# From "Explainable AI" to "Graspable AI"

Maliheh Ghajargar
maliheh.ghajargar@mau.se
Malmö University
Malmö, Sweden

Jeffrey Bardzell
Pennsylvania State University
State College, USA

Alison Smith Renner
Decisive Analytics Corporation
USA

Peter Gall Krogh
Aarhus University
Aarhus, Denmark

Kristina Höök
KTH
Stockholm, Sweden

David Cuartielles
Malmö University
Malmö, Sweden

Laurens Boer
ITU
Copenhagen, Denmark

Mikael Wiberg
Umeå University
Umeå, Sweden

## ABSTRACT

Since the advent of Artificial Intelligence (AI) and Machine Learning (ML), researchers have asked how intelligent computing systems could interact with and relate to their users and their surroundings, leading to debates around issues of biased AI systems, ML black-box, user trust, user's perception of control over the system, and system's transparency, to name a few. All of these issues are related to how humans interact with AI or ML systems, through an interface which uses different interaction modalities. Prior studies address these issues from a variety of perspectives, spanning from understanding and framing the problems through ethics and Science and Technology Studies (STS) perspectives to finding effective technical solutions to the problems. But what is shared among almost all those efforts is an assumption that if systems can explain the how and why of their predictions, people will have a better perception of control and therefore will trust such systems more, and even can correct their shortcomings. This research field has been called Explainable AI (XAI). In this studio, we take stock on prior efforts in this area; however, we focus on using Tangible and Embodied Interaction (TEI) as an interaction modality for understanding ML. We note that the affordances of physical forms and their behaviors potentially can not only contribute to the explainability of ML systems, but also can contribute to an open environment for criticism. This studio seeks to both critique explainable ML terminology and to map the opportunities that TEI can offer to the HCI for designing more sustainable, graspable and just intelligent systems.

## CCS CONCEPTS

• **Human-centered computing** → Interaction paradigms; *Human computer interaction (HCI)*; Interactive systems and tools;

## KEYWORDS

Explainable AI, XAI, Machine Learning, Artificial Intelligence, Tangible Embodied Interaction, TEI, Interaction Design

**ACM Reference Format:**
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 *Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21), February 14–17, 2021, Salzburg, Austria.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3430524.3442704

## 1 INTRODUCTION

Machine Learning (ML) and Artificial Intelligence (AI) are increasingly used to classify data and detect patterns in large and complex data sets, which allows them to make recommendations, detect anomalies, and automate physical systems, among others. From national defense to business to personal entertainment, AI/ML is ubiquitous and used in different areas (e.g., smart homes, smart cities, and autonomous cars). As these intelligent systems affect the everyday lives of people, cities and organizations, concerns such as trust, control, transparency and explainability have been raised [21, 24, 33]. However, the question of explanation as a required form of dialogue between people and computing systems to both build (appropriate) trust and to build more effective and efficient systems is not new [8, 9]. While intelligent systems may appear to operate accurately or as expected, people need to know how and why these systems make decisions, particularly to ensure they can generalize and are operating fairly, or without bias. To that end, the field of Explainable AI (XAI) explores mechanisms for explaining or exposing intelligent systems' inner-workings or outputs to support understanding and increase trust [26]. Such research includes both global explanations (i.e., explaining models as a whole) [7] and local explanations (i.e., explaining—or rationalizing—individual predictions) [23]. Increasing end users' understanding of intelligent systems has another benefit: these users can better influence or improve systems, as they are more aware of how and when they err, and therefore how and what to fix [22]. Interactive (or human-in-the-loop) ML supports rapid, iterative user feedback or guidance to improve or adapt models [2, 3]. Interactive ML techniques have been applied to allow non-expert users to guide AI/ML

