereas the opposite was the case for the transparency need (see study 1). This reflected identified factors of users' transparency experiences, such as the complexity of and familiarity with the technology (Bitzer et al., 2023; Hein & Diefenbach, 2025; Wang & Yin, 2021). Also, our variation of the robot's design in study 3 affected TechTra-P but not TechTra-N ratings. This was reasonable since we just manipulated whether the robot provided an explanation for its behavior but not altered the use context. Moreover, we made only small changes between the groups to test how sensitive TechTra-P is to design modifications. For instance, we implemented slight non-verbal transparency cues such as head movements (e.g., focusing on the user's face for the age estimation) in both conditions and merely added the verbal explanation in the explanation condition. Nevertheless, medium-to-high effect sizes emerged, which support the sensitivity of TechTra.

8.3. Theoretical contributions

Our work provides several theoretical contributions to current research on HCI and technology transparency. First, we contributed to unravelling the term and method confusion by clarifying the understanding of transparency and its role among other terms (e.g., explainability, interpretability). In study 1, we also found that transparency as a term was perceived as highly understandable by users, which further underlines its suitability for UX research. Moreover, we presented an overview of existing UX measures for technology transparency and their shortcomings and developed the TechTra Scales to address the limitations. Thus, our research can serve as guidance for future work, bringing together various lines of research (e.g., on the different terminologies).

Second, we compiled key facets of transparency as the respective research landscape is confusing with overlapping terms, diverse labels for the same facet, and different levels of granularity. We carefully selected and empirically evaluated the facets, forming a set of crucial facets that concretize the definition of technology transparency and should be considered when evaluating transparency perceptions and needs. Moreover, the aspects can serve as an inspiration for transparency design by showing what can be made transparent.

Third, we gained insights into the relations between technology transparency and central constructs in existing research, such as the attractiveness of a technology, its interaction principles, and intuitive use perceptions. This helps to embed transparency into the network of other UX constructs and provides a basis for further tests on construct validity. In addition, it facilitates the selection of appropriate UX measures for concrete research projects and technology developments. Importantly, we also showed that differentiating between transparency perceptions and needs is crucial, which was neglected in prior research.

8.4. Practical application of the TechTra Scales

The TechTra Scales can be used both in research and industry. In research, TechTra-N can be applied to systematically investigate transparency needs for diverse technologies, use contexts, and different stakeholders. More precisely, the TechTra-N items can reveal differences in the relevance of the individual transparency facets from a UX perspective. TechTra-P can separately be integrated into studies to test transparency perceptions of specific technology designs and to compare them. Again, researchers can identify possible reasons for participants' perceptions by focusing on the items. Moreover, combining the scales is

useful for testing the effectiveness of transparency manipulations in explainable AI (xAI) and adjacent research areas. Overall, using the TechTra Scales as a validated and standardized tool should increase the soundness of findings and the comparability between studies.

In the industry, TechTra-N can serve as a need analysis before designing a technological product. The scale shows how relevant a transparent design is for the specific product based on user assessments and which transparency facets should be prioritized accordingly. After developing a prototype, manufacturers can then apply TechTra-P and test the alignment of the users' perceptions with the need analysis. This facilitates identifying target-performance gaps and improving the users' experience. More detailed insights and ideas for improvement can be gained by examining users' qualitative responses alongside the quantitative ratings (see studies 2 and 3). The scales can also be used directly one after the other, as the three studies showed, which is advantageous if a product is already developed and designed or if the product development process needs to be accelerated. Nevertheless, the legally regulated and ethically driven requirements for transparency as well as the accuracy of users' mental models should always be taken into account (e.g., European Commission, 2025; Narayanan, 2023), while the TechTra Scales should additionally be used for a UX-based optimization of the transparency design.

In general, the scales are applicable to previous, accumulated experiences of users (see study 1) and recently experienced technology interactions (see study 3), which allows for a closer look at different applications and functions of the technologies. In research, this can also help to investigate how single transparency experiences influence the overall memory-based experience of the product or how past experiences affect the perception of a new interaction.

8.5. Limitations and future research

The present research has some limitations that should be addressed in future studies. Although we investigated interindividual differences (e.g., experience with the technology, affinity for technology interaction), there are more such variables that may influence TechTra ratings. For example, we did not examine cultural differences that likely impact transparency needs and perceptions, according to previous research (Chien et al., 2024; Kang et al., 2025; Peters & Carman, 2024). Chien et al. (2024) found that Western cultures perceived data privacy as more important and trusted AI services less than Eastern cultures. Considering a greater prevalence of AI in Eastern countries (IBM, 2024), this reflects our finding that transparency perceptions tend to be higher with more technology experience. Moreover, not only technology-related but also domain-related experiences and knowledge (i.e., about the technology's application domain, such as healthcare or finance) could alter scale ratings (Hein & Diefenbach, 2025; Sokol & Flach, 2020). Additionally, the predominantly young age of participants in our studies could have influenced our results, highlighting the importance of further assessing differences between various user groups and characteristics.

