Tangible Explainable AI - an Initial Conceptual Framework

Ashley Colley
University of Lapland
Finland

Kaisa Väänänen
Tampere University
Finland

Jonna Häkkilä
University of Lapland
Finland

ABSTRACT

Artificial Intelligence (AI) solutions are becoming prevalent in almost all aspects of human life. However, their acceptance may be limited by a lack of transparency of how the AI works. Explainable AI (XAI) aims to provide the users of AI systems with an understanding of why decisions are made, increasing trust in the system. To date, research into XAI has focused on the use of graphical user interfaces, presenting numerical, textual or graphical explanations. However, AI is increasingly being used in systems that include physical devices, and hence the need for explainability in physical or tangible user interfaces (TUI) is also increasing. We present an initial conceptual framework for tangible explainable AI (TangXAI), which identifies the potential approaches of communicating XAI through physical artifacts, using the concepts of data physicalization and tangible interaction. The framework provides a basis into which ongoing research of tangible explainable AI can be mapped and related research gaps identified.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Ubiquitous and mobile computing**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Explainable AI, XAI, tangible interaction, TUI, human in the loop, human centered AI, TangXAI, HCI

ACM Reference Format:
Ashley Colley, Kaisa Väänänen, and Jonna Häkkilä. 2022. Tangible Explainable AI - an Initial Conceptual Framework. In *21th International Conference on Mobile and Ubiquitous Multimedia (MUM 2022), November 27–30, 2022, Lisbon, Portugal*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3568444.3568456

1 INTRODUCTION

Artificial Intelligence (AI) has rapidly grown to be a major theme in the research and development of interactive systems. AI is expected to be integrated to virtually all application domains across different life sectors, and will affect people on both individual and societal levels. When considering the field of human-computer interaction, AI is expected to become one of the core components in the design of future interactive systems. The characteristics of AI will drive a shift from reactive information tools to proactive agents, and set new challenges for human-centered design [32].

According to the European Commission Ethics Guidelines [15], AI systems should empower human beings, allowing them to make informed decisions and foster their fundamental rights. Indeed, the European Union data protection law includes a right to explanation. Human-Centered AI (HCAI) can be defined as an approach to strive for ethical AI for common good, putting people and their needs at the center of any AI solution, and considering their wider sociocultural context [10, 36, 46]. One element in HCAI is the Human-in-the-Loop approach, where the human and AI collaborate in the decision making process, e.g. through the human providing feedback on the machine's decisions [48].

Intelligent computing systems are easily perceived as black boxes by people interacting with or affected by them, and transparency is a key quality criterion of human-AI interaction. Explainability is associated with the notion of explanation as an interface between humans and a decision maker that is both an accurate proxy of the decision maker and comprehensible to humans [16]. Thus, explainable AI (XAI) helps users to understand the algorithms and decisions of AI, e.g. giving a reason for a particular decision [46]. Explainability can be considered as a bridge to avoid unwanted or even unethical use of algorithmic outputs. From a social viewpoint, explainability can be seen as the capacity to reach and guarantee fairness in AI [7].

To date, research into XAI has primarily focused on the use of graphical user interfaces, presenting explanations in numeric, textual or graphical format, e.g. [2]. However, the penetration of AI into physical systems – such as smart devices and embedded systems – is increasing, and hence the need for explainability in physical or tangible user interfaces (TUI) is also becoming apparent. Research on tangible interfaces for explainable AI - which we refer to as Tangible XAI (TangXAI) - is only just beginning to emerge. For example, almost the only publication on the topic is a 2021 workshop 'from explainable to graspable AI' [13]. A further challenge for HCI design with XAI is that, so far, only initial research linking the technical aspects of XAI and the concepts of HCI has been made, e.g. [27]. Such mapping is needed in order to enable structured design approaches.

