Matters of Explanation: Rethinking Explainability with Tangible, Embodied, Material Interactions

Goda Klumbytė, Claude Draude
Participatory IT Design, University of Kassel
Pfannkuchstr. 1
34121 Kassel, Germany
goda.klumbyte@uni-kassel.de, claude.draude@uni-kassel.de

Abstract

This paper explores how the notions of explainability and understanding in explainable AI (XAI) and its design can be re-thought with embodied, tangible and material interactions and related theories of new materialism. Contributing to the emerging 'tangible XAI' and 'graspable AI' frameworks, we suggest that XAI design can and should leverage material, tangible qualities, and embodiment to address the emergent and process-oriented approach to understanding and explainability in XAI design. We present two examples that help illustrate this approach: an artwork, created as part of collaborative research that engages affective and felt modalities of AI, and a paper prototype of a pinball machine built during a workshop that explored tangible explainability for large language models. Through these examples, we also show that materiality and embodiment are important agents in the emergence of explainability and understanding and invite to consider ethical explanation design as a material practice of care.

Introduction

Explainability is an important domain within AI design and development. Together with transparency, it is also one of the foundations for trustworthy AI: as the logic goes, if a system can be explained and understood, it is also more likely to be trusted and deemed trustworthy. As a field within AI, explainability is often defined in technical terms, and explainability methods, such as classical SHAP or LIME (Arunika et al. 2024; Holzinger et al. 2022), often target AI systems experts and engineers. The emerging human-centered explainable AI (HCXAI) bridges classical technical approaches to explainability with human-computer interaction and investigates modes of explanation design that are human-focused and also include lay users as the target audience of explanations (Ehsan and Riedl 2020). Methods here include elicitation of mental models and design towards more accurate mental model acquisition (Merry, Riddle, and Warren 2021), as well as leveraging interaction design techniques towards constructing XAI solutions (Ooge and Verbert 2022; Raees et al. 2024). Explanation design cuts across both HCXAI and traditional approaches to XAI and ranges from visualizations to interactive dialogue formats. The overarching goal of explanations is to generate an understanding of a specific AI system, be it for experts or lay users.

Explainability and understandability as fields seem to share a few assumptions, namely: (1) that systems perform objective functions driven by internal logic, which can or at least can be attempted to be explained; (2) that explainability is a property of a system and that systems can therefore be more or less explainable; (3) that explanations can lead to understanding, which is in a sense a property of the human user - a certain measurable characteristic (more/less, better/-worse understanding) that can be acquired; and (4) that both explanation and understanding are fundamentally functions of reason-based cognition. In this paper, we seek to question these assumptions by asking: How might notions of explainability and understandability - and correspondingly XAI design - shift if they are approached through the perspectives of material, tangible, and embodied interactions?

As cognitive science has shown, cognition does not merely take place in the mind but is enactive, extended, embodied, and embedded (the so-called 4E paradigm of cognition - Newen, de Bruin, and Gallagher 2018). Furthermore, understanding is not a simple static state or amount of information or knowledge that a subject possesses but can be seen as a *process* that is in constant development in relation to context (e.g. Blaha et al. 2022). In this paper, we will suggest that this materiality- and process-oriented perspective - let us call it materialist-embodied perspective for the sake of brevity - can be leveraged towards more expansive formulations of explainability, understandability, as well as more embodied and tangible XAI design.

The paper proceeds as follows: first, we introduce XAI and underlying notions of explainability and understandability; then, we will present materialist-embodied conceptual framework that relies on tangible, embodied, material interaction perspectives and new materialist philosophies. We will then report preliminary findings from two projects: a collaborative research project on feeling AI and a workshop on tangible explanations of large language models. We will end the paper with a discussion on possible implications for XAI and explanation design.

Explainability and Understandability in XAI

Engaging with the domain of explainable artificial intelligence (XAI) involves encountering a range of overlapping

and often ambiguously defined concepts. In academic literature, concepts such as explainability, interpretability, understandability, transparency, comprehensibility, and intelligibility are frequently used, but their definitions vary. These notions all deal with the complex socio-technical relationship between human users and algorithmic models, but they differ in terms of who or what initiates the process of explanation, how active or passive the human or model is in this exchange, and what is meant by "understanding" in context. This conceptual ambiguity can lead to different perspectives of how explainability and understandability of AI is to be addressed and operationalized in ethical guidelines and tools. An example of this ambiguity would be the frequent interchangeable use of the terms 'explainability' and 'interpretability', particularly in broader (i.e., not exclusively technical) explainability discourse. Additionally, the same concept can be defined in different ways. For instance, Preece et al. (2018) view interpretability as human-focused, emphasizing the person's role in deciphering the model, whereas Kaur et al. (2022) describe interpretability as stemming from the model's ability to present its logic in a human-understandable form. In Kaur et al.'s framing, explainability then becomes more about the human actively making sense of the system's outputs, thus highlighting a user-centered interpretation.

In our work, we rely on Arrieta et al. (2020), who analyze nearly 400 scholarly contributions to provide a taxonomy that seeks to clarify the conceptual underpinnings of XAI, and Eshan et al (2021) who explore human-centered XAI. As articulated by the latter, explainability refers to the human ability to produce or derive an explanation — either before (ante-hoc) or after (post-hoc) a model's decision — in order to interpret its behavior in a way that aligns with human reasoning. Here, the human plays an active role in generating a narrative or rationale to understand what the model has done. By contrast, interpretability centers on the model's inherent capacity to present its internal processes in a way that is understandable to people. This shifts the focus to the system's design: how well it can be examined and decoded by human observers. Interpretability is thus considered a property of the model that facilitates insight and comprehension.