Also, the construct validity of the TechTra Scales and their relations to other measures should be further investigated, especially for those constructs that showed poor internal validity and Cronbach's alpha below <.60 (Kline, 2016), that is, the Need for Cognition Scale in study 1 and the interaction principle of user error robustness in study 3. For other measures, such as the gut feeling subscale of INTUI, we found questionable values between .60 and .70 (Kline, 2016). Therefore, relations of the TechTra Scales to these measures should be reexamined. Moreover, correlations with further constructs and alternative measures for the included constructs should be explored. For instance, the *Questionnaire for the Subjective Consequences of Intuitive Use (QUESI)* (Naumann & Hurtienne, 2010) can be used in addition to the INTUI. As other constructs, one could include UX and usability (e.g., *User Experience Questionnaire (UEQ)* (Laugwitz et al., 2008), *System Usability Scale (SUS)* (Brooke, 1996)) and technology acceptance (e.g., items of the *Unified Theory of Acceptance and Use of Technology (UTAUT)* (Venkatesh et al., 2003) or *Technology Acceptance Model (TAM)* (Venkatesh & Bala, 2008)). For our studies, we carefully selected measures to avoid overly long surveys that could impair the participants' concentration and data quality, excluding several constructs due to their overlap with integrated scales (e.g., the UEQ measures the same facets as the AttrakDiff), their looser connection to technology transparency, and their inconsistent definition in research (e.g., technology acceptance). Yet, these measures should now be examined to refine the construct network of technology transparency.

Furthermore, TechTra should be applied to other technological products and use cases. We already used a broad range of technologies from mature to emerging and physical to digital products (i.e., washing machine, YouTube, AI-based chatbot, humanoid robot). However, other technology application areas should be investigated to determine the scales' applicability to diverse domains, such as manufacturing (e.g., automation systems), healthcare (e.g., medical imaging systems), automotive (e.g., self-driving vehicles), finance (e.g., fintech solutions), and e-commerce (e.g., recommendation algorithms). Thereby, it would be especially valuable to differentiate between technologies used in lower-stakes and high-risk, probably privacy-sensitive situations (e.g., video games vs. biometric authentications), as we did not include higher-risk use cases. For instance, we will apply the scales in hospital and digital health settings and investigate the influence of perceived technology relevance on transparency experiences. Moreover, future studies should include further technologies with different design challenges, such as Extended Reality (XR), which combines physical and digital components (Norval et al., 2023), and incorporate single- and multi-purpose systems (e.g., calculators vs. smartphones). Accordingly, multiple purposes of one technology should be examined and compared on the basis of hypotheses relating to differences in TechTra ratings.

Moreover, a closer investigation of TechTra-N is needed since we manipulated the transparency perception in study 3 but not the need. This could also be done via different use cases and stakes, as suggested in previous research (Hein & Diefenbach, 2025; Liebherr et al., 2025). For instance, transparency needs should depend on the existence of barriers to goal achievement, such as technology errors (Konopka et al., 2024) or the degree of expectation fulfillment (de Zoeten et al., 2024; Riveiro & Thill, 2022; Springer & Whittaker, 2020). A diverse manipulation of transparency needs could also reveal the extent of users' potential tendency to give socially acceptable answers, which could have influenced the participants' ratings in our studies. Manipulations of transparency needs and perceptions can additionally target specific transparency facets of the TechTra Scales (e.g., if the technology makes it transparent what personal data is used), which should then be reflected in the rating of the respective item. Nevertheless, future research should ideally cover evaluations based on immediate technology interactions (as in study 3) and on prior experiences (as in study 1) to further test the applicability of the scales to both contexts. In addition, field studies are necessary to investigate the scales in real-world settings with more critical consequences of (lacking) transparency for users (Naveed et al., 2024). This should also include the use of the scales during technology development in industry, as this could provide more insights into the critical aspects (e.g., ceiling effects of need ratings, item wording) considered by the professionals in study 2 and show further areas for improvement. This approach would also provide more insights into different perspectives on the practical applicability of the scales, as the sample of professionals in study 2 was rather small.

9. Conclusion

The paper presents the TechTra Scales as a new tool to measure needs for and perceptions of technology transparency (see Supplementary material B for a ready-to-use template of the scales). Addressing prior shortcomings, the scales offer a widely applicable and comprehensive method for low-threshold target-performance comparisons to inform design recommendations from a UX perspective. Developed through

three consecutive studies, the scales showed good reliability and construct validity results with meaningful relations to associated measures. In addition, the non-significant correlation between the need (TechTra-N) and perception (TechTra-P) scales highlights the importance of distinguishing between both constructs. The scales also revealed differences between a broad range of technologies and different technology designs. Moreover, UX and usability practitioners evaluated the scales as easily applicable and would use them in their professional lives, underlining the practical relevance of TechTra for product development in industry. For researchers, the tool and paper are of theoretical and practical interest as they help unravel the term and method confusion in technology transparency research, introduce key transparency facets, clarify relations between technology transparency and other crucial constructs, and provide a validated tool for future studies. Future research should further investigate the scales' validity and use them in diverse contexts to test their applicability and sensitivity to design and situational changes.