In this paper, we present an initial conceptual framework highlighting how the fields of XAI and TUI can be brought together to create intuitive interfaces for a variety of future smart devices. The framework is constructed based on merging of the concepts from existing XAI and TUI frameworks found in literature, specifically the XAI framework by Belle & Papantonis [9] and the TUI framework by Hornecker & Buur [20]. In the following, we first briefly draw together the prior art on XAI and TUIs. After that, we present the positioning of our work within the scope of Human-Centered AI. Finally, we present a conceptual framework how the different XAI approaches can link to TUI design. This initial Tangible XAI

MUM 2022, November 27–30, 2022, Lisbon, Portugal

ACM ISBN 978-1-4503-9820-6/22/11.
https://doi.org/10.1145/3568444.3568456

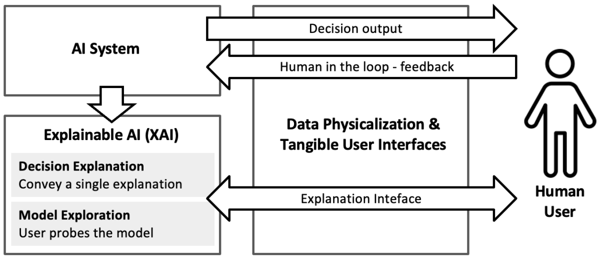


Figure 1: Tangible XAI - highlighting the relevant domains and their interconnections.

(TangXAI) framework can be used to conceptualize and design for different kinds of tangible interactions to help explain AI's decisions to users.

2 PRIOR ART

To provide background to our research, we review prior works on XAI concepts, tangible interaction and human-centered AI. Figure 1 presents the positioning of each of the reviewed domains and their connection with the human user.

2.1 Explainable AI (XAI) Concepts

AI, in the form of intelligent systems, has become a ubiquitous - and, for most people, mysterious - element in our daily lives. This has raised issues of trust, control and transparency, spawning the research field of Explainable AI (XAI) [26, 31]. By explaining, or exposing the inner workings of AI systems, XAI aims to provide an understanding of the reasons behind the system's decisions, such that users identify that the system is operating fairly, without bias and providing correct outputs. Through this, users' trust in the system is increased [33].

The currently extensive amount of research ongoing on XAI primarily focuses on either 1) developing XAI 'algorithms' to address the trade-off between high prediction performance and ease of explainability, e.g. from black-box models to transparent surrogate models [27, 28], or 2) presenting taxonomies of the problem space, e.g. [6, 7, 9, 28]. Some recent works, e.g. Chromik & Butz [12] build on this to present guidelines. However, there are only a few works that present actual novel implementations of XAI interfaces, and fewer still that present their evaluation with end users. Exceptionally, Hohman [19] demonstrates XAI implementations using interactive online articles published in the parametric press. Hohman's works focus on "machine learning interpretability", i.e. explaining the underlying AI models and the potential for bias, rather than addressing explainability during user interaction.

Belle & Papantonis [9] identify 4 general approaches to explainability, simplified rule extraction, feature relevance, local explanations, and visual explanations, and provide examples of the most common techniques used in each category (Figure 2). Several works have presented classification of XAI systems by their input and output formats, e.g. Vilone et al. [43] categorize inputs as outputs as numerical, rules, textual, visual, and mixed. Through a literature review, the authors identified, e.g. that numerical output formats are used primarily for global explanations, whilst textual formats are used for local explanations [43]. At high level, we aim towards finding similar insights when tangible interaction is used as the communication medium. Here, we note that whilst Vilone et al.'s classification decouples the input and output formats from each other, in tangible interaction the user experience often relies on the close combination of input and output modes, i.e. real-time feedback (see e.g. [20]).

2.2 Data Physicalization and Tangible Interaction

Data physicalization transforms data beyond visual representation on paper or screens and gives it a physical form, and, as a consequence, transforms it from the virtual to the physical world [18, 22]. Physical representation of data also creates the possibility for users to interact with the data, leveraging human cognitive skills learnt from the natural world. Lupton [30] discusses the visceral benefits of "feeling data", acknowledging that "... humans, digital technologies and digital data participate and work together in feeling in complex ways" (p. 13). In the context of XAI, data physicalization can be leveraged to provide an intuitive means for users to interact with a physical proxy representing the complex data in the AI system model [39].

The scope of TUIs that may provide novel interface when coupled with AI is broad, stretching from materiality, texture and shape changing to spatial interaction [20]. The materiality of tangible user interfaces has been a focus of study [44], e.g. introducing the concept of computational composites [42]. Shape changing interfaces [34] have, to our knowledge, not yet been connected with XAI systems. However, e.g. Alexander et al. [1] highlight their potential application areas, including creating adaptive affordances, augmenting users, and communicating information. From a study exploring the embodiment of AI enabled voice assistants, such as Amazon Echo, Spallazzo et al. [38] call for "...fostering a more natural interaction, going beyond display-mediated interfaces", echoed in Chromik & Butz [12] XAI interface guideline of "complementary naturalness".

