Organizing Transdisciplinary Research and Innovation: the Case of the Socially-Aware Artificial Intelligence Focus Area

Loïs Vanhée[1,*]

[1]TAIGA Umeå's Center for Transdisciplinary Artificial Intelligence for the Good of All and Department of Computing Science, MIT-Huset, 901 87 Umeå

Abstract

Research, education, and innovation on trust in human-AI teams inherently involves inter/transdisciplinary considerations, which subsequently raises a wide array of challenges on how to produce such research, from networks, funding, to alternate processes for producing science. Whereas the interest around this topic mostly revolves around scientific productions, methods for organizing the required underlying scientific productive system remain limited, from reaching unknowingly relevant researchers to helping interested researchers to unfold, grow, and strive. This paper is dedicated to introducing one of such academic environments through introducing the Socially-Aware Artificial Intelligence focus area of TAIGA, Umeå's center for Transdisciplinary Artificial Intelligence, as a research, education, and innovation environment dedicated to support transdisciplinary AI research and how such an environment can be deployed for serving the development of research on trust within human-artificial intelligence teams in particular.

Keywords

transdisciplinarity, interdisciplinary AI, research organization, socially-aware artificial intelligence

1. Introduction

Creating trust within human-Artificial Intelligence (AI) teams inherently involves crossing a multiplicity of frames, may they arise from different disciplines (e.g. psychology, computer science, sociology, organizational theory, management), or from different sectors (e.g. academia, research institutes, end-users, impacted businesses). This crossing of frames can take various forms such as multidisciplinarity (i.e. applications of methods of different disciplines on the same object of study); interdisciplinarity (i.e. developing new research topics and methods on a given object of study at the interstice of multiple disciplines); and transdisciplinarity (i.e. developing new knowledge about an object of study that combines perspectives of academic and non-academic stakeholders) [1].

Despite the criticality of multi/inter/transdisciplinary research activities for studying trust in Human AI teams, such research is, to a vast extent, carried through and forced to be carried

MultiTTrust: 2nd Workshop on Multidisciplinary Perspectives on Human-AI Team, Dec 04, 2023, Gothenburg, Sweden

*Corresponding author.

lois.vanhee@umu.se (L. Vanhée)

https://www.umu.se/en/staff/lois-vanhee/ (L. Vanhée)

0000-0002-4147-4558 (L. Vanhée)

through disciplinary lenses. From an organizational perspective, these research activities are, for a vast majority, carried by classic scientific institutions, which are themselves structured along disciplines from departments to communities, education, sources of funding, career progression ladders, and recognition schemes (e.g. priority on journals or conference, author ordering). This disciplinary mindset goes beyond social structures, being deeply internalized by the researchers themselves, who, often unconsciously, carry practices, values, and assumptions (on e.g. where to publish, what is quality science, down to the definition of what a well-formated text is) arising from the primary disciplines they studied as undergraduates and their host institutions. Incidentally, such disciplinary-centrism creates visible and invisible boundaries [2] that inherently hinder the innovative mindset needed for transdisciplinary research and its specific manifestation, such as studying trust factors in human-AI teams.

This paper is dedicated to providing an answer to the question: *How can an existing classic academic setting, organized along disciplinary faculties and departments, be adapted for best enabling transdisciplinary research?* Whereas this question is the object of study of entire communities, this paper narrows the problem down to a case-study, i.e. a practical attempt at organizing of research activities for enabling transdisciplinary research on Socially-Aware Artificial Intelligence [3] (SAI), in particular through the studying the SAI focus area of Umeå's center for Transdisciplinary AI for the Good of All [4] (TAIGA). Specifically, this paper details in Section 2 how the center and focus area are structured as to support transdisciplinary AI research. Section 3 details how such organizational premises can be deployed for supporting the initiation of new research track, in particular on how Umeå University research and researchers can be bound to the MULTITTRUST interests and communities, crossing scientific tracks, scientists, and structural prospects.

2. Organizing Transdisciplinary Research Through Centers and Focus Areas

The SAI focus area within TAIGA is organized as to answer to the limitations raised by classic disciplinary research organization. Briefly, TAIGA is dedicated to enable and develop transdisciplinary AI research, education, innovation, and social impact at the university level. TAIGA is established over two framing principles. First, *AI is framed as a transversal object of study*. Unlike classic perspectives that narrow AI down to technical (computer-science) perspectives, TAIGA seeks to provide equal opportunities for all the relevant disciplines to partake in TAIGA's activities with an equal foot, including disciplines that can benefit from using AI methods in their research and praxis (e.g. education, health) [5, 6] and disciplines that can study the ramifications tied to AI (e.g. how people relate to AI systems, AI as sociopolitical object [7]). Second, framing AI research not only as an intellectual endeavour but also as an object of high salience for society for which we are bound to provide answers.