CRediT authorship contribution statement

Ilka Hein: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Daniel Ullrich:** Writing – review & editing, Validation, Methodology, Conceptualization. **Sarah Diefenbach:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Ethics statement

All procedures were performed in compliance with relevant laws and institutional guidelines and have been approved by the university's ethics committee (reference numbers and dates: EK-MIS-2024-280 on 17/07/2024, EK-MIS-2024-304 on 02/09/2024, EK-MIS-2025-0335 on 20/01/2025). The privacy rights of human subjects have been observed, and written informed consent was obtained for experimentation with human subjects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the German Research Foundation (DFG), projects PerforM and TransforM [grant number 425412993], as part of the Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

Supplementary materials

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.chbr.2026.101025.

Data availability

The materials, data sets, and analysis scripts of all studies are available in the OSF repository (see https://osf.io/w8g3k/overview?view_only=b0f496f48cf44818990bb3f58fad147d).

References

Alonso, V., & de la Puente, P. (2018). System transparency in shared autonomy: A mini review. *Frontiers in Neurorobotics, 12*, Article 83. https://doi.org/10.3389/fnbot.2018.00083

Andrada, G., Clowes, R. W., & Smart, P. R. (2022). Varieties of transparency: Exploring agency within AI systems. *AI & Society, 38*, 1321–1331. https://doi.org/10.1007/s00146-021-01326-6

Angelopoulos, G., Lacroix, D., Wullenkord, R., Rossi, A., Rossi, S., & Eyssel, F. (2024). Measuring transparency in intelligent robots. *ArXiv, 1–16*. https://doi.org/10.48550/arXiv.2408.16865

Barman, K. G., Wood, N., & Pawlowski, P. (2024). Beyond transparency and explainability: On the need for adequate and contextualized user guidelines for LLM use. *Ethics and Information Technology, 26*, Article 47. https://doi.org/10.1007/s10676-024-09778-2

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Beißert, H., Köhler, M., Rempel, M., & Beierlein, C. (2014). *Eine deutschsprachige Kurzskala zur Messung des Konstrukts need for cognition: Die need for cognition Kurzskala (NfC-K)*. GESIS - Leibniz-Institut für Sozialwissenschaften. https://www.ssoar.info/ssoar/handle/document/40315.

Bitzer, T., Wiener, M., & Cram, W. A. (2023). Algorithmic transparency: Concepts, antecedents, and consequences – A review and research framework. *Communications of the Association for Information Systems, 52*, 293–331. https://doi.org/10.17705/1CAIS.05214

Brennen, A. (2020). What do people really want when they say they want "Explainable AI?" we asked 60 stakeholders. In R. Bernhaupt, F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. Zhao, B. P. Samson, & R. Kocielnik (Eds.), *Extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1–7). ACM. https://doi.org/10.1145/3334480.3383047.

Brooke, J. (1996). SUS: A 'Quick and Dirty' usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry*. CRC Press.

Buçinca, Z., Lin, P., Gajos, K. Z., & Glassman, E. L. (2020). Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In F. Paternò, N. Oliver, C. Conati, L. D. Spano, & N. Tintarev (Eds.), *Proceedings of the 25th international conference on intelligent user interfaces* (pp. 454–464). ACM. https://doi.org/10.1145/3377325.3377498.

Bujold, A., Parent-Rocheleau, X., & Gaudet, M.-C. (2022). Opacity behind the wheel: The relationship between transparency of algorithmic management, justice perception, and intention to quit among truck drivers. *Computers in Human Behavior Reports, 8*, Article 100245. https://doi.org/10.1016/j.chbr.2022.100245

Cai, C. J., Jongejan, J., & Holbrook, J. (2019). The effects of example-based explanations in a machine learning interface. In W.-T. Fu, S. Pan, O. Brdiczka, P. Chau, & G. Calvary (Eds.), *Proceedings of the 24th international conference on intelligent user interfaces* (pp. 258–262). ACM. https://doi.org/10.1145/3301275.3302289.

Calvaresi, D., Carli, R., Tiribelli, S., Buzcu, B., Aydogan, R., Di Vincenzo, A., Mualla, Y., Schumacher, M., & Calbimonte, J.-P. (2025). Computational persuasion technologies, explainability, and ethical-legal implications: A systematic literature review. *Computers in Human Behavior Reports, 17*, Article 100577. https://doi.org/10.1016/j.chbr.2024.100577

Capel, T., & Brereton, M. (2023). What is human-centered about human-centered AI? A map of the research landscape. In A. Schmidt, K. Väänänen, T. Goyal, P. O. Kristensson, A. Peters, S. Mueller, J. R. Williamson, & M. L. Wilson (Eds.), *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–23). ACM. https://doi.org/10.1145/3544548.3580959.

Cech, F. (2021). Tackling algorithmic transparency in communal energy accounting through participatory design. In F. Cech, & S. Farnham (Eds.), *C&T '21: Proceedings of the 10th international conference on communities & technologies - Wicked problems in the age of tech* (pp. 258–268). ACM. https://doi.org/10.1145/3461564.3461577.