Hornecker and Buur have presented a tangible interaction framework, including the themes of expressive representation, tangible manipulation, spatial interaction and embodied facilitation [20]. *Expressive Representation* focuses on the potential to convey expressive meaning though the material qualities and digital representations of a tangible interaction systems. *Tangible manipulation* focuses on a user's tactile interaction with physical objects, which are coupled to computational systems. Hornecker and Buur highlight grabbing and moving interface elements, rapid feedback during interaction and the importance of metaphor between the interaction and its effect [20]. *Spatial interaction* builds on humans natural understanding of the spatial relationships of objects and our ability to move within in, configuring the space around us. Embodied Facilitation highlights the effect of objects placement and movement in space to influence our social interactions.

2.3 Human Centered AI

Prior work has identified specific challenges in the design process for human-AI interaction, e.g. difficulties in iterative prototyping and testing [47]. Hartikainen et al. [17] found in their study of company practices that there is a gap in developers' end-user understanding due to, among other things, different perspectives of technical and human approaches. Echoing this, Giaccardi and Redström [14] call for a new paradigm in human-centered design methods, to address the challenges of AI. One key approach in a human centered AI system is the inclusion of human-in-the-loop, i.e. the outputs of the AI model are tuned based on user inputs. By including human-in-the-loop approaches as part of an XAI interface, i.e. users providing feedback on the correctness of individual predictions [26] or on the model as a whole [11], the performance of the AI system can be improved [25]. In particular, research has noted the benefits of real-time iterative feedback in AI systems [4, 5] - this is exactly one the strengths of TUIs.

3 CONCEPTUAL FRAMEWORK FOR TANGIBLE XAI RESEARCH

To provide an initial understanding of the potential for tangible interaction as a communication channel for XAI, we present an initial framework that overlays the two domains (Figure 2). Based on this, research can begin to identify the most promising connections between the domains. To build insights into the potential combinations of XAI and TUI we make an initial review for each of the explainability categories identified by Belle & Papantonis [9]. When considering the application of tangible interaction to XAI, we consider initial focus on on the more fundamental themes of expressive representation and tangible manipulation will be most suitable. The themes of spatial interaction and embodied facilitation address higher level interaction concepts and typically build on top of the other themes. In the following we make no distinction regarding the accuracy of the explanation produced by each XAI approach, as this requirement will be dictated by each specific use case (and possibly user). Rather we propose an direction towards 'usable explanation' (c.f. usable security [3]), where the overall user experience with the interactive system is considered, rather than its technical performance. Following this, Table 1 presents examples of possible XAI - TUI combinations for an AI that predicts if an individual is diabetic.

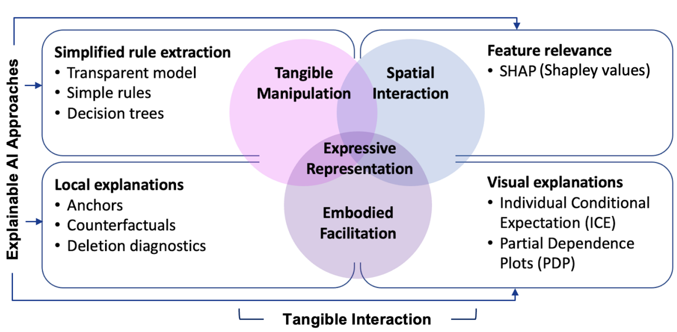


Figure 2: TangXAI conceptual framework combining explainable AI approaches (from Belle & Papantonis [9]) to be communicated by tangible interaction themes (extracted from Hornecker & Buur [20]).

Feature Relevance. With feature relevance based explanation, the influence of each input parameter on the model's outputs is given a score. Those parameters with higher scores are the most important in making the decision. Shapley values (SHAP) are the most popular method for extracting feature relevance [29]. One weakness of feature relevance approaches is that they ignore possible interaction effects between parameters. In some approaches only those parameters with highest scores are presented as an explanation [8].

