Conceptual Alignment in Human-Centered Design Teams through Critical Reflection on Large Language Models

Simone Ciciliano[1,*], Rosella Gennari[1]

[1]Free University of Bozen-Bolzano, via Buozzi 1, 39100 Bolzano, Italy

Abstract

Human-Centered Design (HCD) often requires different types of expertise, ranging from the specialised knowledge of designers and developers (interdisciplinarity) to the domain expertise of stakeholders (transdisciplinarity). Communication between team members is paramount, but often achieving mutual understanding is challenging. HCD also requires mitigation of potential risks related to the implementation of black-box Generative AI tools, such as Large Language Models (LLMs), which provide both interesting opportunities and critical risks. To the best of our knowledge, little attention has been paid to how LLMs can scaffold conceptual alignment among HCD team members in a way that minimises such risks. This paper presents recent work on fostering conceptual alignment among HCD team members by observing misalignment between team members and LLMs. To do that, the paper reports the case study of an interdisciplinary workshop, highlighting that reflecting critically on LLMs can represent a valuable way to foster mutual understanding in HCD processes. Based on the workshop's results, the paper also describes a graphical user interface based on 2*D* projections of LLM-generated sentence embeddings, to foster discussion among HCD team members and progressively attain in-group conceptual alignment. Finally the paper discusses implications and future work, aiming to foster discussion on how, in HCD teams, critical reflection can be leveraged to both minimise risks related to LLMs and how conceptual alignment as a strategy for ensuring LLMs' human-centricity and trustworthiness.

Keywords

Conceptual Alignment, Interdisciplinarity, Large Language Models, Workshop, Graphical User Interface

1. Introduction

Human-Centered Design (HCD) aims to design systems in a way that keeps people at the centre of the design process [1]. To do so, in some cases, HCD teams are interdisciplinary, that is, they involve experts from different fields, ranging from cognitive psychology to usability engineering. In co-design approaches, instead, HCD teams aim at transdisciplinarity, that is, they involve stakeholders as domain experts. Fostering mutual understanding is challenging in both cases and effective communication strategies are needed.

Recent research efforts have considered Pre-trained Large Language Models (LLMs) to facilitate communication in such teams, especially generative ones [2]. However, the blackbox nature of LLMs is problematic. An orthogonal approach focuses instead on the notion of **conceptual alignment**, that is, the coordination of abstract representations of a given phenomenon.

This paper advances the idea that critically reflecting on LLM-generated outputs can safely scaffold conceptual alignment among HCD team members in both interdisciplinary and transdisciplinary contexts. The underlying rationale is that critical reflection serves two complementary functions. First, it encourages team members to actively question LLM-generated outputs, helping to reduce risks such as hallucinations. Second, it allows ambiguities and conceptual differences among team members to surface, creating opportunities for these differences to be identified, addressed, and explored.

To substantiate this claim, the paper reports the case study of the **interdisciplinary workshop** titled *Prompting across Disciplines*, and shows how participants' feedback informed the design of an ad hoc **graphical user interface (GUI)** specifically intended to foster conceptual alignment through

Joint Proceedings of the ACM Intelligent User Interfaces (IUI) Workshops 2026, March 23-26, 2026, Paphos, Cyprus
*Corresponding author.
simone.ciciliano@student.unibz.it (S. Ciciliano); gennari@inf.unibz.it (R. Gennari)
0009-0005-3519-9333 (S. Ciciliano); 0000-0003-0063-0996 (R. Gennari)

critical reflection. This paper is structured as follows. The paper first outlines the most relevant related work, then the workshop and resulting GUI prototype. The conclusions reflect on results and future challenges ahead for conceptual alignment and LLMs in interdisciplinary and transdisciplinary efforts.

