Tangible LLMs: Tangible Sense-Making For Trustworthy Large Language Models

Leonardo Angelini, Goda Klumbytė, Maxime Daniel, Nadine Couture, Elena Mugellini, Claude Draude

Tangible LLMs: Tangible Sense-Making For Trustworthy Large Language Models

Leonardo Angelini
University of Applied Sciences and Arts Western Switzerland
HumanTech Institute
Fribourg, Switzerland
leonardo.angelini@hefr.ch

Goda Klumbyte
Participatory IT Design
University of Kassel
Kassel, Germany
goda.klumbyte@uni-kassel.de

Maxime Daniel
ESTIA-Recherche
ESTIA INSTITUTE OF TECHNOLOGY
Bidart, France
maxaxeldaniel@gmail.com

Nadine Couture
ESTIA-Recherche
ESTIA INSTITUTE OF TECHNOLOGY
Bidart, France
n.couture@estia.fr

Elena Mugellini*
HumanTech Institute
University of Applied Sciences and Arts Western Switzerland (HES-SO)
Fribourg, Switzerland
elena.mugellini@hes-so.ch

Claude Draude
Department of Participatory IT Design
University of Kassel
Kassel, Hessen, Germany
claude.draude@uni-kassel.de

Abstract

The goal of this studio is to explore and discuss the potential benefits and the open challenges of giving a physical form to Artificial Intelligence (AI), towards the definition of tangible, or graspable AI. In this studio, we want to carry out a hands-on exploration of the potential of tangible interaction to help people grasp how transformer-based AI models, such as Large Language Models (LLMs), make predictions. Starting from a basic understanding of the core mechanisms of these models – input/output embedding, positional encoding, self-attention, and multi-head attention, we will explore how tangible interaction properties, such as data physicalization and tangible manipulation, may help designing graspable interfaces for facilitating the understanding and control of LLMs. Fostering AI transparency and human agency, we believe that tangible interaction may help increase the trustworthiness of Large Language Models. Through brainstorming and prototyping activities, we aim at identifying effective tangible representations and manipulation techniques that will help users grasp and trust LLMs.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**.

Keywords

Tangible Embodied Interaction, Explainable AI, Large Language Model

ACM Reference Format:
Leonardo Angelini, Goda Klumbyte, Maxime Daniel, Nadine Couture, Elena Mugellini, and Claude Draude. 2025. Tangible LLMs: Tangible Sense-Making For Trustworthy Large Language Models. In *Nineteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '25), March 04–07, 2025, Bordeaux / Talence, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3689050.3708338

TEI '25, March 04–07, 2025, Bordeaux / Talence, France

ACM ISBN 979-8-4007-1197-8/25/03
https://doi.org/10.1145/3689050.3708338

1 Introduction

In the last decade, we witnessed a tremendous growth of AI models that we use today in our everyday lives. A first breakthrough concerned classification tasks, since the Convolutional Neural Network (CNN) AlexNet that won the ImageNet challenge in 2012. A second breakthrough concerns generative AI, with transformer-based architectures making enormous progress after the introduction of auto-encoders and generative adversarial networks. These recent deep learning models are rapidly spreading in our society, driving innovation in various domains, and reshaping the way we approach complex problems. However, they are inherently complex and opaque. Explainable AI (xAI) techniques for explaining complex deep learning models have emerged in the scientific literature. However, like the models themselves, existing xAI techniques remain often complex, making them difficult to grasp for humans, even experts. Although many xAI techniques have emerged for LLMs [20], including several visual representations, most of them offer limited interactivity to the users, limiting explorations possibilities and hindering human-agency. The goal of this studio is to explore the use of tangible and spatial interaction that can contribute to making LLMs more human-understandable and trustworthy. Through a series of collaborative, hands-on activities, participants will engage with various sensory modalities – tactile, auditory, visual, and even olfactory – to investigate their potential in deepening their comprehension of how LLMs make predictions and generate outputs.