Graphically, feature relevance can be straightforwardly visualized as a bar chart, with each parameter's score presented as the bar's length. Such visualizations maybe directly ported to tangible interfaces, e.g. representing each parameters with a physical object with dimensions based on the parameter's influence. With a tangible manipulation approach, the explainability output maybe combined with the parameter input mechanism. For example, with parameter value input using a set of physical sliders, parameters with higher relevance provide higher friction to slider thumb movement.

Local Explanations. Local explanations explain a particular decision by focusing on the model's behavior nearby the decision, i.e. they do not provide a full explanation of the AI model over its entire range of inputs. In simple terms, a local explanation answers a user's question 'why did the AI give this decision?'. Local explanations may be particularly suited for communication by tangible interaction, as they enable solutions with closely connected input and output modes, which can be designed as naturally engaging experiences.

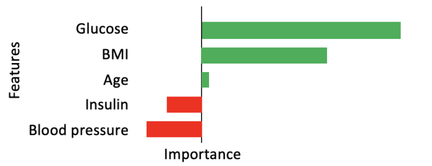
There are several different approaches to probe the model's behavior around a decision, including counterfactuals, anchors and deletion diagnostics. Counterfactuals demonstrate the minimum change in input parameters needed to change the decision, i.e. to cross a decision boundary. In an example TangXAI interface, a tangible 'phicon' [21] could be rotated to move input parameters across decision boundaries. In an anchors based approach, the decision boundary is simplified and a simple rule describing moving from the current decision to cross the decision boundary is created. Another way to demonstrate moving across a decision boundary is by deleting specific data points from the training data, i.e. 'deletion diagnostics'. Tangible interface implementations could, e.g. present a set of physical toggle buttons through which data points could be removed.

Simplified Rule Extraction. With this XAI approach, a simpler transparent proxy model is created from a black box model. The developer of the proxy model can select a suitable balance between the model's accuracy and its complexity to suit the intended use case. In its simplest form, the proxy model can be one or more "if-then" rules, e.g. arranged in a decision tree. For simple proxy models, expressive representation based approaches could include, e.g. a dynamically textured touchable surface, where the user is able to feel the path through the decision tree. Similarly, variations on Ishii and Ullmer's well known marble answering machine could present a tangible manipulation based approach for an interactive decision tree [21]. For communicating more complex proxy models, tangible tabletops, employing a combination of digital displays and tangible manipulation, present perhaps the most promising format,

Table 1: Examples of possible XAI - TUI combinations for an AI that predicts if an individual is diabetic, based on factors including body mass index (BMI) and age. The XAI guides the user to the reasons for the model's classification.

<table>
<tr><th>XAI Approach</th><th>Potential TUI</th></tr>
<tr><td>

Feature Relevance

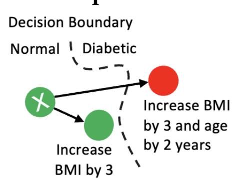


</td><td>

Tangible Bar Chart [40]

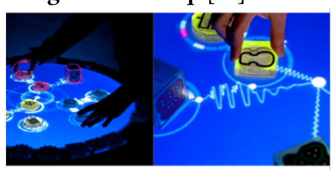
</td></tr>
<tr><td colspan="2">The influence of each input parameter on the model's outputs is scored. Those parameters with higher scores are the most important in making the decision, e.g. Shapley values (SHAP) [29]. Such values could be presented through tangible interaction using a physical bar chart interface [40]. Explainability output may be combined with the parameter input mechanism, e.g. parameters with higher relevance have higher friction resistance to movement.</td></tr>
<tr><td>

Local Explanations



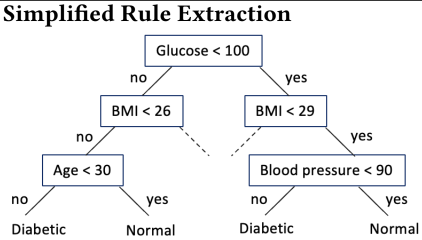
</td><td>

Tangible Tabletop [23]



</td></tr>
<tr><td colspan="2">From the AI model's output value (x) a minimal set of changes needed to change the output category are identified (counterfactuals), i.e. why did the AI give this decision? An interactive tabletop such as Jordà et al.'s ReacTable [23] could provide an intuitive interface for exploring local explanations. Local explanations may be particularly suited for communication by tangible interaction, as they enable solutions with closely connected input and output modes, which can be designed as engaging experiences. The related XAI approach of deletion diagnostics illustrates the effect of removing a data point.</td></tr>
<tr><td>