Understandability can be understood as the broader degree to which a model's operations can be mentally grasped by users and to what extent the model's behavior makes sense to a human observer. As Arrieta et al. discuss, 'understandability measures the degree to which a human can understand a decision made by a model. [...] understandability is a two-sided matter: model understandability and human understandability' (2020). Others, however, define understandability (and intelligibility) as a model-characteristic that defines how much the model allows the user to grasp its function (how the model works) without explaining its internal structure or data processing operations (Samek et al. 2019). Understandability is related to the concept of transparency, which represents an aspirational ideal, especially in the context of complex systems like deep learning models. Transparency refers to the quality of a model being understandable without requiring external aids or explanations (Arrieta et al. 2020). It implies that the system's logic and processes are directly accessible and legible to human observers. This notion can be broken down into dimensions such as simulatability (the ability to mentally simulate the model), decomposability (clarity of individual components), and algorithmic transparency (clarity in the functioning of learning procedures). These dimensions reflect varying levels of interpretability, depending on how much of the model's structure and behavior can be scrutinized and understood.

The three concepts of explainability-understandability-transparency form a nexus when placed in ethical discourses: as policies (such as EU Trustworthy AI agenda) show, it is widely understood and accepted that the way towards ethically responsible and trustworthy AI use is grounded in the idea that systems can gain trustworthiness when they are properly explained and understood, which in turn rests on the general (i.e. non-technical) idea of transparency as the quality of being explainable/understandable, directly or indirectly. Put simply: explainability - and transparency - are directly related to ethics in AI (Maclure 2021).

As we approach XAI from the perspective of human-centredness and sociotechnical design, we are particularly interested in the *effects* and *human-centered* methods of explanation design. That means that we see explainability not as a matter of technical methods alone but a question of generating explanations of systems' structure, rationale, and operations that potentially lead to an understanding of how said system works - i.e., understanding is the goal of explanation. Towards that end, human-centered explanation design methods have been developed. These include, for instance, mental model approach (Schulz 2023; Merry, Riddle, and Warren 2021) to probe and influence user's mental imaginaries of a specific system and its functions; interaction-based methods (Bertrand et al. 2023), such as interactive visualisations (Ooge and Verbert 2022), dialogues (Mindlin et al. 2025), and other interaction modalities; narrative-based explanation design (Hartmann et al. 2022); and frameworks such as End-User-Centered Explainable AI (EUCA) that summarizes some user-friendly explanator forms and links them to existing technical explainability methods (Jin et al. 2021) and conceptual frameworks such as approaching explaination as a social practice (Rohlfing et al. 2021).

While both technical and human-centred approaches to explainable AI have different methods and foci, what they seem to share is an approach to explainability as a communication question. In her work on algorithms as communication partners and explainability as communicative act (Esposito 2023), Esposito proposes that 'Explanations as communicative processes do not imply any disclosure of thoughts or neural processes, but only reformulations that provide the partners with additional elements and enable them to understand (from their perspective) what has been done and why'. In a similar vein, Liao et al. also propose that communication perspective - how the model communicates its operations through specific transparency and interaction features, is important for trustworthy AI (Liao and Sundar 2022). As we mentioned in the introduction, there are some intuitive assumptions that ideas and implementations of ex-

plainability and understandability often rest on, namely: (1) that systems perform objective functions driven by internal logic, which can or at least can be attempted to be explained; (2) that explainability is a property of a system and that systems can therefore be more or less explainable; (3) that explanations can lead to understanding, which is in a sense a property of the human user - a certain measurable characteristic (more/less, better/worse understanding) that can be acquired; and (4) that both explanation and understanding are fundamentally functions of reason-based cognition. Without denying the usefulness of these assumptions, we, however, would like to open the question: how might explainability and understandability be approached and rendered from tangible, embodied, material interaction perspective?

Embodiment, Materiality, Tangibility

Tangible, embodied, and material interaction is a field in computing that explores how users engage with computational systems through physical and sensory-rich interfaces. It sits at the intersection of Human-Computer Interaction (HCI), design, and cognitive science, and emphasizes the role of the body, the materiality of objects, and spatial context in shaping user experience and cognition (Dourish 2004; Ishii and Ullmer 1997; Wiberg 2018; Hornecker and Buur 2006).

Briefly put, tangible interaction leverages physical objects and environments as interfaces and often works with physical manipulation of various objects (not only keyboards, mice and touch-screens) for computational feedback, focusing significantly on haptic and other forms of sensory feedback as well as spatial interaction. Embodied interaction emphasizes the role of the body in cognitive processes and therefore its centrality to interaction, including physical and social contexts of embodiment. Material interaction highlights the agency of various materials that shape and participate in interaction and develops design strategies with and through materials. Furthermore, emerging perspectives such as soma design (Höök 2018) straddles all three of these fields with the focus on somatic and somaesthetic qualities in interaction and design processes.

Much of this field rests on the 4E perspective on cognition, i.e., the understanding of cognition as an embodied, embedded, extended, and enactive process, and thus positions cognition as an action-oriented process that is situated in physical and sociocultural environment (Newen, de Bruin, and Gallagher 2018). This has implications also for how meaning and knowing is conceptualized: meaning, according to the 4E paradigm, emerges in situated and embodied interaction, whereas knowing is not something that is contained in the brain but cognitively extends throughout and emerges within the environment and thus also various devices in it (Varela, Thompson, and Rosch 2016; Clark and Chalmers 1998; Menary 2010).

Scholars such as Jaegher and Di Paolo (2007) also emphasize that sense-making, which is similar yet not synonymous to understanding, is a participatory process, i.e., it is a process that is enacted and emerges collectively through social interactions. While these scholars specifically theorized human interactions, others have explored participatory sense making in technology design (van Dijk 2024; Smit et al. 2022; Davis et al. 2016; Deshpande and Magerko 2024) as well as a paradigm for re-envisioning explainability and interpretability (Kaur et al. 2022).