*TEI '21, February 14–17, 2021, Salzburg, Austria*
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8213-7/21/02.
https://doi.org/10.1145/3430524.3442704

algorithms through user-model interaction, which in return increases model transparency and interpretability [27, 31, 32]. Thus, these two forms of interactivity (explanations and control) go hand in hand to improve understanding, experience, and system performance. Intelligent systems, which may support explanations, feedback, or both, facilitate the interaction of non-expert users with systems for the purpose of improving the learning process, bettering system and human-machine task performance, and increasing system transparency. This is an approach that can benefit from designing interactions based not only on user needs and perceptions, but also on ways to empower them to grasp the complexities and dynamics of AI systems [10, 15, 25]. Importantly, these systems often rely on interaction modalities to interact with people —visual, audible, tactual, etc. One of the interaction modalities that has been less explored in this area, is Tangible Embodied Interaction (TEI).

**Tangible Embodied Interaction (TEI).** Tangible User Interfaces (TUI) are a way to grasp and manipulate computing systems by merging systems with everyday physical objects and spaces [17, 18]. TUIs bridge the gap between the virtual and the real world, hence between bits and atoms. This concept of TEI has evolved over time, to include not only the controlling and manipulating aspects of computing systems, but also the representation and form giving of the data—by not only visualizing but also physicalizing it (i.e., Data Physicalization [12, 19]). Data Physicalization benefits the user interaction with data, by leveraging natural human perceptual and cognitive skills, and it relates back to the history of written manuscripts using tokens and tangibles [29]. Over the years the TEI field has evolved, and related concepts have emerged: Embodied Interaction concerns how computing systems and our interaction with them can change our perception of the physical reality [10], a material-centered approach to Interaction Design calls to emphasize the material manifestations of the interaction [34–36], and Soma Design, a body-centric approach to interaction design, focuses on a holistic approach to interaction design, incorporating bodies and movements into the design and use [16]. Further, the forms and form giving practices are prominent mediums in design processes to convey meaning for artifacts or how artifacts can be used through their perceived affordances [20, 28, 30]. Aesthetic criticism informed by aesthetic philosophy considers the form of an artifact as an unifying principle (sometimes known as "significant form" after Clive Bell [6]) that composes the work's disparate parts into a whole that is replete with meaning [4, 5]. Building upon the variety of perspectives on TEI and physicality, we aim to map the opportunities they can offer to XAI, or as we call it here, Graspable AI.

## 2 GRASPABLE AI

This overall objective of this studio is to map the opportunities that TEI in its broadest sense, including TUI, embodied interaction, body-centric and forms and materiality of the interaction with computing systems can offer to human-in-the-loop and XAI systems. We use the phrase "Graspable AI," which deliberately plays with two senses of the word "to grasp," one referring to taking something into one's hand, and the other when the mind "grasps" an idea. To this aim, we first seek to challenge the terminology of XAI. After coming to an understanding that physical and tangible interaction can offer more spaces for designing a more understandable and transparent ML/AI, we realized the terminology may fall short in that regard. So, we use the term Graspable AI as a way to approach XAI through Tangible and Embodied Interaction perspective, since it refers to something that is not only understandable and perceivable, but it also is coherent and accessible as a unified form, and which can be held our bodies (e.g. by hands). The term Graspable inherently conveys the meaning of being understandable intellectually, meaningfully and physically. More specifically this studio will focus on three challenges of Graspable AI, considering the entire process of AI/ML from classification of the data to the explanation of decisions to the users [33]: forms, behaviors and interaction criticism of Graspable AI. These themes serve to articulate and to cluster the contributions of TEI to the XAI space.

### 2.1 Graspable Forms

Forms are the outcome of the design process and are the result of the pragmatic synthesis of multiple factors (e.g. context, user needs, materials, etc.) [1]. We frame graspable forms as synthesized and unified wholes capable of conveying a meaning, a message or a state (classifying, learning or explaining) manifested in physical forms. Further graspable forms are often self-explanatory and intuitively understandable and relatable. We aim to explore graspable forms of ML/AI process by mapping different ways the unified form of the ML/AI models within intelligent systems can become graspable through TEI.