Simplified Rule Extraction

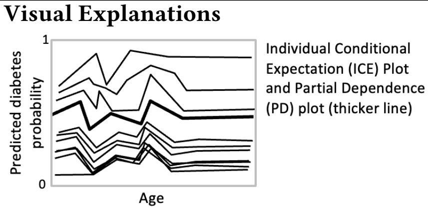


</td><td>

Smart Playing Cards [35]

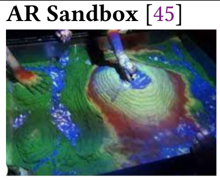
</td></tr>
<tr><td colspan="2">To explain the black box AI model, a simplified proxy decision tree model is created. An interactive card game with smart playing cards could provide an intuitive tangible interface for exploring the decision tree, e.g. [35]. For communicating more complex proxy models, tangible tabletops combining digital displays and tangible manipulation, present a promising format, e.g. Kubicki et al.'s TangiSense tabletop [24].</td></tr>
<tr><td>

Visual Explanations



</td><td>

AR Sandbox [45]



</td></tr>
<tr><td colspan="2">The Individual Conditional Expectation plot (ICE) fixes all factors except the one of interest (age in the example plot) and plots one line per instance. The mean of all instances is the partial dependence (PD). Such plots could be presented through tangible interaction, e.g. using an AR sandbox type interface [45], where the user can physically manipulate piles of sand which are augmented with projected colors based on their height.</td></tr>
</table>

e.g. employing interaction similar to the TangiSense interactive table [24].

Another approach to provide an overall understanding of the model is the presentation of representative example data points from the training data [41]. For example, with this approach a pressure based interface could be used to dynamically vary the number or representative data points presented.

Visual Explanations. In principle, plotting the AI model's output for the full range of input values can provide a picture from which some insights into the model's operation can be drawn. However, in most cases the resulting plot is too complex to be understood and hence some limitations need to be applied. In individual conditional expectation (ICE) plots, all the input parameters except one are fixed to a certain instance, and thus a simple chart of decision output as a function of the parameter of interest is created. Similarly, partial dependence plots (PDPs) average out all the model's parameters except the one of interest. Tangible chart representations have been a common form of tangible interface e.g. Taher et al.'s EMERGE 10x10 tangible bar chart [40]. Such TUIs can support both expressive representation and tangible manipulation aspects of tangible interaction.

The four approaches of XAI and the possibilities of tangible interactions within those form an initial framework we have named TangXAI. This conceptual framework can help designers and developers choose suitable combinations of tangible XAI concepts and implementing them for the users and their tasks in the contexts at hand. Table 1 provides examples of possible XAI - TUI combinations for an AI that predicts if an individual is diabetic, based on factors including body mass index (BMI) and age. The XAI guides the user to the reasons for the model's classification.

4 DISCUSSION AND CONCLUSION

AI solutions are becoming prevalent in almost all areas of human life, but people using or affected by those solutions are not always in control of the AI. Human-in-the-loop approach has been promoted in recent years, e.g. [37], to ensure that users of AI applications can guide and direct the functioning of the AI. AI has special characteristics such as proactiveness and dynamic outcomes, leading to unpredictability from the user's viewpoint [32]. Such characteristics caused by complex AI algorithms and vast amounts of data introduce non-transparency to the user interfaces of AI systems (ibid.). This calls for new human-entered design approaches - however, guiding frameworks for this are largely missing [17].

Explainable AI (XAI) is advocated as the approach to overcome the challenges of opaque AI systems [12]. While the recent XAI research has focused on graphical user interfaces, the everyday world is still largely tangible. The domain of tangible user interfaces (TUI) has been well established as a research field for several decades, and an extremely broad set of prior work, as well as commercial products exists. The capabilities of tangible interaction to create natural, intuitive user interfaces for complex data can be applied to provide interfaces to the complex and often abstract domain of explainable AI systems. As a topic within TUI, data physicalization is showing promise in making abstract concepts concrete to the users [1, 30].