In the field of explainability, understandability and transparency of AI, tangible, embodied, material interaction perspective is slowly gaining traction. Colley et al. (2022) have introduced a tangible XAI (tangXAI) framework that bridges technical explainability methods with data physicalization and tangible user interface methods, arguing that 'Tangibility can provide a new dimension to embodied and multisensory human-AI interaction, provide means to adapt to user expectations and behaviour, as well as to contexts of use'. They report that indeed tangible interaction can be helpful in users' reflections of AI explanations but it is important to gear tangible XAI design towards not simply tangible interaction but specifically also explanation of AI (Kaisa Väänänen, Ashley Colley, and Jonna Häkkilä 2024).

Ghajargar et al. have also explored tangible XAI and introduce the idea of 'graspable AI' instead of explainable AI (Ghajargar et al. 2021, 2022; Ghajargar and Bardzell 2022), which 'conveys the meaning of being understandable intellectually, meaningfully and physically'. In their empirical work, Ghajargar and Bardzell utilized product semantics and aesthetics in design theory to explore how forms (both physical and abstract) become communicative of function and could be utilized towards graspable AI design (Ghajargar et al. 2022). They highlight that designing tangible explanations is a complex task and requires technological as well as ontological and aesthetic considerations.

Building on this background, we seek to contribute to the tangible, embodied, and material approaches to explainable AI. Next to existing focus on materiality and embodiment, we offer to add another theoretical perspective to the conceptual vocabulary around tangible and graspable AI that is rooted in new materialist philosophies.

New materialism approaches matter as active and agentive, and draws attention to how matter and meaning mutually co-define each other (Coole and Frost 2010; Fox and Alldred 2015). New materialists suggest that matter is not static but unfolds as a process of *mattering* that is intermingled with processes of meaning-making (Fox and Alldred 2015). In this light then, meaning making or sense making itself can be understood as a process that emerges at the interaction between various materialities that include human bodies and non-human technologies, as well as sociocultural contextual factors. Such a perspective, when applied to explainability and understandability in AI, helps draw attention to the situation and process of interaction as a place where explanation and understanding 'happen'.

Together with tangible and embodied interaction, this focus on materiality as agentive and process-oriented constitutes what we called in the introduction a 'materialist-embodied' perspective towards XAI. We will illustrate this perspective and some of the re-orientations that it can produce in XAI in the next section with two examples from our ongoing research.

Felt and Tangible: Two Examples of Materialist-Embodied Approach to XAI

In this section we describe two examples from our research work that addresses material, embodied, and tangible interaction qualities in relation to explainability and understandability of AI models. First example presents a collaborative research project on 'feeling AI', conducted by Goda Klumbytė with collaborators Mika Satomi and Daniela K. Rosner - an artist and a design scholar, respectively - and specifically the artistic work that Mika Satomi created in response to and as part of this collaboration. Second example briefly describes the process and result of a group work conducted during a workshop on tangible large language model (LLM) explanations that Goda Klumbytė and Claude Draude were part of and co-organised (Angelini et al. 2025). Both of these examples are not full-fledged case studies but rather vignettes that, we suggest, help think about and illustrate the emergent, material, and embodied quality of explainability and understanding. We present insights from an autoethnographic perspective as an invitation for further exploration.

Feeling AI: Developing Affective and Embodied Understanding

This project started off as a conversation between Goda Klumbytė, Mika Satomi and Daniela K. Rosner in 2022 and has since the beginning of 2024 been sustained by regular monthly meetings and a few longer workshops to investigate the question of what might it mean to *feel AI*. G. Klumbytė was particularly motivated to explore this question following works of scholars such as Eubanks (2018) and O'Neil (2017), which aptly point out that oftentimes AI systems operate in the background, without users and other affected people being explicitly aware of them - for instance, one's job application might be evaluated by an automated hiring model without one's knowledge. This lack of knowledge of when one might be subject to AI's evaluations leads to a further lack of understanding of the potential implications and effects of such systems not only societally but also individually. Nonetheless, these effects, while not necessarily consciously registered as the results of AI operations, are still acutely *felt* - through, for example, opening or closing of specific life chances, such as a job position. As Eubanks and O'Neil demonstrate, the negative effects, particularly of evaluative AI systems, disproportionately affect already marginalized populations. Simultaneously, feeling for and of AI models, as a kind of tacit modality of knowledge, especially of the nuances and subtle specificities of the system, is something that engineers as well as users develop through extensive engagement with AI systems (Fridland and Stichter 2021; Cha et al. 2023).

As a response to our collaborative ongoing work and discussions, Mika Satomi developed an artwork (figure 1) that pieced together used stuffed animals filled with conductive wool as a touch sensitive instrument. The instrument 'vocalized' touch by using a machine learning tool *Gesture Recognition Toolkit* (https://github.com/nickgillian/grt) to analyze and map the sensor data into parameters for speech synthesis with *Pink Trombone* (https://dood.al/pinktrombone/). The instrument would map the sounds to specific touch inputs (that registered not only position of the touch points but also pressure levels), whereas the new, unmapped inputs that were generated through touching different parts of the instrument were predicted and synthesized by the network. As one interacted with the artwork, it created human-like sounds of vowels, producing slightly eerie *aaaa*'s, *ooooo*'s, *eeee*'s, melting and blending into each other.

Figure 1: G. Klumbytė 'playing' the instrument-artwork created by Mika Satomi. Video material available at https://www.nerding.at/felt-ai/.