To sum up, this studio invites participants to re-imagine how AI systems can be explained and understood, moving away from purely abstract, visual methods and embracing a multi-sensory, tangible and experiential approach. Through interdisciplinary collaboration, the studio aims to open new paths for making AI more accessible, interpretable, and embedded in human experience. This, we believe, contributes to discussions on sustainability as a broader framework that entails careful material engagement as well as positions social sustainability through inclusion of diverse users as one of the core concerns.

1.1 Grounding in theory

The first explorations of Graspable AI were carried out in a studio at TEI'21, where participants investigated whether the affordances of physical forms could contribute to the explainability of intelligent systems of our everyday life, such as a movie recommender system, a self-tracking device or a robotic vacuum cleaner [9]. Graspable explanations for recommender systems were further explored in a followup studio at TEI'22 [8]. However, these two workshops considered AI mainly as a black box, designing high-level explanations that were agnostic of the model architecture. On the other hand, Colley et al. [5] proposed to leverage existing XAI techniques, typically based on Graphical User Interfaces, in order to design tangible AI explanations. The authors posited that existing XAI techniques, such as feature relevance, local explanations and decision rules may be easier to understand and to control when embodied in a tangible interface. This should increase AI transparency, the understandability of AI decision making, and should enable new forms of human-in-the-loop AI. Besides explaining decisions, model exploration is often important to better understand and trust AI. In a previous workshop at ETIS'22 [2], some of the organizers of this studio investigated how tangible interaction could facilitate the understanding of Convolutional Neural Networks (CNNs) through the physicalization of the model architecture and of critical parameters of the algorithm. Participants' design concepts showed that a physical representation of the model architecture could encourage exploration and help understanding the structure and functioning of deep learning algorithms, facilitating group discussion and giving multiple controls for shared interaction.

In this studio, we aim at bridging tangible and interaction properties and AI design principles, to design tangible explanations for Large Language Models.

Similarly to Colley et al [5], we will ground our studio on the Hornecker and Buur framework, exploring not only tangible manipulation and physical data representation but also embodied and spatial interactions in real spaces [13]. Embodied and spatial interaction is a broad umbrella term for interactions that exploit the role of physical manipulation and perception and movement in 3D space, including the physicality of our own bodies, the materiality of objects, the physical world in general, and the physicality of space [12]. Embodied and spatial interaction move the interaction from a computer screen to the world leveraging innate human abilities such as physical manipulation and spatial reasoning. Spatial reasoning, the human ability to understand, reason about, and manipulate objects in three-dimensional space, has been shown to be helpful not only in science, technology, engineering and mathematics (STEM) fields [14, 18], but also in social sciences [10]. In particular, interactivity plays an important role in understanding complex problems. Liang and Sedig [16] showed that providing learners with multiple interactions with 3D mathematical visualization was increasing the learner's performance in spatial thinking. Tangible interfaces could therefore foster the interactive explorations of LLM architectures and decision explanations, facilitating the understanding of LLM functioning.

At the same time, during the workshop, we will leverage the AI requirements proposed by the European Commission [1] to reflect on the desired characteristics of trustworthy AI. The proposed guidelines will be helpful to align the studio participants' ideas to the strategic goals of the European Commisions towards the development of a fair and sustainable AI. In particular we believe that tangible interaction could be particularly useful for meeting the following AI requirements: Human agency and oversight (Requirement 1), Transparency (Requirement 4), Diversity, non-discrimination and fairness (Requirement 5), Accountability (Requirement 7). Furthermore, through leveraging human embodiment and sensorium towards design of explainable AI, this workshop also contributes to the field of Human-Centred Explainable AI (HCXAI) [7], which has been gathering pace in the last years.

The first goal of this studio is giving a glimpse of LLM architectures and their underlying mechanisms such as attention, embeddings and latent space, to foster discussion on the possible role of tangible and embodied interaction as a mediator for trustworthy AI. We will leverage existing visual explanations and representations of LLMs to allow studio participants to explore and get a grasp of the most important concepts that are behind LLMs, such as the attention mechanisms, the input embeddings and the latent space. We will then ask participants to identify potential use-cases where tangible interaction could be particularly beneficial for exploring and explaining a large language model. Participants having an initial idea before the workshop will have the opportunity to pitch their use-case, that will eventually be selected for the following brainstorming and paper prototyping phases. During the brainstorming phase, other potential use-cases will be identified while reflecting on different research questions:

- How can a LLM be materialized?
- How can we physically explore the functioning (input and output space, embeddings, attention mechanism) of a LLM algorithm?
- How can we leverage tangible interaction properties to foster AI trustworthiness?