In this short paper, we introduce a "merging" of the concepts of XAI and tangible interactions with AI, and propose an initial framework, TangXAI to help create understanding between the more technical explainability approaches and the viewpoint of the human user who acts in the world in an embodied manner. Hornecker and Buur's tangible interaction framework [20] overlaid on Belle & Papantonis's four general approaches to explainability [9] provides a comprehensive platform for understanding the possibilities of tangible and physical interactions for XAI. To our knowledge, this is the first time that the concepts of XAI and TUI have been connected on a conceptual level. In our framework we have illustrated the addition of an explainability layer to an AI system and how it could be combined with physical and tangible interaction. Through this framework, we aim to help developers improve the transparency and understandability of AI decision making, as well as enabling new forms of human-in-the-loop AI.

To further develop and evaluate the framework, a set of tangible interactions should be implemented and evaluated to identify those best suited to convey explanations of decisions of AI systems in our physical realm. In out future work we aim to conduct studies of real-world cases. Two especially prominent application domain for tangible XAI user interfaces are industrial machinery and personal health, in which AI outputs are based on a large and dynamic sets of input data, and in which the users have high stakes to remain in control.

ACKNOWLEDGMENTS

This work has been supported by the Lapland Robotics project, funded by the European Regional Development Fund (ERDF). We are also grateful to the funders of the KITE project, ERDF, Business Tampere and University of Tampere.

REFERENCES

[1] Jason Alexander, Anne Roudaut, Jürgen Steimle, Kasper Hornbæk, Miguel Bruns Alonso, Sean Follmer, and Timothy Merritt. 2018. Grand challenges in shape-changing interface research. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.
[2] Gulsum Alicioglu and Bo Sun. 2022. A survey of visual analytics for Explainable Artificial Intelligence methods. *Computers & Graphics* 102 (2022), 502–520.
[3] Florian Alt and Emanuel von Zezschwitz. 2019. Emerging trends in usable security and privacy. *i-com* 18, 3 (2019), 189–195.
[4] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *Ai Magazine* 35, 4 (2014), 105–120.
[5] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
[6] Plamen P Angelov, Eduardo A Soares, Richard Jiang, Nicholas I Arnold, and Peter M Atkinson. 2021. Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11, 5 (2021), e1424.
[7] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115.
[8] Lidia Auret and Chris Aldrich. 2012. Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering* 35 (2012), 27–42.
[9] Vaishak Belle and Ioannis Papantonis. 2021. Principles and practice of explainable machine learning. *Frontiers in big Data* (2021), 39.
[10] Raymond R Bond, Maurice D Mulvenna, Hui Wan, Dewar D Finlay, Alexander Wong, Ansgar Koene, Rob Brisk, Jennifer Boger, and Tameem Adel. 2019. Human

Centered Artificial Intelligence: Weaving UX into Algorithmic Decision Making.. In *RoCHI*. 2–9.

[11] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. 2015. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1721–1730.

[12] Michael Chromik and Andreas Butz. 2021. Human-xai interaction: A review and design principles for explanation user interfaces. In *IFIP Conference on Human-Computer Interaction*. Springer, 619–640.

[13] Maliheh Ghajargar, Jeffrey Bardzell, Alison Smith Renner, Peter Gall Krogh, Kristina Höök, David Cuartielles, Laurens Boer, and Mikael Wiberg. 2021. From" explainable ai" to" graspable ai". In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–4.

[14] Elisa Giaccardi and Johan Redström. 2020. Technology and more-than-human design. *Design Issues* 36, 4 (2020), 33–44.

[15] Bryce Goodman and Seth Flaxman. 2017. European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine* 38, 3 (2017), 50–57.

[16] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)* 51, 5 (2018), 1–42.

[17] Maria Hartikainen, Kaisa Väänänen, Thomas Olsson, Anu Lehtiö, and Saara Ala-Luopa. 2022. Human-Centered AI Design in Reality: A Study of Developer Companies' Practices. In *Proceedings of NordiCHI 2022. (To appear)*. ACM.

[18] Trevor Hogan, Eva Hornecker, Simon Stusak, Yvonne Jansen, Jason Alexander, Andrew Vande Moere, Uta Hinrichs, and Kieran Nolan. 2016. Tangible Data, explorations in data physicalization. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. 753–756.