### 2.2 Graspable Behaviors

One of the salient characteristics of ML/AI algorithms that distinguishes them from other kinds of algorithms (or programs) is that they classify data, detect patterns and learns from them. Therefore, their outcome depends on what the algorithms learn over time from data (e.g., from human behaviors, environment, or their own decisions). In a learning system, we expect that the unified wholes or forms changes over time, as such, the outcomes or the decisions become very complex and sometimes not fully understandable by humans. We aim to unpack this area by exploring how temporality influences the tangible forms of algorithms? What are the possible behaviors that make the AI/ML models more graspable? And, what are the suitable graspable and familiar metaphors (e.g., growing, etc.)?

### 2.3 Graspable Interaction Criticism

Inspired by interaction criticism [4, 5], which suggests that the form of an artifact is one of four primary considerations, along with intentionality, the individual experience, and the sociocultural context in which it was produced/consumed, we believe the Graspable AI can provide opportunities for an open and more democratic environment for criticism. Hence physical and tangible forms of ML/AI can contribute to designing better, equal and more fair intelligent systems by facilitating the participation and mutual understanding between humans and AI/ML. Further, Graspable interaction criticism seeks to reveal the relations between physical and intellectual dimensions of Graspable AI. It seeks to raise the question of whether the interactions can be made intentionally difficult to grasp (physically) because they are intellectually hard to grasp for humans, and by that increasing the transparency of the AI/ML systems.

## 3 STUDIO STRUCTURE

In order to facilitate the participation of larger number of people from different continents and time-zones, this studio will be organized in two 4-hour rounds (Table 1). The studio will be a combination of presentations, discussions and analytical activities in groups of 3 to 5 participants.

### 3.1 Kick-off short presentations

We will kick off the studio by presenting the topic and activities of the studio. Then we ask all participants to introduce themselves and their position paper or interactive object demo or a physical object they brought to the studio whose graspability is relevant to the studio topic in a 10-minute presentation. We will ask each participant to end their presentation by stating that how their positions contribute to explainability in AI and ML models and how TEI can be included in a future development of the position.

### 3.2 Making Graspable AI

Participants will form groups based on their submissions, and its relation to the three areas of Graspable forms, Graspable behaviors and Graspable interaction criticism. If the submission is related to a specific application area, and does not fit perfectly within just one of the above-mentioned categories, then it will be considered and situated within the closest thematic category. Each group will then go through an ideation process using digital cards inspired by two methodologies of Inspiration Cards [11] and The Card Brainstorming Game [13, 14]. Overall, it consists of four themes and related concepts, which are based on Hornecker and Buur's Tangible Interaction theoretical framework. The themes are (1) Tangible Manipulation, (2) Spatial Interaction, (3) Embodied Facilitation, (3) Expressive Representation, which are presented with 'provocative questions' that help participants to concretize the concepts related to the TEI explainable ML.

## 4 DISCUSSIONS AND PRESENTATIONS

A discussion session then will follow, during which participants will discuss the challenges and opportunities of designing TEI for XAI in general (Graspable forms, behaviors and interaction criticism), and the effectiveness and usefulness of the ideation cards for designing Graspable AI interfaces in specific. (e.g. what are the challenges of designing Tangible Interactions when the computer is able to learn and interactions with the user are aimed to enhance the learning outcome? What are the social-technical implications/challenges? etc. The groups will then summarize their ideas and analysis and will present in 5-minute presentation.

## 5 PLANS FOR DISSEMINATION OF STUDIO RESULTS

The studio will explore the state of the art and possible future developments concerning the Tangible Embodied Interaction perspectives of Explainable AI, including what is currently known, where research is ongoing, and where more research is needed. These results will be submitted as an article to a magazine (e.g., ACM interactions) to inspire and invite more HCI and design researchers to engage this research space. In addition, the website of the studio will be updated with the outcome of the workshop and its future developments. We will also invite interested participants to sign up to a mailing list in order to keep updated on the future developments and activities related to the Graspable AI topic.