The interaction with the artwork entailed 'playing' it by positioning it against one's body and using one's arms to press, touch, squeeze and stroke the stuffed animal composite body. As one engages in this interaction, one soon starts noticing certain vocalization patterns in relation to one's movement. However, these patterns were sometimes hard to reproduce, since neither the location of sensors nor the map of inputs were explicitly marked. This invited an open and curious engagement. As G. Klumbytė was 'playing' the work, they could vary the strength of the sound as well as the kind of sound that was made, while also engaging in a tactile, deeply embodied and *affective* experience - the later element strengthened both by the human-like vocalizations, as well as by the materiality and emotional associativeness of stuffed animals. This allowed G. Klumbytė to develop a kind of embodied 'feeling for the machine' (cf. Barbara McClintock's 'feeling for the organism' - Fox Keller 1983) which was both pragmatic (understanding how to 'play'), as well as functional (it did communicate something about the internal workings of the machine learning tool as an engine behind the instrument's vocalizations). Additionally, by engaging the 'player's' embodiment and presenting instrument-like qualities of interaction, this work also drew author's attention to tacit and affective dimension of knowledge that emerges through practice.

Explaining LLMs through Tangible Play

The second example we describe is a process of building a tangible explanation prototype for an LLM. This process took place during a workshop, which was co-organised by the authors with colleagues Leonardo Angelini, Maxime Daniel, Nadine Couture and Elena Mugellini, presented during ACM conference on Tangible, Embedded, Embodied In-

Figure 2: The LLM pinball machine composed of cardboard structure and pins representing temperature

teraction (Angelini et al. 2025). The goal of the workshop was to explore how tangible, embodied, and material interaction modalities can facilitate explainable AI design as well as lead to better understanding of how LLMs work. The process was facilitated by providing participants with two sets of cards: one set introducing different properties of LLMs that might merit explanation, such as embeddings, latent space, temperature, attention, overall function, and representations and control; and another set that listed different tangible and embodied aspects of interaction, such as exploitation of human skills and senses, persistency, spatial interactions and collaboration, physical constraints to the interaction space, immediacy and intuitiveness, and others (the card set was based on Angelini et al. 2018). Participants were also provided a table with a broad set of physical materials, from string and paper, to fabric, carving stone, heat-sensitive paper, and others.

The authors were part of a smaller group of five people that decided to explore the tangible, embodied, and material possibilities of explaining temperature - a feature that describes the 'creativity' of LLM or the randomness and unpredictability of the model's output. The higher the temperature, the more random the output selection, thus the more possibility of surprising and unexpected result. The lower the temperature, the more 'conservative' the output selection, since LLM picks the output with the highest likelihood. The group settled on temperature feature after a brainstorming phase, during which we discussed different modalities, such as touch, temperature, color, voice, and other physical possibilities of expression and how they would connect to the different technical characteristics of LLMs and their overall function. Focusing on temperature, the idea of constructing some kind of physical game, such as a pinball machine (figure 2), emerged, followed by further discussions about how might a pinball machine illuminate the function of temperature in an LLM and what kind of tangible interaction might be needed for that. In the end, we constructed a machine out of cardboard, sponges, and wooden sticks that would illustrate how LLM picks a specific word out of different word options to complete different sentences (figure 3).

To demonstrate this, we replicated the structure of a poetry generator created by artist Alison Knowles, in association with James Tenney at CalArts in 1967, which was written for the early programming language Fortran (Taylor 2009). Knowles suggested that computer could generate poetry using a specific formula:

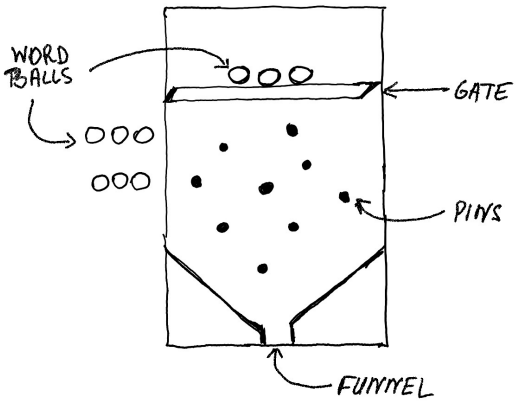

Figure 3: Schematic of the LLM pinball machine

a house of (list material)
in (list location)
using (list light source)
inhabited by (list inhabitants)

The beginning of each sentence was written on the machine, while the differently colored felt balls represented groups of three words that could be selected to complete each sentence. The user would pick up the group of felt balls and place them against the gate. Then they could set the temperature by adding or removing the number of wooden pins. The fewer the pins - the lower the temperature and the more likely the middle ball would reach the bottom of the funnel first, thus representing the more 'conservative' and less 'creative' (i.e., randomized) selection of the word to complete the sentence. Once the desired temperature was set, the user would lift the cardboard gate and allow the balls to drop through the pins towards the funnel made of sponge. The ball that reached the middle first was the one selected to complete the sentence.

What was striking about this particular experience for the authors, was that the process of building a tangible machine and trying out its operations generated more nuanced understanding of how LLM temperature feature and the LLM itself works. We started thinking about the LLM itself as a kind of 'probability machine' of word generators, contrary to the popular imaginary of LLMs as search engines or truth machines (Duh, Gomez, and Bethard 2024; Steyvers et al. 2025). Additionally, with each step of building the pinball machine, with each test of how it might work, discussions ensued on whether or not it represents the functioning of temperature accurately, whether it even should do so or if representation of 'general idea' of temperature is enough,

thus deepening our own understanding of LLMs and its possible effects. Working with tangible materials, building a physical representation of an LLM, and doing it collaboratively through trial and error and tangible play, allowed us to notice how all of these aspects of interaction - tangibility, materiality, collaboration - contributed to us developing a better grasp of LLM function overall as well as the role of temperature specifically.

Discussion

While the above examples are not fully-fledged case studies, we nonetheless think that they provide interesting starting points and illustrations to think about how explainability and understanding are currently addressed and designed in XAI and the new directions for design and conceptualization of these terms that can open up when taking materiality, embodiment, and tangibility into account.