After selecting one use-case per group, participants will work on refining the initial ideas, trying to materialize the proposed tangible interfaces through rapid prototyping, discussing implementation challenges (such as the difficulty to physically represent models that are complex and large by definitions, the large amount of time required for training such networks and therefore to show effects of model updates on the outputs) and ideas to overcome these challenges. In both brainstorming and prototyping phases, design cards will be used to reflect on tangible interaction properties and trustworthy AI principles, and discover links between tangible properties and AI principles. The final discussions around the design concepts and paper prototype explored, will help to further explore how tangible interaction could help building trust in trustworthy LLMs. Future activities and collaboration opportunities in this field will be discussed at the end of the studio. Indeed, the overarching goal of this studio is to foster and sustain an interdisciplinary research community at the intersection of tangible interaction and

AI. To this purpose, the invitation to this studio will be extended to researchers in AI.

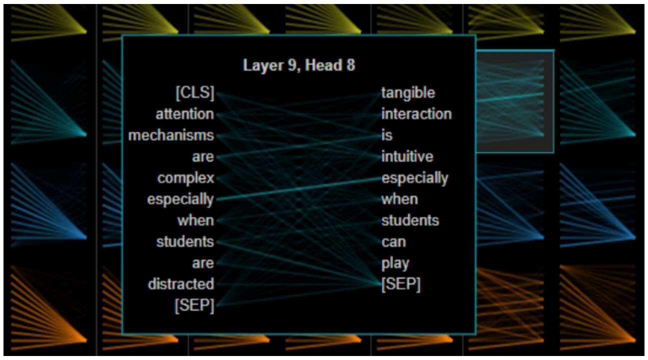

Figure 1: Visualization of the attention for a specific layer through the BertViz tool [17].

2 Supporting documents/figures/materials

During the workshop we will provide high-level explanations of some concepts that are at the foundations of large language models, such as input embeddings, the attention mechanisms and the latent space. To this purpose, during the workshop, we will allocate time to first explain these foundations, and second to test and explore existing visualization tools that have been developed to provide graphical representation of transformer-based AI. We will explore, for example, BertViz [17] for understanding the attention mechanism (see Figure 1); Bbycroft [4] to have a 3D overview of a LLM architecture; and LatentScope [15] to explore how embeddings are organized in the latent space.

During the brainstorming and prototyping phases we will provide a new design card set that will be conceived on purpose for exploring tangible interaction properties (see Figure 2) in relation to the requirement principles of large language models. The design cards used in the studio will be based on existing cardssets, i.e., Hornecker's cards set for design Tangibles [11], and Angelini et al.'s cards set for design Tangible Internet of Things [3] for tangible interaction properties and Wang et al. [19] cards set for ethics in AI and the EU guidelines requirements for the AI principles [6].

Furthermore, participants will be provided with a set of materials supporting various sensory experiences. Tactile materials will include:

- *Fabrics*: Including a variety of textures (e.g., silk, felt, cotton, mesh, wool) to convey different states of AI behavior or model dynamics.
- *Clay and Play-Dough*: For molding data structures, highlighting concepts like clustering or data transformations.
- *Magnetic tiles, blocks, Lego*: To build and break apart models, demonstrating concepts like feature attribution or network layers.
- *Foam or soft materials*: Representing soft boundaries in AI models or data points that are on the threshold.
- *String, yarn, wire*: For mapping out relationships between different variables, showing causal links, or weaving network connections.

Materials to support movement-based experiences will include:

- *Elastic bands or springs*: To symbolize relationships between input features and their tension or influence on outputs.
- *Choreography cards, ribbons, scarves (or else)*: To encourage participants to create "data dances," using cards or ribbons to signify how data points move through a model.
- *Body maps or space layouts*: Using large sheets or physical boundaries to guide participants to map AI structures through spatial organization.