2. Related Work

2.1. HCD Teams and Communication

For keeping communication as accessible and unambiguous as possible in HCD teams, several toolkits rely on predominantly non-linguistic forms of communication, such as icons and play cards, e.g. [3]. Other HCD tools or techniques aim instead at reflecting through language, such as Brainwriting and SCRAMPER [4, 5]. However, they avoid discipline-specific language (i.e., **jargon**) [6]. Although avoiding jargon can be useful to facilitate communication, its removal or over-simplification, boundaries are crossed, might hamper communication, as recently highlighted by Bao et al. [7]. It can prevent participants from exploring different conceptual frameworks and notions, and hence they may overlook relevant exchange possibilities.

To the best of our knowledge, no contribution has so far focused on how reflecting critically on LLMs can foster the conceptual alignment of interdisciplinary or transdisciplinary teams.

2.2. Conceptual Alignment

The notion of conceptual alignment is two-fold, and therefore provides a particularly interesting perspective to address the issue of communication in HCD teams. In the context of Artificial Intelligence (**AI**) research, conceptual alignment can be defined as the state in which two communicating agents achieve a common representation of abstract knowledge through information exchange [8]. While this can be applied to several different paradigms (e.g., emergent communication [9] or semantic heterogeneity in ontologies [8]), it has gained significant popularity with LLMs based on transformer neural networks [10]. In the domain of **cognitive science**, conceptual alignment can be defined as the condition in which two or more people's mental models become sufficiently aligned to allow for mutual understanding through iterative, abductive processes [11].

In this paper, the former is leveraged to achieve the latter: by bringing HCD team members to reflect critically on conceptual alignment between them through LLMs, the goal is to support team members in understanding each other and jointly constructing a shared conceptual space for the the HCD process.

Lack of alignment may between human users and LLMs may lead to several issues, such as ignoring cultural differences [12], highlighting gender stereotypes [13] and ignoring domain-specific knowledge [14]. In order to safely leverage LLMs in the context of HCD, these risks need to be minimised. In the rest of the paper, these are referred to as LLM-related risks

2.3. Relevant Tools and Interfaces

Recent works to overcome barriers across disciplines include that of Bao et al. [7]. In this work, the authors develop a search tool prototype based on static word embeddings. This tool leverages word embeddings trained on discipline-specific corpora, therefore facilitating researchers in looking for information in scientific literature they may not be familiar with. While the authors frame conceptual alignment as a translation problem, this paper addresses it from an interactional, situated perspective: conceptual alignment, in this case, is addressed as the goal of an iterative, contextualised interaction.

While not specifically focused on conceptual alignment, previous research on group recommender systems also focused on supporting the kind of interaction considered in this paper: the work of Delic et al. [15] highlighted that groups achieve consensus in an iterative and context-specific way. The AI recommender systems serve to give a structure to the group's discussion, so that relevant aspects always remain in focus and the discussion can move back and forth across stages. Most notably, it was also highlighted that tools meant to support groups in achieving consensus (e.g., chat-based interfaces)

should not attempt to provide a solution, but rather support all involved parties in reflecting collectively and getting autonomously to a final decision [16].

3. Workshop Case-Study and Interface Prototype

This section presents the case study of the interdisciplinary workshop *Prompting across Disciplines* and the GUI prototype designed in accordance with the workshop results. The workshop involved 8 PhD students, all coming from different backgrounds (see Table 1). The purpose of the workshop was to allow participants to acknowledge differences and commonalities among themselves through critical reflection on domain-specific language. The activities focused on each participant's jargon and the way in which it related to that of all others. For further details on the workshop, see [17]. The GUI prototype has been consequently developed with a double intention: on the one hand, it incorporates feedback from workshop participants to ensure that the use of LLMs focuses on fostering participants' critical reflections. On the other hand, it aims to adapt the proposed activities from an interdisciplinary to a transdisciplinary context, therefore also attempting to make the activity accessible to stakeholders from society.

3.1. The *Prompting Across Disciplines* Workshop

The workshop consisted in three progressive activities (Activity 1–3), to enable participants to reflect critically on their domain-specific language and how it relates to that of different disciplines, interacting both with an LLM and among themselves. Each activity was followed by shared reflection moments, where participants shared thoughts and impressions about the activity they just performed. Each activity is described in the following subsections.