**Table 1: Studio Schedule**

| Time | Activity-virtual participation |
|---|---|
| 08:00 – 08:15 | Introduction to the Studio (Zoom) |
| 08:15 – 09:45 | Participants presentations (Zoom) |
| 09:45 – 10:15 | Group creation based on the Studio themes (Zoom and Miro) / Coffee Break |
| 10:15 – 11:00 | Ideation and discussion in groups (Zoom and Miro) |
| 11:00 – 11:45 | Reflection and presentation (Zoom and Miro) |
| 11:45 – 12:00 | Conclusions (Zoom) |
| 2nd round | |
| 14:00 – 14:15 | Introduction to the Studio (Zoom) |
| 14:15 – 15:45 | Participants presentations (Zoom) |
| 15:45 – 16:15 | Group creation based on the Studio themes (Zoom and Miro) / Coffee Break |
| 16:15 – 17:00 | Ideation and discussion in groups (Zoom and Miro) |
| 17:00 – 17:45 | Reflection and presentation (Zoom and Miro) |
| 17:45 – 18:00 | Conclusions (Zoom) |

## REFERENCES

[1] Christopher Alexander. 1964. *Notes on the Synthesis of Form.* Harvard University Press, Cambridge, Mass.

[2] 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 (Dec. 2014), 105–120. https://doi.org/10.1609/aimag.v35i4.2513 Number: 4.

[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* Association for Computing Machinery, Glasgow, Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300233

[4] Jeffrey Bardzell. 2009. Interaction criticism and aesthetics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09).* Association for Computing Machinery, New York, NY, USA, 2357–2366. https://doi.org/10.1145/1518701.1519063

[5] Jeffrey Bardzell. 2011. Interaction criticism: An introduction to the practice. *Interacting with Computers* 23, 6 (Nov. 2011), 604–621. https://doi.org/10.1016/j.intcom.2011.07.001

[6] Clive Bell. 1949. *Art.* Chatto & Windus.

[7] 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 (KDD '15).* Association for Computing Machinery, New York, NY, USA, 1721–1730. https://doi.org/10.1145/2783258.2788613

[8] Alison Cawsey. 1989. Explanatory dialogues. *Interacting with Computers* 1, 1 (April 1989), 69–92. https://doi.org/10.1016/0953-5438(89)90008-8

[9] Alison Cawsey. 2003. *Explanation and Interaction (ACL-MIT Series in Natural Language Processing): The Computer Generation of Explanatory Dialogues.* A Bradford Book, Place of publication not identified.

[10] Elisa Giaccardi and Johan Redström. 2020. Technology and More-Than-Human Design. *Design Issues* 36, 4 (Sept. 2020), 33–44. https://doi.org/10.1162/desi_a_00612 Publisher: MIT Press.

[11] Kim Halskov and Peter Dalsgård. 2006. Inspiration card workshops. In *Proceedings of the 6th conference on Designing Interactive systems (DIS '06)*. Association for Computing Machinery, New York, NY, USA, 2–11. https://doi.org/10.1145/1142405.1142409

[12] 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 (TEI '16)*. Association for Computing Machinery, New York, NY, USA, 753–756. https://doi.org/10.1145/2839462.2854112

[13] Eva Hornecker. 2010. Creative idea exploration within the structure of a guiding framework: the card brainstorming game. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction (TEI '10)*. Association for Computing Machinery, New York, NY, USA, 101–108. https://doi.org/10.1145/1709886.1709905

[14] 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 (CHI '06)*. Association for Computing Machinery, Montréal, Québec, Canada, 437–446. https://doi.org/10.1145/1124772.1124838

[15] K. Höök. 2000. Steps to take before intelligent user interfaces become real. *Interacting with Computers* 12, 4 (Feb. 2000), 409–426. https://doi.org/10.1016/S0953-5438(99)00006-5

[16] Kristina Höök. 2018. *Designing with the Body: Somaesthetic Interaction Design*. The MIT Press, Cambridge, Massachusetts.