Chien, S.-Y., Wang, Y.-F., Cheng, K.-T., & Chen, Y.-C. (2024). Comparative study of XAI perception between eastern and Western cultures. *International Journal of Human-Computer Interaction*, 1–17. https://doi.org/10.1080/10447318.2024.2441015

Clinciu, M.-A., & Hastie, H. (2019). A survey of explainable AI terminology. In J. M. Alonso, & A. Catala (Eds.), *Proceedings of the 1st workshop on interactive natural language technology for explainable artificial intelligence (NL4XAI 2019)* (pp. 8–13). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-8403.

Colin, C., & Martin, A. (2023). The user experience of low-techs: From user problems to design principles. *Journal of User Experience, 18*, 68–85. https://doi.org/10.5555/3604890.3604892

Corbett, E., & Dove, G. (2024). Signs of the smart city: Exploring the limits and opportunities of transparency. In F. F. Mueller, P. Kyburz, J. R. Williamson, C. Sas, M. L. Wilson, P. T. Dugas, & I. Shklovski (Eds.), *Proceedings of the CHI conference on human factors in computing systems* (pp. 1–14). ACM. https://doi.org/10.1145/3613904.3641931.

Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction, 18*(5), 455–496. https://doi.org/10.1007/s11257-008-9051-3

de Zoeten, M., Ernst, C.-P., & Rothlauf, F. (2024). The effect of explainable AI on AI-Trust and its antecedents over the course of an interaction. In M. Avital, E. Karahanna, & M. Themistocleous (Eds.), *ECIS 2024 proceedings* (16 pages). AIS. https://aisel.aisnet.org/ecis2024/track03_ai/track03_ai/10.

Deters, H., Droste, J., Obaidi, M., & Schneider, K. (2025). Exploring the means to measure explainability: Metrics, heuristics and questionnaires. *Information and Software Technology, 181*, Article 107682. https://doi.org/10.1016/j.infsof.2025.107682

Deters, H., Droste, J., & Schneider, K. (2023). A means to what end? Evaluating the explainability of software systems using goal-oriented heuristics. In M. A. Babar, R. Chitchyan, J. Li, & H. Zhang (Eds.), *Proceedings of the 27th international conference on evaluation and assessment in software engineering* (pp. 329–338). ACM. https://doi.org/10.1145/3593434.3593444.

Diefenbach, S., Christoforakos, L., Ullrich, D., & Butz, A. (2022). Invisible but understandable: In search of the sweet spot between technology invisibility and transparency in smart spaces and beyond. *Multimodal Technologies and Interaction, 6*(10), Article 95. https://doi.org/10.3390/mti6100095

DIN Deutsches Institut für Normung e.V.. (2023). *Artificial intelligence – Life cycle processes and quality requirements – Part 3: Explainability (DIN SPEC 92001-3)*. Beuth Verlag GmbH. https://www.dinmedia.de/de/technische-regel/din-spec-92001-3/369799101.

Endsley, M. R. (2017). From here to autonomy. *Human Factors, 59*(1), 5–27. https://doi.org/10.1177/0018720816681350

European Commission. (2019). *Ethics guidelines for trustworthy AI. EU Publications Office.* Advance online publication. https://doi.org/10.2759/346720

European Commission. (2025). *AI act.* https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

Eyert, F., & Lopez, P. (2023). Rethinking transparency as a communicative constellation. In S. Fox, C. Harrington, A. Z. Huq, & C. Tan (Eds.), *2023 ACM conference on fairness, accountability, and transparency* (pp. 444–454). ACM. https://doi.org/10.1145/3593013.3594010.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics, 26*(6), 3333–3361. https://doi.org/10.1007/s11948-020-00276-4

Fox, S., & Rey, V. F. (2024). A cognitive load theory (CLT) analysis of machine learning explainability, transparency, interpretability, and shared interpretability. *Machine Learning and Knowledge Extraction, 6*(3), 1494–1509. https://doi.org/10.3390/make6030071

Gaborov, M., & Ivetić, D. (2022). The importance of integrating thinking design, user experience and agile methodologies to increase profitability. *Journal of Applied Technical and Educational Sciences, 12*, Article 286. https://doi.org/10.24368/jates286

Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision making and a "right to explanation". *AI Magazine, 38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Gumber, S. (2023). Minimalism in design: A trend of simplicity in complexity. *ShodhKosh: Journal of Visual and Performing Arts, 4*(2), 357–365. https://doi.org/10.29121/shodhkosh.v4.i2.2023.539

Gunning, D., Vorm, E., Wang, J. Y., & Turek, M. (2021). DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters, 2*(4), 1–11. https://doi.org/10.1002/ail2.61

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus, & J. Ziegler (Eds.), *Berichte des German Chapter of the ACM. Mensch & Computer 2003* (Vol. 57, pp. 187–196). Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-80058-9_19.