Activity 1: Co-generating Sentences. Participants were given a sheet of paper with a set of incomplete sentences and were then asked to add one word to each sentence. Then, every participant had to pass on the sheet of paper and repeat the activity on the sheet they had received, similarly to what happens in autoregressive text generation. For example, with the incomplete sentence "Today it is", a participant would continue the sentence adding the word "sunny", the following participant might at that point add "and" and the following one "warm" and so on. The goal was not in-so-much to form a complete sentence, but rather to familiarise participants with the modalities used by current LLMs to generate text, as well as introducing participants to the diversity within the group by observing the different continuations that each sentence may have.

Activity 2: Mapping Keywords. In the second activity, participants received a set of sticky notes and were asked to write one word per note that they considered relevant for their research project, i.e., primary beneficiaries, topic, goals. Afterwards, emulating the visualisation of a 2-dimensional distributional language model, they were asked to arrange the sticky notes on the blackboard, interacting among themselves to determine how different or similar the keywords were. This way participants created an affinity diagram by arranging the sticky notes on a blackboard. See Figure 1.

Activity 3: Prompting and Evaluating ChatGPT. In this last activity, participants used some of the keywords from Activity 2 to prompt an LLM (ChatGPT [18]), to assess its ability to understand discipline-specific jargon. This was used as a strategy to have participants elaborate on the definition of concepts relevant for them, so that also other participants could gain such understanding.

Workshop Results. All participants provided very positive feedback. Everyone agreed on the fact that the activities were clearly explained and helped them gain more awareness about other participants' work and background, as well as how LLMs might respond to discipline-specific jargon. That is the case of Participant 6, who claimed to be satisfied with the LLM's definition of *information*. However, some

Participant	Background	Stakeholders
P1	Computer Science	HCI Research Community
P2	Human-Computer Interaction	Vulnerable Populations
P3	Electronic Engineering	Cultural Sites Visitors
P4	Cognitive Science	Professionals in Creative Professions
P5	Computer Science	Professionals in Industrial Production
P6	Computer Engineering	Professionals in Industrial Production
P7	Computer Engineering	Professionals in Industrial Production
P8	Neuro and Psychomotor Rehabilitation	People with Intellectual Disabilities

Table 1
Labelling and description of participants

Figure 1: Activities of the first workshops. **Left image**: the Co-generating Sentences Activity 1. **Right image**: result of the Mapping Keywords Activity 3.

participants highlighted potential problems in the use of the LLM proposed in Activity 3. That is the case of Participant 8, who attempted to obtain a definition of *disorders*, in the domain of psychomotor rehabilitation, deeming the LLM's definition too simplistic and unclear. Although cases like this created the opportunity to discuss the issues with such definitions, these interactions focused heavily on the alignment between the LLM and individual participants, rather than among them. This might me undesirable for future iterations of the workshop. Moreover, LLM-generated answers might which may induce participants to just accept the correctness of the answer instead of elaborating on it with team members. This is also another aspect that should be avoided in future iterations.

3.2. Interface Design

The workshop's results were analysed and relevant lessons (**L**) for the following iterations of the workshop were extracted as follows:

L1: Ensure that HCD team members can address conceptual alignment regardless of their background and with as little cognitive load as possible.

L2: Ensure that LLM-related risks in HCD contexts are limited and easy to address for HCD team members.

L3: Ensure that the focus stays on conceptual alignment among team members rather than between them and the LLM, individually.

To address these points, a GUI prototype was developed, focusing on $2D$ projections of sentence embeddings, so that HCD team members could jointly interact with them through a large touchscreen surface. The GUI prototype focuses not on facilitating the organisation among team members or their brainstorming activities, as several widely available software tools do, but specifically on conceptual alignment among them.