[19] Frederick Hohman. 2020. *Interactive Scalable Interfaces for Machine Learning Interpretability*. Ph.D. Dissertation. Georgia Institute of Technology.

[20] Eva Hornecker and Jacob Buur. 2006. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 437–446.

[21] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. 234–241.

[22] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and challenges for data physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3227–3236.

[23] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*. 139–146.

[24] Sébastien Kubicki, Yoann Lebrun, Sophie Lepreux, Emmanuel Adam, Christophe Kolski, and René Mandiau. 2013. Simulation in contexts involving an interactive table and tangible objects. *Simulation Modelling Practice and Theory* 31 (2013), 116–131.

[25] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell me more? The effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1–10.

[26] Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2016. Rationalizing Neural Predictions. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Austin, Texas, 107–117. https://doi.org/10.18653/v1/D16-1011

[27] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[28] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. 2020. Explainable ai: A review of machine learning interpretability methods. *Entropy* 23, 1 (2020), 18.

[29] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[30] Deborah Lupton. 2017. Feeling your data: Touch and making sense of personal digital data. *New Media & Society* 19, 10 (2017), 1599–1614.

[31] Tim Miller, Piers Howe, and Liz Sonenberg. 2017. Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences. *arXiv preprint arXiv:1712.00547* (2017).

[32] Thomas Olsson and Kaisa Väänänen. 2021. How does AI challenge design practice? *Interactions* 28, 4 (2021), 62–64.

[33] Alun Preece. 2018. Asking 'Why'in AI: Explainability of intelligent systems–perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management* 25, 2 (2018), 63–72.

[34] Majken K Rasmussen, Esben W Pedersen, Marianne G Petersen, and Kasper Hornbæk. 2012. Shape-changing interfaces: a review of the design space and open research questions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 735–744.

[35] Kay Römer and Svetlana Domnitcheva. 2002. Smart playing cards: A ubiquitous computing game. *Personal and Ubiquitous Computing* 6, 5 (2002), 371–377.

[36] Ben Shneiderman. 2020. Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 4 (2020), 1–31.

[37] Ben Shneiderman. 2020. Human-centered artificial intelligence: three fresh ideas. *AIS Transactions on Human-Computer Interaction* 12, 3 (2020), 109–124.

[38] Davide Spallazzo, Martina Sciannamè, and Mauro Ceconello. 2019. The domestic shape of AI: a reflection on virtual assistants. In *11th Proceedings of Design and Semantics of Form and Movement International Conference (DeSForM) MIT Boston*. 52–59.

[39] Martin Spindler, Christian Tominski, Heidrun Schumann, and Raimund Dachselt. 2010. Tangible views for information visualization. In *ACM International Conference on Interactive Tabletops and Surfaces*. 157–166.

[40] Faisal Taher, John Hardy, Abhijit Karnik, Christian Weichel, Yvonne Jansen, Kasper Hornbæk, and Jason Alexander. 2015. Exploring interactions with physically dynamic bar charts. In *Proceedings of the 33rd annual acm conference on human factors in computing systems*. 3237–3246.

[41] Sarah Tan, Matvey Soloviev, Giles Hooker, and Martin T Wells. 2020. Tree space prototypes: Another look at making tree ensembles interpretable. In *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference*. 23–34.

[42] Anna Vallgårda, Laurens Boer, Vasiliki Tsaknaki, and Dag Svanæs. 2016. Material programming: A design practice for computational composites. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. 1–10.

[43] Giulia Vilone and Luca Longo. 2021. Classification of explainable artificial intelligence methods through their output formats. *Machine Learning and Knowledge Extraction* 3, 3 (2021), 615–661.

[44] Mikael Wiberg. 2018. *The materiality of interaction: Notes on the materials of interaction design*. MIT press.

[45] Terri L Woods, Sarah Reed, Sherry Hsi, John A Woods, and Michael R Woods. 2016. Pilot study using the augmented reality sandbox to teach topographic maps and surficial processes in introductory geology labs. *Journal of Geoscience Education* 64, 3 (2016), 199–214.

[46] Wei Xu. 2019. Toward human-centered AI: a perspective from human-computer interaction. *interactions* 26, 4 (2019), 42–46.

[47] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

[48] Fabio Massimo Zanzotto. 2019. Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research* 64 (2019), 243–252.