Explainability and Understanding as Process

First, with regards to conceptual implications, we suggest that these examples point to how explainability and understanding emerge as *processes* instead of system properties. Both in the case of interacting with the AI artwork, as well as in constructing the LLM pinball machine, the very process of material and embodied interaction and crafting allowed for more nuanced, layered understanding to emerge. In the case of the artwork-instrument, G. Klumbytė felt like they are acquiring a *feeling* for the machine, intuitive and tacit knowledge. The instrument's very explainability - as a characteristic conventionally describes the system itself - was unfolding and increasing during the interaction. They experienced it as a kind of 'settling into the body' of understanding of how the instrument works, which in turn increased their perception of the explainability of the instrument. In the case of making a tangible game to represent LLM and its temperature function, it was literally the process of building it, collaboratively, that allowed understanding to unfold, whereas the explainability of LLM's temperature characteristic became a tangible quality to be played with and engaged in interaction.

Furthermore, if explainability and understanding emerge as a process, and especially as a process of *interaction*, then it is also important to highlight that this interaction spills over and encompasses more than just a 'human' and an 'AI', but also other agents, such as workshop and project collaborators, and the physical materials that constituted the artwork and the pinball machine - in other words, it is extended and collective. For instance, holding the little felt ball with the word in one's hands, playing with the various distributions of balls and pin constellations, while at the same time engaging in a lively discussion with workshop participants about temperature of LLM could all be considered part of the interaction setting, contributing to the emergence of understanding and explanation. This points to the possibility of XAI design and particularly tangible, embodied, and material XAI design to take into account not only the specific properties to be explained, the users and stakeholders who need explanation, and the interface modalities, but also the broader material setting and relations among different agents participating in the interaction space.

While the importance of interaction in generating understanding is not in itself a new insight (Selbst and Barocas for instance noted this already in 2018), it is specifically the material, embodied and collective/extended qualities of interaction that our work opens as promising for future explainable AI design. The artistic work around feeling AI raises conceptual questions around what happens when understanding of AI systems is acquired through interaction, when internal workings of the system might not be known - which might be especially the case for lay users. It also encourages to ask what the role of affective embodied engagement and knowledge might be and how this knowledge can be invited into interactive XAI and explainability design. The LLM pinball machine meanwhile points to practical possibilities of leveraging collective forms of material interaction and physicalization for XAI design. Specifically, we think that materialist-embodied approaches could bear fruit in cases of designing human-centred explanations when working with lay users and other affected stakeholders of AI systems as they allow to expand explanation modalities and make explanations potentially more accessible and graspable for more diverse target audiences.

Materiality and Embodiment as Agents in XAI

This leads us to our next insight, which is that materiality and (in this case human) embodiment are important agents in the process of explainability and understanding. On the one hand, quite straightforwardly, in tangible, embodied, and material interaction, the various properties and agencies of materials themselves play a role in the design. For instance, different materials in the workshop (such as rocks, thread, wood, paper, felt) have different physical communicative capacities (i.e. they can express different things), and our group spent some time discussing indeed what kind of materials can express which kind of qualities. Additionally, as people who work with physical materials, such as craftspeople and artists, know well, materials have their own agency, e.g., they can behave in ways that are stubborn or unexpected, and they therefore demand that we acknowledge them and find ways to collaborate. For instance, during the building of the pinball machine we figured that felt balls do not roll as smoothly or fast as a granite ball would, and we considered adding weights to them, but that would have complicated the representation of probabilities and their distributions.

In the artwork-instrument, the physicality and texture of the stuffed toys, as well as emotional charge of what stuffed toys represent (conventionally, they are associated with childhood, yet for each person they might have additional personal emotional associations), made itself felt in the interaction and guided the kind of engagement that became possible (touching, squeezing, varying intensity). Furthermore, the interaction itself was very intimate and required an embodied engagement: G. Klumbytė reflected that the interaction invited a kind of engagement that felt like *caressing* the instrument, whereas its vocalizations reminded them of a theremin - an electronic instrument that is played

spatially, without direct touch. This drew the author's and collaborators' attention to the role of affective and emotional aspects of interaction with AI. Engaging in an embodied and affective way with this AI artwork allowed us to rise and probe questions about the affective and embodied *effects* of AI in our lives and in the lives of other users - a dimension that is often utilized in emotion AI design and chatbot design, but not necessarily included in explanation of these systems. This affective and materially embodied emotional dimension was therefore present in multiple ways - physical, metaphorical, discursive - in this interaction. Additionally, the composition of this artwork also showed that AI itself can be seen as a specific kind of digital material, with its own capacities and limitations (Dourish and Mazmanian 2013; Reichert and Richterich 2015).

Ethical Explanation Design as Material Practice of Care

Last but not least, materiality, embodiment, and tangibility in XAI can help us think about the kind of careful and attentive labor that explainability and understanding require. In her book *Matters of Care: Speculative Ethics in More Than Human Worlds*, Puig de la Bellacasa (2017) develops a care-based approach to thinking about ethics, responsibility, and relationality in an interconnected world that includes humans and more-than-human actors. Referencing Latour's concept of 'matters of concern' (2004) that he uses to describe how certain objects and topics become political and significant (as 'matters of concern'), Bellacasa re-directs the question and asks how to think not only about concern but about matters of care - which is to say, an affective, ethical, practical labor of maintaining, repairing, and enabling life-sustaining relations. This labor entails material practices, emotional labor as well as ethical commitment.

The above examples illustrate how what requires explanation and the process through which explanation and understanding emerges are constructed through careful attention and negotiation. For example, the artwork instrument coupled machine learning tool to a specific physicality of the instrument, and it is precisely this whole coupling that can become known and understood - not just the logic of the machine learning model alone. This draws attention to how AI and its various embodiments (in software and hardware), as well as the physical and sociocultural settings of these embodiments, the specific domain within which AI operates, and the stakeholders that will be affected by these operations are interlinked. We therefore argue that it is also this specific linking, this interconnection and its effects - and not just the internal logic of the model - that requires explanation and understanding. Tangible, material and embodied interactions can be particularly good means of making these interconnections perceptible and graspable.