Heerink, M., Krose, B., Evers, V., & Wielinga, B. (2009). Measuring acceptance of an assistive social robot: A suggested toolkit. In *RO-MAN 2009 - The 18th IEEE international symposium on robot and human interactive communication* (pp. 528–533). IEEE. https://doi.org/10.1109/ROMAN.2009.5326320.

Hein, I., & Diefenbach, S. (2025). Towards a comprehensive view on technology transparency: A cross-technology investigation of psychological and design factors around users' transparency need and perception. *International Journal of Human-Computer Interaction, 41*(24), 15862–15879. https://doi.org/10.1080/10447318.2025.2502983

Hein, I., Diefenbach, S., & Ullrich, D. (2023). Designing for technology transparency – Transparency cues and user experience. *Usability Professionals, 23*, 1–5. https://doi.org/10.18420/muc2023-up-448

Hein, I., Ullrich, D., & Diefenbach, S. (2024). Creating subtle transparency in smart spaces: Visual cues to communicate attention and intention of assistive systems. In Å. Cajander, & M. Wiberg (Eds.), *Adjunct proceedings of the 2024 nordic human-computer interaction conference* (pp. 1–6). ACM. https://doi.org/10.1145/3677045.3685435.

Hoffman, R. R., Miller, T., Klein, G., Mueller, S. T., & Clancey, W. J. (2023). Increasing the value of XAI for users: A psychological perspective. *KI - Künstliche Intelligenz, 37*, 237–247. https://doi.org/10.1007/s13218-023-00806-9

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *ArXiv*, 1–50. https://doi.org/10.48550/arXiv.1812.04608

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers of Computer Science, 5*, Article 1096257. https://doi.org/10.3389/fcomp.2023.1096257

Holzinger, A., Carrington, A., & Müller, H. (2020). Measuring the quality of explanations: The system causability scale (SCS): Comparing human and machine explanations. *KI - Künstliche Intelligenz, 34*(2), 193–198. https://doi.org/10.1007/s13218-020-00636-z

Hu, P., Zeng, Y., Wang, D., & Teng, H. (2024). Too much light blinds: The transparency-resistance paradox in algorithmic management. *Computers in Human Behavior, 161*, Article 108403. https://doi.org/10.1016/j.chb.2024.108403

Huff Jr, E. W., Day Grady, S., & Brinnkley, J. (2021). Tell me what I need to know: Consumers' desire for information transparency in self-driving vehicles. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 65*(1), 327–331. https://doi.org/10.1177/1071181321651240

Hughes, J. E. (2013). Descriptive indicators of future teachers' technology integration in the PK-12 classroom: Trends from a laptop-infused teacher education program. *Journal of Educational Computing Research, 48*(4), 491–516. https://doi.org/10.2190/EC.48.4.e

IBM. (2024). *IBM global AI adoption index 2023.* https://www.multivu.com/players/English/9240059-ibm-2023-global-ai-adoption-index-report/.

Kang, S., Potinteu, A.-E., & Said, N. (2025). ExplainitAI: When do we trust artificial intelligence? The influence of content and explainability in a cross-cultural comparison. In N. Yamashita, V. Evers, K. Yatani, & X. Ding (Eds.), *Proceedings of the extended abstracts of the CHI conference on human factors in computing systems* (pp. 1–7). ACM. https://doi.org/10.1145/3706599.3720222.

Kaushal, R., van de Kerkhof, J., Goanta, C., Spanakis, G., & Iamnitchi, A. (2024). Automated transparency: A legal and empirical analysis of the digital services act transparency database. In R. Binns, F. Calmon, A. Olteanu, & M. Veale (Eds.), *The 2024 ACM conference on fairness, accountability, and transparency* (pp. 1121–1132). ACM. https://doi.org/10.1145/3630106.3658960.

Kelley, J. F. (2018). Wizard of Oz (WoZ) - A yellow brick journey. *Journal of Usability Studies, 13*(3), 119–124. http://uxpajournal.org/wp-content/uploads/sites/7/pdf/jus_kelley_may2018.pdf.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kohl, M. A., Baum, K., Langer, M., Oster, D., Speith, T., & Bohlender, D. (2019). Explainability as a non-functional requirement. In S.-W. Lee (Ed.), *2019 IEEE 27th international requirements engineering conference (RE)* (pp. 363–368). IEEE. https://doi.org/10.1109/RE.2019.00046.

Konopka, B., Hönemann, K., & Wiesche, M. (2024). Augmented reality, trust and technology malfunctions: An initial theory and empirical study. In S. Zaza, C. Riemenschneider, I. R. Guzman, M. Bantan, T. Monroe-White, R. Sundrup, N. Brooks, J. Williams, & D. Lee (Eds.), *Proceedings of the 2024 computers and people research conference* (pp. 1–8). ACM. https://doi.org/10.1145/3632634.3655884.

Koverola, M., Kunnari, A., Sundvall, J., & Laakasuo, M. (2022). General attitudes towards robots scale (GAToRS): A new instrument for social surveys. *International Journal of Social Robotics, 14*(7), 1559–1581. https://doi.org/10.1007/s12369-022-00880-3

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice and using software.* SAGE Publications.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In A. Holzinger (Ed.), *HCI and usability for education and work* (pp. 63–76). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-89350-9_6.