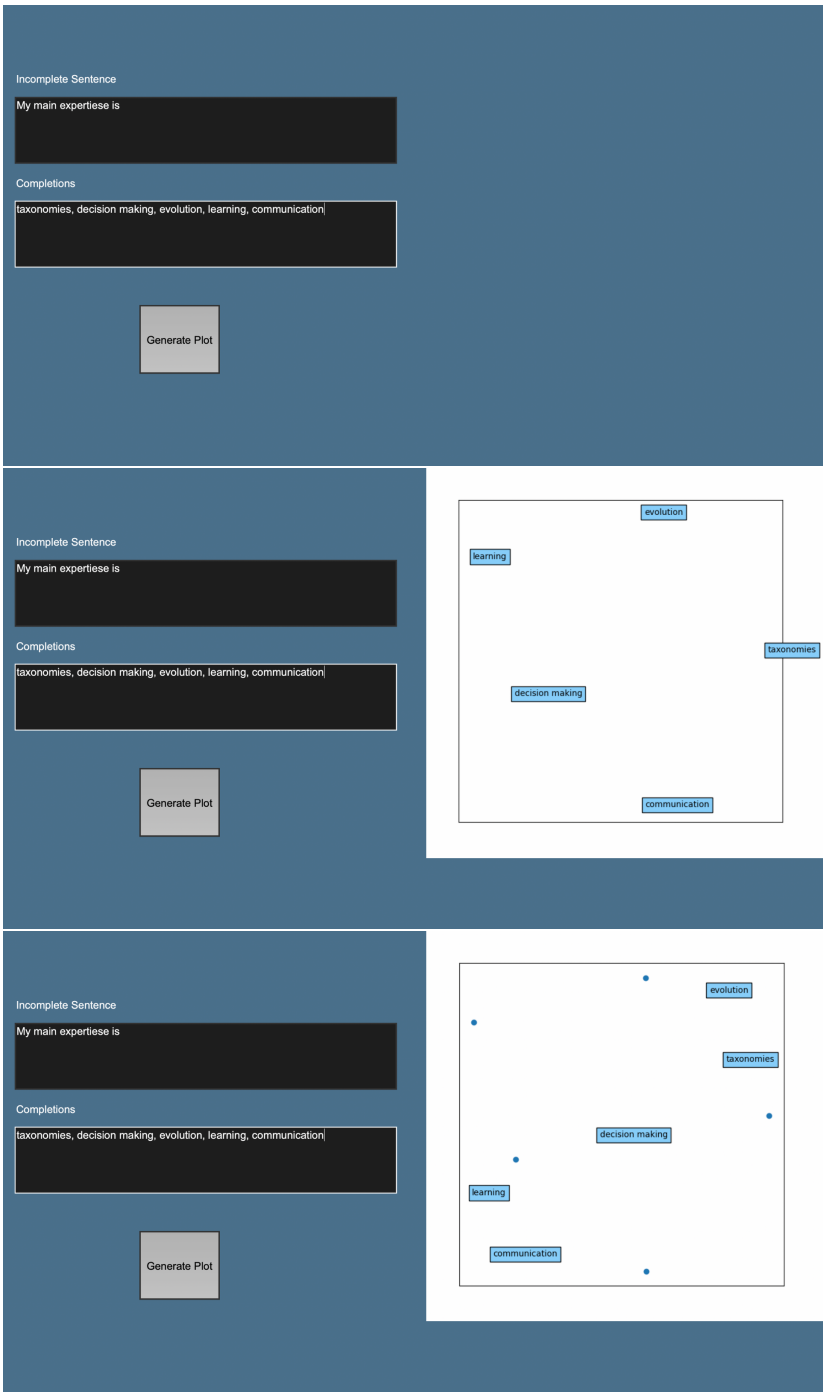

Figure 2: An example of the described GUI prototype: First, users write an incomplete sentence and several completions (**top image**), then the GUI generates a plot of $2D$ sentence embeddings (**center image**) which users can modify (**bottom image**).

Interface Description. The GUI prototype was entirely developed in `Python` version 3.9.23), using the libraries `Tkinter` (version 8.6), `Matplotlib` (version 3.9.4), `UMAP` (version 0.5.9) and `SentenceTransformers` (version 5.1.2). For reference, see Figure 2. It provides two distinct text input boxes, one asking for a sentence to be completed with a final part (a word or a short phrase), the other asking for several comma-separated completions (much like in Activity 1). The incomplete sentence is meant to elicit relevant concepts from participants, so that each concept is then merged with the incomplete sentence and transformed into a context-sensitive vector embedding through an LLM. The vectors' dimensions are then reduced to 2 through UMAP and represented visually in the

GUI as a scatter plot (similar to the affinity diagram obtained from Activity 2). Each vector in the plot is represented with a text label (the corresponding sentence completion), which HCD team members can move and replace through a drag-and-drop interaction with the GUI.