Furthermore, paying attention to the 'matters of explanation', to paraphrase Puig de la Bellacasa, also allows to understand explanation - and the ethical dimension of explainability - as a *material practice* (Bellacasa herself understands ethics indeed as such a material practice of care). To construct interactions within which explainability and understanding can emerge requires a material practice of paying attention to who and what is included in the explanation, what kind of material access to knowledge and understanding different stakeholders might have or lack, what kind of concrete situations of interactions are germinal to or prevent understanding to emerge. Making things tangible, embodied, materially graspable can go a long way to literally 'make things matter' and draw attention to 'what matters' and to whom in an ethical explainable AI design. In other words, what we wish to point out here is that explanation and XAI design is an ethico-political question and the scene where it unfolds is not only ethical guidelines and regulations but also the very material design practices and decisions.

Conclusion

In this paper we argued that material, embodied, tangible approach that incorporates new materialist philosophical perspectives - what we called the materialist-embodied approach - allows to address explainability and understandability not as system properties but as *processes* that emerge in interaction. This interaction entails both human and non-human actors and agencies, and invites to think about the ethical work of XAI as a *material practice of care*: a kind of care-full labor of creating explanations and understanding in ways that matter, paying close attention to who is involved and who is addressed in this process and how. In that sense, we concur with insights from previous tangible XAI research (Colley, Väänänen, and Häkkilä 2022; Kaisa Väänänen, Ashley Colley, and Jonna Häkkilä 2024; Ghajargar and Bardzell 2022) that materialist-embodied approach to XAI design can help expand access to explanations, knowledge, and understanding by providing a more diverse array of modalities of engagement. Based on our experience and ongoing work, we would add that this approach more generally opens a domain of questions around relations between embodiment, affect, and knowledge that are not yet widely explored in XAI and its ethics.

To conclude this paper, we would like to reflect back on the intuitive assumptions that we mentioned in the beginning that undergird explainability and understandability in XAI, namely: (1) that systems perform objective functions driven by internal logic, which can or at least can be attempted to be explained; (2) that explainability is a property of a system and that systems can therefore be more or less explainable; (3) that explanations can lead to understanding, which is in a sense a property of the human user - a certain measurable characteristic (more/less, better/worse understanding) that can be acquired; and (4) that both explanation and understanding are fundamentally functions of reason-based cognition. Materialist-embodied perspective subtly shifts these assumptions by drawing focus on process, material agency, and embodiment towards: (1a) addressing systems as performing internal logic that is linked, through their material embodiments, to specific contexts and domains, and positions these links as in need of explanation; (2a) that explainability is an emergent property of interaction which can and often does entail broader physical and social contexts and collective practices; (3a) that explanations can lead to understanding, which is a process that also emerges interaction; and (4a) that both explanation and understanding are

dynamic processes of cognition that is embodied, embedded, enactive, and extended. These new assumptions can help open further possibilities for XAI design.

The challenge, of course, is how to implement these new assumptions into pragmatic XAI design situations. As AI applications are broad, so the situations and interactions are indeed varied and we appreciate the difficulty of trying to design a context-specific, material, embodied and tangible interaction for these situations. Nonetheless, we contend that designing with material-embodied perspective allows for an expansive and inclusive approach to XAI design that opens the field for more diverse range of bodies, embodiments, and experiences. We therefore hope that this paper can contribute to further research and design in this direction.

Acknowledgments

We are grateful to our former research assistant Hannah Piehl for her support in the initial stages of this research. We are thankful to our collaborators Mika Satomi and Daniela K. Rosner, and co-authors of the workshop - Leonardo Angelini, Maxime Daniel, Nadine Couture, and Elena Mugellini, as well as participants of the workshop. Our thanks also goes to anonymous reviewers for their thoughtful feedback and suggestions. This research has been supported by Volkswagen Foundation funding for the project 'AI Forensics: Accountability through Interpretability in Visual AI Systems', within the funding line 'Artificial Intelligence and the Society of the Future'.

References

Angelini, L.; Klumbyte, G.; Daniel, M.; Couture, N.; Mugellini, E.; and Draude, C. 2025. Tangible LLMs: Tangible Sense-Making For Trustworthy Large Language Models. In *Proceedings of the Nineteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '25. New York, NY, USA: Association for Computing Machinery. ISBN 9798400711978.

Angelini, L.; Mugellini, E.; Couture, N.; and Abou Khaled, O. 2018. Designing the Interaction with the Internet of Tangible Things. In Fernaeus, Y.; McMillan, D.; Jonsson, M.; Girouard, A.; and Tholander, J., eds., *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, 299–306. New York, NY, USA: ACM. ISBN 9781450355681.

Arrieta, A. B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R.; and Herrera, F. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58: 82–115.

Arunika, M.; Saranya, S.; Charulekha, S.; Kabilarajan, S.; and Kesavan, G. 2024. A Survey on Explainable AI Using Machine Learning Algorithms Shap and Lime. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–6. IEEE. ISBN 979-8-3503-7024-9.

Bertrand, A.; Viard, T.; Belloum, R.; Eagan, J. R.; and Maxwell, W. 2023. On Selective, Mutable and Dialogic XAI: a Review of What Users Say about Different Types of Interactive Explanations. In Schmidt, A., ed., *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ACM Digital Library, 1–21. New York,NY,United States: Association for Computing Machinery. ISBN 9781450394215.

Blaha, L. M.; Abrams, M.; Bibyk, S. A.; Bonial, C.; Hartzler, B. M.; Hsu, C. D.; Khemlani, S.; King, J.; St Amant, R.; Trafton, J. G.; and Wong, R. 2022. Understanding Is a Process. *Frontiers in systems neuroscience*, 16: 800280.