Lenzner, T., & Menold, N. (2016). *Question Wording. GESIS - Leibniz-Institut für Sozialwissenschaften.* GESIS Survey Guidelines. https://www.gesis.org/fileadmin/admin/Dateikatalog/pdf/guidelines/question_wording_lenzner_menold_2016.pdf.

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. In R. Bernhaupt, F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. Zhao, B. P. Samson, & R. Kocielnik (Eds.), *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–15). ACM. https://doi.org/10.1145/3313831.3376590.

Liao, Q. V., & Varshney, K. R. (2022). Human-centered explainable AI (XAI): From algorithms to user experiences. *ArXiv, 1–18*. https://doi.org/10.48550/arXiv.2110.10790

Liebherr, M., Gößwein, E., Kannen, C., Babiker, A., Al-Shakhsi, S., Staab, V., Li, B. J., Ali, R., & Montag, C. (2025). Working memory and the need for explainable AI - Scenarios from healthcare, social media and insurance. *Heliyon, 11*(2), Article e41871. https://doi.org/10.1016/j.heliyon.2025.e41871

Lim, B. Y., & Dey, A. K. (2009). Assessing demand for intelligibility in context-aware applications. In S. Helal, H. Gellersen, & S. Consolvo (Eds.), *Proceedings of the 11th international conference on ubiquitous computing* (pp. 195–204). ACM. https://doi.org/10.1145/1620545.1620576.

Lopes, P., Silva, E., Braga, C., Oliveira, T., & Rosado, L. (2022). XAI systems evaluation: A review of human and computer-centred methods. *Applied Sciences, 12*(19), Article 9423. https://doi.org/10.3390/app12199423

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In *Proceedings of the 11th australasian conference on information systems* (Vol. 53, pp. 6–8). https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b8eda9593fbcb63b7ced1866853d9622737533a2.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems, 11*(3-4), 1–45. https://doi.org/10.1145/3387166

Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion.* Springer. https://doi.org/10.1007/978-3-662-61532-4

Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. *ArXiv*, 1–204. https://doi.org/10.48550/arXiv.1902.01876

Narayanan, D. (2023). Welfarist moral grounding for transparent AI. In S. Fox, C. Harrington, A. Z. Huq, & C. Tan (Eds.), *2023 ACM conference on fairness, accountability, and transparency* (pp. 64–76). ACM. https://doi.org/10.1145/3593013.3593977.

Naumann, A., & Hurtienne, J. (2010). Benchmarks for intuitive interaction with Mobile devices. In M. de Sá, L. Carriço, & N. Correia (Eds.), *Proceedings of the 12th international conference on human computer interaction with mobile devices and services* (pp. 401–402). ACM. https://doi.org/10.1145/1851600.1851685.

Naveed, S., Stevens, G., & Robin-Kern, D. (2024). An overview of the empirical evaluation of explainable AI (XAI): A comprehensive guideline for user-centered evaluation in XAI. *Applied Sciences, 14*(23), Article 11288. https://doi.org/10.3390/app142311288

Nomura, T. (2014). Influences of experiences of robots into negative attitudes toward robots. In P. A. Vargas, R. Aylett, & F. Amirabdollahian (Eds.), *The 23rd IEEE international symposium on robot and human interactive communication* (pp. 460–464). IEEE. https://doi.org/10.1109/ROMAN.2014.6926295.

Norval, C., Cloete, R., & Singh, J. (2023). Navigating the audit landscape: A framework for developing transparent and auditable XR. In S. Fox, C. Harrington, A. Z. Huq, & C. Tan (Eds.), *2023 ACM conference on fairness, accountability, and transparency* (pp. 1418–1431). ACM. https://doi.org/10.1145/3593013.3594090.

Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction, 27*(3-5), 393–444. https://doi.org/10.1007/s11257-017-9195-0

Ooge, J., Kato, S., & Verbert, K. (2022). Explaining recommendations in E-Learning: Effects on adolescents' trust. In G. Jacucci, S. Kaski, C. Conati, S. Stumpf, T. Ruotsalo, & K. Gajos (Eds.), *27th international conference on intelligent user interfaces* (pp. 93–105). ACM. https://doi.org/10.1145/3490099.3511140.

Papenmeier, A., Kern, D., Englebienne, G., & Seifert, C. (2022). It's complicated: The relationship between user trust, model accuracy and explanations in AI. *ACM Transactions on Computer-Human Interaction, 29*(4), 1–33. https://doi.org/10.1145/3495013

Pataki, K., Sachse, K., Prümper, J., & Thüring, M. (2006). ISONORM 9241/110-Short: Kurzfragebogen zur Software-Evaluation. In F. Lösel (Ed.), *Berichte über den 45. Kongress der Deutschen Gesellschaft für Psychologie* (pp. 258–259). Pabst Science Publishers.