Critical Reflection on LLMs. The most notable difference between the use of LLMs in the outlined workshop and the GUI prototype is the shift from a textual LLM-generated output to a graphical one, shifting from natural language generation to natural language understanding. This allows to address **L1** and **L2**: For **L1**, the visualisation of vector embeddings, where spatial proximity serves as an indicator for semantic similarity, provides an intuitive representation of semantic relations between concepts, which allows also non-experts to understand and address such relations, jointly with others. For **L2**, the GUI only projects what users give as an input, and in case of associations between concepts that might seem biased or wrong, HCD team members can address it and modify it. HCD team members to leverage LLMs as tools to critically reflect on differences among themselves, rather than focusing on LLM-generated answers (and all related potential risks).

Fostering Conceptual Alignment. The shift from a one-at-a-time, text-based interaction with an LLM to a collective, visual one posits the GUI prototype to directly foster conceptual alignment among team members, directly addressing **L3**. More specifically, providing a single visual representation for the whole group to interact with through a large touchscreen can ensure both a low barrier for all team members to actively participate and an engaging, safe way to leverage LLMs for this task.

4. Conclusion

Final Remarks. This paper claims that critical reflection on LLM-generated outputs helps scaffold conceptual alignment among HCD team members in interdisciplinary and transdisciplinary contexts.

It illustrated the idea through an interdisciplinary workshop, in which participants critically reflected on conceptual alignment among themselves and between them and a generative LLM. The paper concludes by describing the design of a GUI prototype, informed by the results of the workshop. This visually displays LLM-generated vector embeddings. By critically and jointly reflecting on this visualisation, participants can work towards their conceptual alignment.

The conclusive message of the reported case study is that, in the context of HCD teams, reflecting critically on concepts and their linguistic expression represents both a valuable strategy to counteract the lack of transparency of LLMs, and a way to scaffold conceptual alignment among team members.

Future Work. Future directions for this work involve, first and foremost, testing the GUI prototype in a real-world scenario with different stakeholders.

In particular, while this paper argues that critical reflection can scaffold conceptual alignment among HCD team members, it still has to be tested to what extent this is the case. Lastly, relevant future work might involve improving the GUI prototype, for example testing different dimensionality reduction algorithms or improving the visual organisation of the different elements.

Acknowledgments

This contribution benefited from the financial support of the TORUS Interdisciplinary research project, granted by the Free University of Bozen-Bolzano.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] I. D. Foundation, What is human-centered design?, https://www.interaction-design.org/literature/topics/humanity-centered-design, 2021. Accessed: 2025-07-07.

[2] J. Cowin, B. Oberer, C. Leon, Trans-disciplinary communication in the chatgpt age: A systems perspective, in: Proceedings of the 17th International Multi-Conference on Society, Cybernetics and Informatics, International Institute of Informatics and Cybernetics, 2023. doi:https://doi.org/10.54808/IMSCI2023.01.138.

[3] R. Gennari, M. Matera, A. Melonio, M. Rizvi, How to enable young teens to design responsibly, Future Generation Computer Systems 150 (2024) 303–316. URL: https://www.sciencedirect.com/science/article/pii/S0167739X23003321. doi:https://doi.org/10.1016/j.future.2023.09.004.

[4] A. B. VanGundy, Brain writing for new product ideas: an alternative to brainstorming, Journal of Consumer Marketing 1 (1984). doi:https://doi.org/10.1108/eb008097.

[5] R. Friis Dam, T. Yu Siang, Introduction to the essential ideation techniques which are the heart of design thinking, 2022. URL: https://www.interaction-design.org/literature/article/introduction-to-the-essential-ideation-techniques-which-are-the-heart-of-design-thinking, accessed: 2025-07-07.