Cha, I.; Oh, J.; Park, C. Y.; Han, J.; and Lee, H. 2023. Unlocking the Tacit Knowledge of Data Work in Machine Learning. In Schmidt, A.; Väänänen, K.; Goyal, T.; Kristensson, P. O.; and Peters, A., eds., *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–7. New York, NY, USA: ACM. ISBN 9781450394222.

Clark, A.; and Chalmers, D. 1998. The Extended Mind. *Analysis*, 58(1): 7–19.

Colley, A.; Väänänen, K.; and Häkkilä, J. 2022. Tangible Explainable AI - an Initial Conceptual Framework. In Döring, T.; Boll, S.; Colley, A.; Esteves, A.; and Guerreiro, J., eds., *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*, 22–27. New York, NY, USA: ACM. ISBN 9781450398206.

Coole, D. H.; and Frost, S., eds. 2010. *New materialisms: Ontology, agency, and politics*. Durham, N.C. and London: Duke University Press. ISBN 978-0-8223-4772-9.

Davis, N.; Hsiao, C.-P.; Yashraj Singh, K.; Li, L.; and Magerko, B. 2016. Empirically Studying Participatory Sense-Making in Abstract Drawing with a Co-Creative Cognitive Agent. In Nichols, J.; Mahmud, J.; O'Donovan, J.; Conati, C.; and Zancanaro, M., eds., *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 196–207. New York, NY, USA: ACM. ISBN 9781450341370.

de Jaegher, H.; and Di Paolo, E. 2007. Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6(4): 485–507.

Deshpande, M.; and Magerko, B. 2024. Embracing Embodied Social Cognition in AI: Moving Away from Computational Theory of Mind. In Mueller, F. F.; Kyburz, P.; Williamson, J. R.; and Sas, C., eds., *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–7. New York, NY, USA: ACM. ISBN 9798400703317.

Dourish, P. 2004. *Where the Action Is: The Foundations of Embodied Interaction*. Cambridge (Mass.): MIT Press. ISBN 9780262541787.

Dourish, P.; and Mazmanian, M. 2013. Media as Material: Information Representations as Material Foundations for Organizational Practice. In Carlile, P. R.; Nicolini, D.; Langley, A.; and Tsoukas, H., eds., *How Matter Matters*, 92–118. Oxford University Press. ISBN 9780199671533.

Duh, K.; Gomez, H.; and Bethard, S., eds. 2024. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human*

Language Technologies (Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics.

Ehsan, U.; and Riedl, M. O. 2020. Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach. In Stephanidis, C.; Kurosu, M.; Degen, H.; and Reinerman-Jones, L., eds., *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, 449–466. Springer, Cham.

Ehsan, U.; Wintersberger, P.; Liao, Q. V.; Mara, M.; Streit, M.; Wachter, S.; Riener, A.; and Riedl, M. O. 2021. Operationalizing Human-Centered Perspectives in Explainable AI. In Kitamura, Y., ed., *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM Digital Library, 1–6. New York,NY,United States: Association for Computing Machinery. ISBN 9781450380959.

Esposito, E. 2023. Does Explainability Require Transparency? *Sociologica, Vol. 16 No. 3*.

Eubanks, V. 2018. *Automating Inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press. ISBN 9781250074317.

Fox, N. J.; and Alldred, P. 2015. New materialist social inquiry: designs, methods and the research-assemblage. *International Journal of Social Research Methodology*, 18(4): 399–414.

Fox Keller, E. 1983. *A Feeling for the Organism: The Life and Work of Barbara McClintock*. San Francisco: W.H. Freeman and Co.

Fridland, E.; and Stichter, M. 2021. It just feels right: an account of expert intuition. *Synthese*, 199(1-2): 1327–1346.

Ghajargar, M.; and Bardzell, J. 2022. Making AI Understandable by Making it Tangible: Exploring the Design Space with Ten Concept Cards. In Sweetser, P.; Taylor, J. L.; Martin, C.; Mckay, D.; Rogerson, M.; Cumbo, B.; Wadley, G.; Hespanhol, L.; Tsimeris, J.; Xi, M.; Turner, J.; Yoo, S.; Cooper, N.; Rahman, J.; Andres, J.; Pillai, A. G.; and Kutay, C., eds., *Proceedings of the 34th Australian Conference on Human-Computer Interaction*, 74–80. New York, NY, USA: ACM. ISBN 9798400700248.

Ghajargar, M.; Bardzell, J.; Renner, A. S.; Krogh, P. G.; Höök, K.; Cuartielles, D.; Boer, L.; and Wiberg, M. 2021. From "Explainable AI" to "Graspable AI". In Wimmer, R., ed., *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, ACM Digital Library, 1–4. New York,NY,United States: Association for Computing Machinery. ISBN 9781450382137.

Ghajargar, M.; Bardzell, J.; Smith-Renner, A. M.; Höök, K.; and Krogh, P. G. 2022. Graspable AI: Physical Forms as Explanation Modality for Explainable AI. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 1–4. New York, NY, USA: ACM. ISBN 9781450391474.

Hartmann, M.; Du, H.; Feldhus, N.; Kruijff-Korbayová, I.; and Sonntag, D. 2022. XAINES: Explaining AI with Narratives. *KI - Künstliche Intelligenz*, 36(3-4): 287–296.

Holzinger, A.; Saranti, A.; Molnar, C.; Biecek, P.; and Samek, W. 2022. Explainable AI Methods - A Brief Overview. In Holzinger, A.; Goebel, R.; Fong, R.; Moon, T.; Müller, K.-R.; and Samek, W., eds., *xxAI - Beyond Explainable AI*, volume 13200 of *Lecture Notes in Computer Science*, 13–38. Cham: Springer International Publishing. ISBN 978-3-031-04082-5.

Höök, K. 2018. *Designing with the Body*. The MIT Press. ISBN 9780262348324.