Peters, U., & Carman, M. (2024). Cultural bias in explainable AI research: A systematic analysis. *Journal of Artificial Intelligence Research, 79*, 971–1000. https://doi.org/10.1613/jair.1.14888

Procter, R., Tolmie, P., & Rouncefield, M. (2023). Holding AI to account: Challenges for the delivery of trustworthy AI in healthcare. *ACM Transactions on Computer-Human Interaction, 30*(2), 1–34. https://doi.org/10.1145/3577009

Riveiro, M., & Thill, S. (2022). The challenges of providing explanations of AI systems when they do not behave like users expect. In A. Bellogin, L. Boratto, O. C. Santos, L. Ardissono, & B. Knijnenburg (Eds.), *Proceedings of the 30th ACM conference on user modeling, adaptation and personalization* (pp. 110–120). ACM. https://doi.org/10.1145/3503252.3531306.

Roesler, E., Rieger, T., & Manzey, D. (2022). Trust towards human vs. automated agents: Using a multidimensional trust questionnaire to assess the role of performance, utility, purpose, and transparency. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 66*(1), 2047–2051. https://doi.org/10.1177/1071181322661065

Roets, A., & van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the need for closure scale. *Personality and Individual Differences, 50*(1), 90–94. https://doi.org/10.1016/j.paid.2010.09.004

Schecter, A., Hohenstein, J., Larson, L., Harris, A., Hou, T.-Y., Lee, W.-Y., Lauharatanahirun, N., DeChurch, L., Contractor, N., & Jung, M. (2023). Vero: An accessible method for studying human–AI teamwork. *Computers in Human Behavior, 141*, Article 107606. https://doi.org/10.1016/j.chb.2022.107606

Schmude, T., Koesten, L., Möller, T., & Tschiatschek, S. (2025). Information that matters: Exploring information needs of people affected by algorithmic decisions. *International Journal of Human-Computer Studies, 193*, Article 103380. https://doi.org/10.1016/j.ijhcs.2024.103380

Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A., & van den Bosch, K. (2021). Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies, 154*, Article 102684. https://doi.org/10.1016/j.ijhcs.2021.102684

Schott, K., Papenmeier, A., Hienert, D., & Kern, D. (2024). What did I say again? Relating user needs to search outcomes in conversational commerce. In A. Maedche, M. Beigl, K. Gerling, & S. Mayer (Eds.), *Proceedings of Mensch und Computer 2024* (pp. 129–139). ACM. https://doi.org/10.1145/3670653.3670680.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies, 146*, 10. https://doi.org/10.1016/j.ijhcs.2020.102551, 102551.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences? *International Journal of Information Management, 52*, Article 102061. https://doi.org/10.1016/j.ijinfomgt.2019.102061

Silva, A., Schrum, M., Hedlund-Botti, E., Gopalan, N., & Gombolay, M. (2022). Explainable artificial intelligence: Evaluating the objective and subjective impacts of xAI on human-agent interaction. *International Journal of Human-Computer Interaction, 39*(7), 1390–1404. https://doi.org/10.1080/10447318.2022.2101698

Sokol, K., & Flach, P. (2020). Explainability fact sheets. In M. Hildebrandt, C. Castillo, E. Celis, S. Ruggieri, L. Taylor, & G. Zanfir-Fortuna (Eds.), *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 56–67). ACM. https://doi.org/10.1145/3351095.3372870.

Speith, T., & Langer, M. (2023). A new perspective on evaluation methods for explainable artificial intelligence (XAI). In K. Schneider, F. Dalpiaz, & J. Horkoff (Eds.), *2023 IEEE 31st international requirements engineering conference workshops (REW)* (pp. 325–331). IEEE. https://doi.org/10.1109/REW57809.2023.00061.

Springer, A., & Whittaker, S. (2020). Progressive disclosure: When, why, and how do users want algorithmic transparency information? *ACM Transactions on Interactive Intelligent Systems, 10*(4), 1–32. https://doi.org/10.1145/3374218

Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in likert scales: A questionable practice. *Psicothema, 30*(2), 149–158. https://doi.org/10.7334/psicothema2018.33

Suh, A., Hurley, I., Smith, N., & Siu, H. C. (2025). Fewer than 1% of explainable AI papers validate explainability with humans. In N. Yamashita, V. Evers, K. Yatani, & X. Ding (Eds.), *Proceedings of the extended abstracts of the CHI conference on human factors in computing systems* (pp. 1–7). ACM. https://doi.org/10.1145/3706599.3719964.

Turri, V., Morrison, K., Robinson, K.-M., Abidi, C., Perer, A., Forlizzi, J., & Dzombak, R. (2024). Transparency in the wild: Navigating transparency in a deployed AI system to broaden need-finding approaches. In R. Binns, F. Calmon, A. Olteanu, & M. Veale (Eds.), *The 2024 ACM conference on fairness, accountability, and transparency* (pp. 1494–1514). ACM. https://doi.org/10.1145/3630106.3658985.