[6] S. Shashwat, X. Xue, Z. Yu, J. Aboky-Djanty, common::ground: A gamified toolkit for facilitating meaningful interactive sessions, in: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25, Association for Computing Machinery, New York, NY, USA, 2025. URL: https://doi.org/10.1145/3706599.3719894. doi:10.1145/3706599.3719894.

[7] C. Bao, Y.-T. Shiue, M. Carpuat, J. Chan, Words as bridges: Exploring computational support for cross-disciplinary translation work, in: Proceedings of the 30th International Conference on Intelligent User Interfaces, IUI '25, Association for Computing Machinery, New York, NY, USA, 2025, p. 1598–1623. URL: https://doi.org/10.1145/3708359.3712110. doi:10.1145/3708359.3712110.

[8] M. Atencia, M. Schorlemmer, An interaction-based approach to semantic alignment, Journal of Web Semantics 12-13 (2012) 131–147. URL: https://www.sciencedirect.com/science/article/pii/S1570826811001016. doi:https://doi.org/10.1016/j.websem.2011.12.001, reasoning with context in the Semantic Web.

[9] A. Lazaridou, K. M. Hermann, K. Tuyls, S. Clark, Emergence of linguistic communication from referential games with symbolic and pixel input, 2018. URL: https://arxiv.org/abs/1804.03984. arXiv:1804.03984.

[10] S. Rane, M. Ho, I. Sucholutsky, T. L. Griffiths, Concept alignment as a prerequisite for value alignment, 2023. URL: https://arxiv.org/abs/2310.20059. arXiv:2310.20059.

[11] A. Stolk, L. Verhagen, I. Toni, Conceptual alignment: How brains achieve mutual understanding, Trends in Cognitive Sciences 20 (2016) 180–191. URL: https://www.sciencedirect.com/science/article/pii/S1364661315002867. doi:https://doi.org/10.1016/j.tics.2015.11.007.

[12] B. AlKhamissi, M. ElNokrashy, M. Alkhamissi, M. Diab, Investigating cultural alignment of large language models, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 12404–12422. URL: https://aclanthology.org/2024.acl-long.671/. doi:10.18653/v1/2024.acl-long.671.

[13] V. Neplenbroek, A. Bisazza, R. Fernández, Reading between the prompts: How stereotypes shape llm's implicit personalization, 2025. URL: https://arxiv.org/abs/2505.16467. arXiv:2505.16467.

[14] S. Palta, N. Chandrasekaran, R. Rudinger, S. Counts, Speaking the right language: The impact of expertise (mis)alignment in user-AI interactions, in: K. Inui, S. Sakti, H. Wang, D. F. Wong, P. Bhattacharyya, B. Banerjee, A. Ekbal, T. Chakraborty, D. P. Singh (Eds.), Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics, The Asian Federation of Natural Language Processing and The Association for Computational Linguistics, Mumbai, India,

2025, pp. 58–69. URL: https://aclanthology.org/2025.ijcnlp-short.5/.

[15] A. Delic, J. Neidhardt, T. N. Nguyen, F. Ricci, L. Rook, H. Werthner, M. Zanker, Observing group decision making processes, in: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 147–150. URL: https://doi.org/10.1145/2959100.2959168. doi:10.1145/2959100.2959168.

[16] T. N. Nguyen, F. Ricci, A chat-based group recommender system for tourism, Information Technology & Tourism 18 (2018) 24. doi:10.1007/s40558-017-0099-y.

[17] S. Ciciliano, R. Gennari, Workshops and hybrid toolkits for communicating across disciplines, in: Proceedings of the 16th Biannual Conference of the Italian SIGCHI Chapter, CHItaly '25, Association for Computing Machinery, New York, NY, USA, 2025, pp. 1–6. URL: https://doi.org/10.1145/3750069.3750087. doi:10.1145/3750069.3750087.

[18] OpenAI, Introducing chatgpt, https://openai.com/index/chatgpt/, 2022. Accessed: 2025-07-07.