3 Learning Goals

The aim of this studio is to provide participants with high-level understanding of fundamental concepts of transformer-based LLM as well as actionable knowledge on tangible interaction design for grasping and trusting LLM through hands-on explorations as well as brainstorming and prototyping activities. We aim to further ground this by having a critical interdisciplinary discussion on the interplay between tangible interaction properties, AI transparency and human agency and trustworthiness of Large Language Models. Therefore the studio will allow participants to:

- Get a grasp of large language models, thanks to model explorations and physical prototyping.
- Identify links between tangible interaction properties and trustworthy AI requirements through group brainstorming and paper prototyping.
- Explore and reflect on design ideas that leverage tangible interaction properties for making large language models more graspable.

Ultimately the studio aims to promote discussions on the emerging field of human-centered XAI and gain valuable insights that can drive the field forward.

4 Schedule

- **9:00 AM - 9:10 AM: Icebreaker**
  - Brief tactile, auditory, visual or embodied immersion exercise to heighten sensory awareness and set the tone for the studio.
- **9:10 AM - 9:20 AM: Studio Presentation**
  - Explanation of studio goals, structure, and expected outcomes.
- **9:20 AM - 9:30 AM: Participant Introductions**
  - Brief introduction (name, affiliation, research area, why participate in this studio) by each participant.
- **9:30 AM - 10:00 AM: Provocation Talk**
  - Brief talk on the foundational mechanisms of large language models (LLMs) and the limitations of existing explainability techniques. Current techniques often emphasize graphical user interfaces (screen, keyboard, mouse), which, while valuable, can overlook the potential of tangible, embedded, and embodied interaction.
- **10:00 AM - 10:30 AM: Visualization tools Exploration**
  - Invite participants to delve into existing visualization tools for core LLM mechanisms, including BertViz, Bbycroft, and LatentScope. The session will include an interactive

Figure 2: Example of Tangible properties to be explored during the workshop. From Angelini et al. [3].

walkthrough led by the organizers, offering hands-on experience and insights into these tools' capabilities.
- **10:30 PM - 10:45 PM: Break**
- **10:45 AM - 12:00 AM: Creativity Session**
  - Participants will divide into groups to brainstorm potential tangible prototypes making LLMs more interpretable and explainable through multi-sensory experiences. Each group is given the same card set and a different material focus, drawing inspiration from a range of provided materials.
- **12:00 PM - 1:00 PM: Lunch Break**
- **1:00 PM - 3:00 PM: Prototyping Session**
  - This part emphasizes iteration and testing, encouraging participants to think critically about how these multisensory experiences can enhance AI explainability through feedback from the group. During this time, participants can experiment with:
    - Multi-Sensory Feedback: Combining tactile, auditory, and visual outputs into a single prototype.
    - Narrative and Embodiment: Exploring how storytelling or bodily engagement with materials can enhance the clarity of AI processes.
- **3:00 PM - 3:30 PM: Break**
- **3:30 PM - 4:30 PM: Group Pitch & User Testing**
  - Each group will present their prototype to the rest of the participants, showcasing how their tangible, multi-sensory design offers a new perspective on explainability in AI. Question to cover:
    - What AI process or model does your prototype aim to explain?
    - How do the sensory modalities or material choices enhance understanding?
    - What are the challenges and opportunities discovered through your design?
  - Peer feedback: Facilitated session where each group receives constructive feedback from peers on how to further refine their concepts.
- **4:00 PM - 4:45 PM: Group reflection & Future Direction**
  - Sharing insights and takeaways from the studio, reflecting on the experience with material-based interactions and embodied sense-making in AI.
  - Discussion on future research: How can these insights be incorporated into ongoing research in XAI and HCI? Participants discuss possible follow-up collaborations, publications, or future studios/workshops.
- **4:45 PM - 5:00 PM: Closing Remarks & Next Steps**
  - Studio wrap-up: Summarizing key points, outlining next steps for documentation, sharing insights from the day.

References

[1] HLEG AI. 2019. High-level expert group on artificial intelligence. *Ethics guidelines for trustworthy AI* 6 (2019).

[2] Leonardo Angelini, Nadine Couture, Mira El Kamali, Quentin Meteier, and Elena Mugellini. 2022. Workshop on Tangible xAI. *ETIS 2022* (2022).