Ullrich, D., & Diefenbach, S. (2010). INTUI. Exploring the facets of intuitive interaction. In J. Ziegler, & A. Schmidt (Eds.), *Mensch & computer 2010* (pp. 251–260). Oldenbourgh Wissenschaftsverlag. https://doi.org/10.1524/9783486853483.251.

Usmani, U. A., Happonen, A., & Watada, J. (2023). Human-centered artificial intelligence: Designing for user empowerment and ethical considerations. In T. Özseven, E. Yaşar, & S. Yevseiev (Eds.), *2023 5th international congress on human-computer interaction, optimization and robotic applications (HORA)* (pp. 1–7). IEEE. https://doi.org/10.1109/HORA58378.2023.10156761.

van Berlo, Z. M., & Breves, P. L. (2025). Disclosing the virtual nature of virtual influencers: The effect of disclosure prominence and the role of product digitality. *Computers in Human Behavior Reports, 19*, Article 100742. https://doi.org/10.1016/j.chbr.2025.100742

Vasconcelos, H., Bansal, G., Fourney, A., Liao, Q. V., & Vaughan, J. W. (2024). Generation probabilities are not enough: Uncertainty highlighting in AI code completions. *ACM Transactions on Computer-Human Interaction, 32*(1), Article 3702320. https://doi.org/10.1145/3702320

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion, 76*, 89–106. https://doi.org/10.1016/j.inffus.2021.05.009

Vorm, E. S., & Combs, D. J. Y. (2022). Integrating transparency, trust, and acceptance: The intelligent systems technology acceptance model (ISTAM). *International Journal of Human-Computer Interaction, 38*(18-20), 1828–1845. https://doi.org/10.1080/10447318.2022.2070107

Vorm, E. S., & Miller, A. D. (2020). Modeling user information needs to enable successful human-machine teams: Designing transparency for autonomous systems. In D. D. Schmorrow, & C. M. Fidopiastis (Eds.), *Augmented cognition. Human cognition and behavior, 12197* pp. 445–465). Springer International Publishing. https://doi.org/10.1007/978-3-030-50439-7_31.

Wang, R., Bush-Evans, R., Arden-Close, E., Bolat, E., McAlaney, J., Hodge, S., Thomas, S., & Phalp, K. (2023). Transparency in persuasive technology, immersive technology, and online marketing: Facilitating users' informed decision making and practical implications. *Computers in Human Behavior, 139*, 15. https://doi.org/10.1016/j.chb.2022.107545, 107545.

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–15). ACM. https://doi.org/10.1145/3290605.3300831.

Wang, X., & Yin, M. (2021). Are explanations helpful? A comparative study of the effects of explanations in AI-Assisted decision-making. In T. Hammond, K. Verbert, D. Parra, B. Knijnenburg, J. O'Donovan, & P. Teale (Eds.), *26th international conference on intelligent user interfaces* (pp. 318–328). ACM. https://doi.org/10.1145/3397481.3450650.

Wanner, J., Herm, L.-V., Heinrich, K., & Janiesch, C. (2022). The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study. *Electronic Markets, 32*(4), 2079–2102. https://doi.org/10.1007/s12525-022-00593-5

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weber, T., Hußmann, H., & Eiband, M. (2021). Quantifying the demand for explainability. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, & K. Inkpen (Eds.), *Human-computer interaction – interact 2021* (pp. 652–661). Springer International Publishing. https://doi.org/10.1007/978-3-030-85616-8_38.

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research, 49*(5), 737–747. https://doi.org/10.1509/jmr.11.0368

Weitz, K., Zellner, A., & André, E. (2022). What do end-users really want? Investigation of human-centered XAI for mobile health apps. *ArXiv*, 1–19. https://doi.org/10.48550/arXiv.2210.03506

Werz, J. M., Borowski, E., & Isenhardt, I. (2024). Explainability as a means for transparency? Lay users' requirements towards transparent AI. In W. Karwowski, & T. Ahram (Eds.), *15th international conference on applied human factors and ergonomics*

(AHFE 2024) (pp. 1–11). AHFE International. https://doi.org/10.54941/ahfe1004712.

Wessel, D., Attig, C., & Franke, T. (2019). ATI-S - An ultra-short scale for assessing affinity for technology interaction in user studies. In F. Alt, A. Bulling, & T. Döring (Eds.), *Proceedings of Mensch und Computer 2019* (pp. 147–154). https://doi.org/10.1145/3340764.3340766. ACM.

Wijekoon, A., Wiratunga, N., Corsar, D., Martin, K., Nkisi-Orji, I., Díaz-Agudo, B., & Bridge, D. (2024). XEQ scale for evaluating XAI experience quality. *ArXiv*, 1–25. https://doi.org/10.48550/arXiv.2407.10662

You, Y., Tsai, C.-H., Li, Y., Ma, F., Heron, C., & Gui, X. (2023). Beyond self-diagnosis: How a chatbot-based symptom checker should respond. *ACM Transactions on Computer-Human Interaction, 30*(4), 1–44. https://doi.org/10.1145/3589959

Zednik, C. (2021). Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology, 34*(2), 265–288. https://doi.org/10.1007/s13347-019-00382-7