[17] Hiroshi Ishii, Dávid Lakatos, Leonardo Bonanni, and Jean-Baptiste Labrune. 2012. Radical atoms: beyond tangible bits, toward transformable materials. *Interactions* 19, 1 (Jan. 2012), 38–51. https://doi.org/10.1145/2065327.2065337

[18] 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 (CHI '97)*. Association for Computing Machinery, New York, NY, USA, 234–241. https://doi.org/10.1145/258549.258715

[19] 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 (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3227–3236. https://doi.org/10.1145/2702123.2702180

[20] Peter Gall Krogh and Ilpo Koskinen. 2020. *Drifting by Intention: Four Epistemic Traditions from within Constructive Design Research* (1st ed. 2020 edition ed.). Springer.

[21] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*. Association for Computing Machinery, Atlanta, Georgia, USA, 126–137. https://doi.org/10.1145/2678025.2701399

[22] 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 (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2207676.2207678

[23] 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

[24] Brian Lim, Advait Sarkar, Alison Smith-Renner, and Simone Stumpf. 2019. ExSS: explainable smart systems 2019. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion (IUI '19)*. Association for Computing Machinery, Marina del Ray, California, 125–126. https://doi.org/10.1145/3308557.3313112

[25] 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:1712.00547 [cs]* (Dec. 2017). http://arxiv.org/abs/1712.00547 arXiv: 1712.00547.

[26] 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. https://doi.org/10.1002/isaf.1422 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/isaf.1422.

[27] Alison Marie Renner. 2020. Designing for the Human in the Loop: Transparency and Control in Interactive Machine Learning. (2020). https://doi.org/10.13016/ze3u-bfbq Accepted: 2020-07-08T05:34:15Z.

[28] Marco C. Rozendaal, Maliheh Ghajargar, Gert Pasman, and Mikael Wiberg. 2018. Giving Form to Smart Objects: Exploring Intelligence as an Interaction Design Material. In *New Directions in Third Wave Human-Computer Interaction: Volume 1 - Technologies*, Michael Filimowicz and Veronika Tzankova (Eds.). Springer International Publishing, Cham, 25–42. https://doi.org/10.1007/978-3-319-73356-2_3

[29] Denise Schmandt-Besserat. 1997. *How Writing Came About* (abridged edition ed.). University of Texas Press, Austin.

[30] Gerda Smets, Kees Overbeeke, and William Gaver. 1994. Form-giving: Expressing the Nonobvious. In *Conference Companion on Human Factors in Computing Systems (CHI '94)*. ACM, New York, NY, USA, 204–. https://doi.org/10.1145/259963.260289 event-place: Boston, Massachusetts, USA.

[31] Alison Smith, Varun Kumar, Jordan Boyd-Graber, Kevin Seppi, and Leah Findlater. 2018. Closing the Loop: User-Centered Design and Evaluation of a Human-in-the-Loop Topic Modeling System. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. Association for Computing Machinery, Tokyo, Japan, 293–304. https://doi.org/10.1145/3172944.3172965

[32] Alison Smith-Renner, Varun Kumar, Jordan Boyd-Graber, Kevin Seppi, and Leah Findlater. 2020. Digging into user control: perceptions of adherence and instability in transparent models. In *Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI '20)*. Association for Computing Machinery, Cagliari, Italy, 519–530. https://doi.org/10.1145/3377325.3377491

[33] Stefano Teso and Kristian Kersting. 2019. Explanatory Interactive Machine Learning. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)*. Association for Computing Machinery, Honolulu, HI, USA, 239–245. https://doi.org/10.1145/3306618.3314293

[34] 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 (NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2971485.2971554

[35] Anna Vallgårda and Johan Redström. 2007. Computational Composites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 513–522. https://doi.org/10.1145/1240624.1240706 event-place: San Jose, California, USA.

[36] Mikael Wiberg. 2018. *The Materiality of Interaction: Notes on the Materials of Interaction Design* (illustrated edition ed.). The MIT Press, Cambridge, Massachusetts.