[3] Leonardo Angelini, Elena Mugellini, Nadine Couture, and Omar Abou Khaled. 2018. Designing the Interaction with the Internet of Tangible Things: A Card Set. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction* (Stockholm, Sweden) *(TEI '18)*. Association for Computing Machinery, New York, NY, USA, 299–306. https://doi.org/10.1145/3173225.3173288

[4] Brendan Bycroft. 2024. LLM Visualization. https://bbycroft.net/llm. Accessed: 2024-10-31.

[5] Ashley Colley, Kaisa Väänänen, and Jonna Häkkilä. 2022. Tangible Explainable AI - an Initial Conceptual Framework. In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia (MUM '22)*. Association for Computing Machinery, New York, NY, USA, 22–27. https://doi.org/10.1145/3568444.3568456

[6] Natalia Díaz-Rodríguez. 2023. Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion* (2023).

[7] Upol Ehsan, Philipp Wintersberger, Q. Vera Liao, Elizabeth Anne Watkins, Carina Manger, Hal Daumé III, Andreas Riener, and Mark O. Riedl. 2022. Human-Centered Explainable AI (HCXAI): Beyond Opening the Black-Box of AI. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (ACM Digital Library)*, Simone Barbosa (Ed.). Association for Computing Machinery, New York,NY,United States, 1–7. https://doi.org/10.1145/3491101.3503727

[8] 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*. ACM, Salzburg Austria, 1–4. https://doi.org/10.1145/3430524.3442704

[9] Maliheh Ghajargar, Jeffrey Bardzell, Alison Smith-Renner, Peter Gall Krogh, Kristina Höök, and Mikael Wiberg. 2021. Tangible XAI. *ACM Interactions* (2021).

[10] Michael F Goodchild and Donald G Janelle. 2010. Toward critical spatial thinking in the social sciences and humanities. *GeoJournal* 75 (2010), 3–13. https://doi.org/10.1007/s10708-010-9340-3

[11] 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* (Cambridge, Massachusetts, USA) *(TEI '10)*. Association for Computing Machinery, New York, NY, USA, 101–108. https://doi.org/10.1145/1709886.1709905

[12] Eva Hornecker. 2011. The role of physicality in tangible and embodied interactions. *Interactions* 18, 2 (March 2011), 19–23. https://doi.org/10.1145/1925820.1925826

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

[14] Sherry Hsi, Marcia C. Linn, and John E. Bell. 1997. The Role of Spatial Reasoning in Engineering and the Design of Spatial Instruction. *Journal of Engineering Education* 86, 2 (1997), 151–158. https://doi.org/10.1002/j.2168-9830.1997.tb00278.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2168-9830.1997.tb00278.x

[15] Ian Johnson. [n. d.]. LatentScope. https://enjalot.github.io/latent-scope/. Accessed: 2024-10-31.

[16] Hai-Ning Liang and Kamran Sedig. 2010. Role of interaction in enhancing the epistemic utility of 3D mathematical visualizations. *International Journal of Computers for Mathematical Learning* 15 (2010), 191–224. https://doi.org/10.1007/s10758-010-9165-7

[17] Jesse Vig. 2019. BertViz: A tool for visualizing multihead self-attention in the BERT model. *ICLR workshop: Debugging machine learning models* 3 (2019).

[18] Jonathan Wai, David Lubinski, and Camilla P. Benbow. 2009. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology* 101, 4 (Nov. 2009), 817–835. https://doi.org/10.1037/a0016127

[19] Zezhong Wang, Shan Hao, and Sheelagh Carpendale. 2024. Card-Based Approach to Engage Exploring Ethics in AI for Data Visualization. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 69, 7 pages. https://doi.org/10.1145/3613905.3650972

[20] Haiyan Zhao, Hanjie Chen, Fan Yang, Ninghao Liu, Huiqi Deng, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, and Mengnan Du. 2024. Explainability for Large Language Models: A Survey. *ACM Transactions on Intelligent Systems and Technology* 15, 2 (April 2024), 1–38. https://doi.org/10.1145/3639372