Hornecker, E.; and Buur, J. 2006. Getting a grip on tangible interaction. In Grinter, R.; Rodden, T.; Aoki, P.; Cutrell, E.; Jeffries, R.; and Olson, G., eds., *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 437–446. New York, NY, USA: ACM. ISBN 1595933727.

Ishii, H.; and Ullmer, B. 1997. Tangible bits. In Pemberton, S., ed., *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 234–241. New York, NY, USA: ACM. ISBN 0897918029.

Jin, W.; Fan, J.; Gromala, D.; Pasquier, P.; and Hamarneh, G. 2021. EUCA: the End-User-Centered Explainable AI Framework. *arXiv*, arXiv:2102.02437.

Kaisa Väänänen; Ashley Colley; and Jonna Häkkilä. 2024. Towards Adaptive AI Explanations with Tangible User Interfaces. In *2024 ACM International Conference on Intelligent User Interfaces Workshops, IUI-WS 2024*, CEUR Workshop Proceedings. CEUR-WS.

Kaur, H.; Adar, E.; Gilbert, E.; and Lampe, C. 2022. Sensible AI: Re-imagining Interpretability and Explainability using Sensemaking Theory. In Association for Computing Machinery, ed., *FAccT 2022*, ICPS, 702–714. New York, New York: The Association for Computing Machinery. ISBN 9781450393522.

Latour, B. 2004. Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry*, 30(2): 225–248.

Liao, Q.; and Sundar, S. S. 2022. Designing for Responsible Trust in AI Systems: A Communication Perspective. In Association for Computing Machinery, ed., *FAccT 2022*, ICPS, 1257–1268. New York, New York: The Association for Computing Machinery. ISBN 9781450393522.

Maclure, J. 2021. AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind. *Minds and Machines*, 31(3): 421–438.

Menary, R. 2010. *The Extended Mind*. Cambridge Mass.: MIT Press.

Merry, M.; Riddle, P.; and Warren, J. 2021. A mental models approach for defining explainable artificial intelligence. *BMC medical informatics and decision making*, 21(1): 344.

Mindlin, D.; Beer, F.; Sieger, L. N.; Heindorf, S.; Esposito, E.; Ngonga Ngomo, A.-C.; and Cimiano, P. 2025. Beyond one-shot explanations: a systematic literature review of dialogue-based xAI approaches. *Artificial Intelligence Review*, 58(3).

Newen, A.; de Bruin, L.; and Gallagher, S. 2018. *The Oxford Handbook of 4E Cognition*. Oxford University Press. ISBN 9780198735410.

O'Neil, C. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: B/D/W/Y Broadway Books. ISBN 0553418831.

Ooge, J.; and Verbert, K. 2022. Explaining Artificial Intelligence with Tailored Interactive Visualisations. In Association for Computing Machinery, ed., *IUI '22 Companion*, 120–123. New York, NY, United States: Association for Computing Machinery. ISBN 9781450391450.

Preece, A.; Harborne, D.; Braines, D.; Tomsett, R.; and Chakraborty, S. 2018. Stakeholders in Explainable AI.

Puig de la Bellacasa, María. 2017. *Matters of Care: Speculative Ethics in More than Human Worlds*. Minneapolis: University of Minnesota Press. ISBN 9781517900656.

Raees, M.; Meijerink, I.; Lykourentzou, I.; Khan, V.-J.; and Papangelis, K. 2024. From explainable to interactive AI: A literature review on current trends in human-AI interaction. *International Journal of Human-Computer Studies*, 189: 103301.

Reichert, R.; and Richterich, A. 2015. Introduction: Digital Materialism. *Digital Culture & Society*, 1(1): 5–18.

Rohlfing, K. J.; Cimiano, P.; Scharlau, I.; Matzner, T.; Buhl, H. M.; Buschmeier, H.; Esposito, E.; Grimminger, A.; Hammer, B.; Hab-Umbach, R.; Horwath, I.; Hullermeier, E.; Kern, F.; Kopp, S.; Thommes, K.; Ngonga Ngomo, A.-C.; Schulte, C.; Wachsmuth, H.; Wagner, P.; and Wrede, B. 2021. Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3): 717–728.

Samek, W.; Montavon, G.; Vedaldi, A.; Hansen, L. K.; and Müller, K.-R., eds. 2019. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer, Cham. ISBN 1611-3349.

Schulz, C. 2023. From Mental Models to Algorithmic Imaginaries to Co-Constructive Mental Models. *: Navigationen - Zeitschrift für Medien- und Kulturwissenschaften*, 23(2): 66–76.

Selbst, A. D.; and Barocas, S. 2018. The Intuitive Appeal of Explainable Machines. *Fordham Law Review*, 87(3): 1085–1139.

Smit, D.; Hengeveld, B.; Murer, M.; and Tscheligi, M. 2022. Hybrid Design Tools for Participatory, Embodied Sensemaking: An Applied Framework. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 1–10. New York, NY, USA: ACM. ISBN 9781450391474.

Steyvers, M.; Tejeda, H.; Kumar, A.; Belem, C.; Karny, S.; Hu, X.; Mayer, L. W.; and Smyth, P. 2025. What large language models know and what people think they know. *Nature Machine Intelligence*, 7(2): 221–231.

Taylor, S. 2009. Alison Knowles, James Tenney and the House of Dust at CalArts.

van Dijk, J. 2024. Making sense with things in participatory design. In Nimkulrat, N.; and Groth, C., eds., *Craft and Design Practice from an Embodied Perspective*, 183–194. New York: Routledge. ISBN 9781003328018.

Varela, F. J.; Thompson, E.; and Rosch, E. 2016. *The embodied mind: Cognitive science and human experience*. Cambridge Massachusetts and London England: MIT Press, revised edition edition. ISBN 9780262529365.

Wiberg, M. 2018. *The materiality of interaction: Notes on the materials of interaction design*. Cambridge Massachusetts: The MIT Press. ISBN 9780262037518.