From ‘Explainable AI’ to ‘Graspable AI’

MALIHEH GHAJARGAR, Malmö University, Sweden

JEFFREY BARDZELL, Pennsylvania State University, USA

ALISON RENNER, ML Visualization Lab for Decisive Analytics Corporation, USA

KRISITINA HÖÖK, KTH, Sweden

PETER GALL KROGH, Aarhus University, Denmark

DAVID CUARTIELLES, Malmö University, Sweden

LAURENS BOER, ITU, Denmark

MIKAEL WIBERG, Umeå University, Sweden

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 proposal, 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. Our proposal 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 • Insert your first CCS term here • Insert your second CCS term here • Insert your third CCS term here

Additional Keywords and Phrases: Insert comma delimited author-supplied keyword list, Keyword number 2, Keyword number 3, Keyword number 4

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 BACKGROUND

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 [22, 25, 34, 34]. 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 [27]. Such research includes both global explanations (i.e., explaining models as a whole) [7] and local explanations (i.e., explaining—or rationalizing—individual predictions) [24]. 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 [23]. 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 algorithms through user-model interaction, which in return increases model transparency and interpretability [28, 32, 33]. 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 [11, 15, 26]. 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).

1.1 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 [18, 19]. 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 [13, 20]). 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 [30].

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 [35–37], 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 [14].

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 [21, 29, 31]. 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 for, 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 [34]: 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 WORKSHOP STRUCTURE

We propose a one-day, 4-hour long studio. The studio will be a combination of presentations, discussions and analytical activities in groups of 3 to 5 people.

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 [12] and The Card Brainstorming Game [16, 17]. 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.

3.3 Discussions and presentation

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.

Table 1. Preliminary studio schedule (CET)

Time	Activity – virtual participation
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] Alexander, C. 1964. *Notes on the Synthesis of Form*. Harvard University Press.

[2] Amershi, S., Cakmak, M., Knox, W.B. and Kulesza, T. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine*. 35, 4 (Dec. 2014), 105–120. DOI:https://doi.org/10.1609/aimag.v35i4.2513.

[3] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R. and Horvitz, E. 2019. Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk, May 2019), 1–13.

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

[5] Bardzell, J. 2009. Interaction criticism and aesthetics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2009), 2357–2366.

[6] Bell, C. 1949. *Art*. Chatto & Windus.

[7] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M. and Elhadad, N. 2015. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, Aug. 2015), 1721–1730.

[8] Cawsey, A. 2003. *Explanation and Interaction (ACL-MIT Series in Natural Language Processing): The Computer Generation of Explanatory Dialogues*. A Bradford Book.

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

[10] Dourish, P. 2004. *Where the Action Is: The Foundations of Embodied Interaction*. The MIT Press.

[11] Giaccardi, E. and Redström, J. 2020. Technology and More-Than-Human Design. *Design Issues*. 36, 4 (Sep. 2020), 33–44. DOI:https://doi.org/10.1162/desi_a_00612.

[12] Halskov, K. and Dalsgård, P. 2006. Inspiration card workshops. *Proceedings of the 6th conference on Designing Interactive systems* (New York, NY, USA, Jun. 2006), 2–11.

[13] Hogan, T., Hornecker, E., Stusak, S., Jansen, Y., Alexander, J., Moere, A.V., Hinrichs, U. and Nolan, K. 2016. Tangible Data, explorations in data physicalization. *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (New York, NY, USA, Feb. 2016), 753–756.

[14] Höök, K. 2018. *Designing with the Body: Somaesthetic Interaction Design*. The MIT Press.

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

[16] Hornecker, E. 2010. Creative idea exploration within the structure of a guiding framework: the card brainstorming game. *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction* (New York, NY, USA, Jan. 2010), 101–108.

[17] Hornecker, E. and Buur, J. 2006. Getting a grip on tangible interaction: a framework on physical space and social interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada, Apr. 2006), 437–446.

[18] Ishii, H., Lakatos, D., Bonanni, L. and Labrune, J.-B. 2012. Radical atoms: beyond tangible bits, toward transformable materials. *Interactions*. 19, 1 (Jan. 2012), 38–51. DOI:https://doi.org/10.1145/2065327.2065337.

[19] Ishii, H. and Ullmer, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (New York, NY, USA, Mar. 1997), 234–241.

[20] Jansen, Y., Dragicevic, P., Isenberg, P., Alexander, J., Karnik, A., Kildal, J., Subramanian, S. and Hornbæk, K. 2015. Opportunities and Challenges for Data Physicalization. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2015), 3227–3236.

[21] Krogh, P.G. and Koskinen, I. 2020. *Drifting by Intention: Four Epistemic Traditions from within Constructive Design Research*. Springer.

[22] Kulesza, T., Burnett, M., Wong, W.-K. and Stumpf, S. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA, Mar. 2015), 126–137.

[23] Kulesza, T., Stumpf, S., Burnett, M. and Kwan, I. 2012. Tell me more? the effects of mental model soundness on personalizing an intelligent agent. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2012), 1–10.

[24] Lei, T., Barzilay, R. and Jaakkola, T. 2016. Rationalizing Neural Predictions. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (Austin, Texas, Nov. 2016), 107–117.

[25] Lim, B., Sarkar, A., Smith-Renner, A. and Stumpf, S. 2019. ExSS: explainable smart systems 2019. *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion* (Marina del Ray, California, Mar. 2019), 125–126.

[26] Miller, T., Howe, P. and Sonenberg, L. 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).

[27] Preece, A. 2018. Asking 'Why' in AI: Explainability of intelligent systems – perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management*. 25, 2 (2018), 63–72. DOI:https://doi.org/10.1002/isaf.1422.

[28] Renner, A.M. 2020. Designing for the Human in the Loop: Transparency and Control in Interactive Machine Learning. (2020). DOI:https://doi.org/10.13016/ze3u-bfbq.

[29] Rozendaal, M.C., Ghajargar, M., Pasman, G. and Wiberg, M. 2018. Giving Form to Smart Objects: Exploring Intelligence as an Interaction Design Material. *New Directions in Third Wave Human-Computer Interaction: Volume 1 - Technologies*. M. Filimowicz and V. Tzankova, eds. Springer International Publishing. 25–42.

[30] Schmandt-Besserat, D. 1997. *How Writing Came About*. University of Texas Press.

[31] Smets, G., Overbeeke, K. and Gaver, W. 1994. Form-giving: Expressing the Nonobvious. *Conference Companion on Human Factors in Computing Systems* (New York, NY, USA, 1994), 204–.

[32] Smith, A., Kumar, V., Boyd-Graber, J., Seppi, K. and Findlater, L. 2018. Closing the Loop: User-Centered Design and Evaluation of a Human-in-the-Loop Topic Modeling System. *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan, Mar. 2018), 293–304.

[33] Smith-Renner, A., Kumar, V., Boyd-Graber, J., Seppi, K. and Findlater, L. 2020. Digging into user control: perceptions of adherence and instability in transparent models. *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy, Mar. 2020), 519–530.

[34] Teso, S. and Kersting, K. 2019. Explanatory Interactive Machine Learning. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA, Jan. 2019), 239–245.

[35] Vallgårda, A., Boer, L., Tsaknaki, V. and Svanæs, D. 2016. Material Programming: a Design Practice for Computational Composites. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (New York, NY, USA, Oct. 2016), 1–10.

[36] Vallgårda, A. and Redström, J. 2007. Computational Composites. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2007), 513–522.

[37] Wiberg, M. 2018. *The Materiality of Interaction: Notes on the Materials of Interaction Design*. The MIT Press.