As an organizational structure, TAIGA is organized around a coordination core and eight focus areas that arise from eight different disciplines, titled: "AI in health and medicine"; "AI and art"; "education and AI"; "understanding and explaining AI"; "critical, ethical, legal, social AI"; "AI management"; "embodied interactive AI"; and "Socially-Aware Artificial intelligence" (SAI). This organization is set as to enable the best connectivity across all the relevant disciplines and

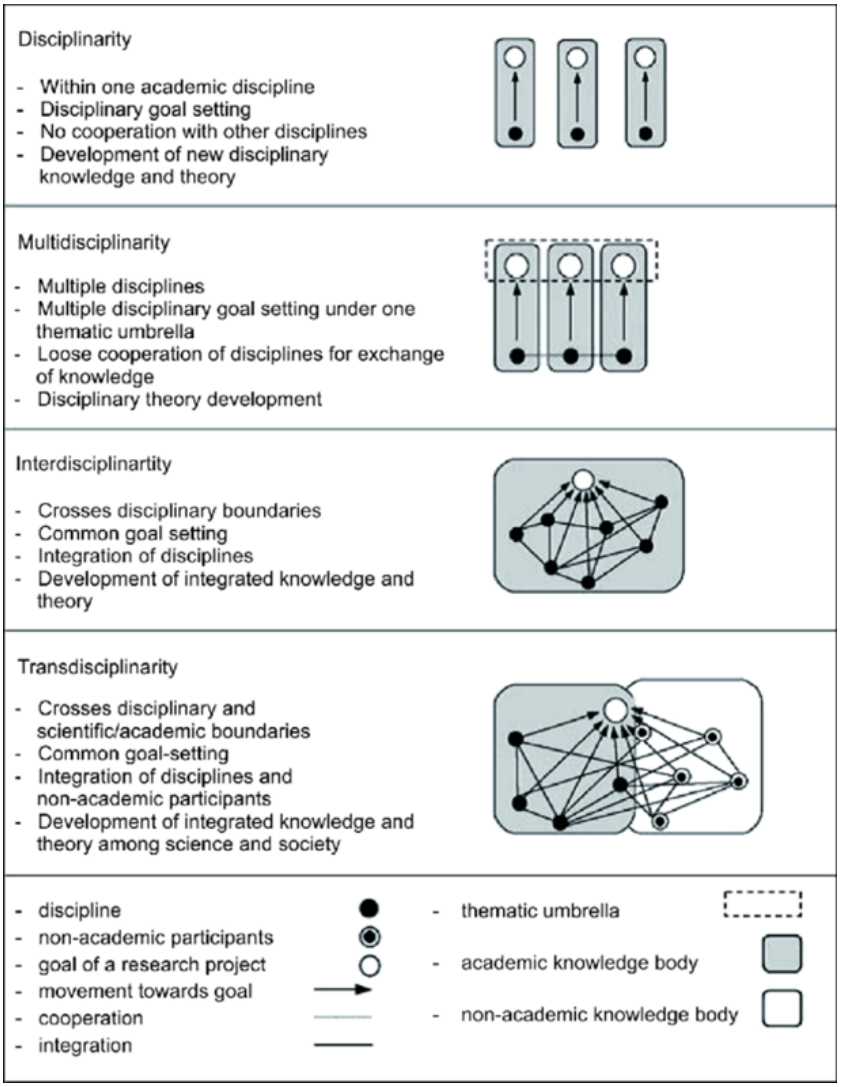

Figure 1: Taxonomy of disciplinary, multidisciplinary, interdisciplinary, and transdisciplinary research. Source: [1] (reprinted by creative commons permission)

the university.

The SAI focus area is primarily hosted by the computing sciences department. The SAI focus area the most centred around the study of new AI methods that allow accounting for and adapting to human social behavior (i.e. human-centered SAI) and/or accurately simulating and replicating human social behavior (human-like SAI). Beyond technical and applicative interests, the SAI focus area is dedicated to raising operational capacity at the university level by providing state-of-the-art AI-powered competencies and research methods to interested other departments (e.g. psychology, pedagogy, medicine, arts) as well as for fostering a university-wide discussion of the ramifications of such technologies on society (e.g. sociology, organizational science, business, law). The SAI focus area is also involved in reaching out to sectors and

actors beyond the university (e.g. businesses, region, other centers). As a general strategy for achieving this mission, the SAI focus area is dedicated to identifying, enabling, and supporting the development of all the prospective scientific niches available to Umeå University, crossing all relevant pre-existing available resources (e.g. staff, skills, networks) and seeking to cover and synergise all the university's missions (i.e. research, education, innovation, social impact, communication). This action is articulated along six activities, including fostering internal collaborations through opportunities for academics to reach each other (A1), supporting their academic missions (i.e. research (A2), education (A3), grant-writing and innovation projects (A4)) tied to SAI, and reaching out to wider communities both within academia (A5) and beyond (A6).

A1: Stimulating SAI Networks Growth through internal events (e.g. seminars, workshops, pitch events, coffee gatherings); internal diffusion of available SAI competences and methods (e.g. connecting computer scientists specialized in Natural Language Processing to pedagogy and psychology); and workshops for identifying and spreading interests related to SAI (e.g. regulation of SAI) in visits tailored to the interests of the disciplinary departments of the university.

A2: Supporting Local SAI Research through identifying and organizing the internal activities carried out by the university. This activity is structured along two axes. First, an achievement-centered approach, through a research-focused bibliometric analysis of all Umeå University's SAI activities (e.g. listing all retrievable research papers, listing of research trends, trends over time, productivity over time, key researchers of the topic and their contributions and interests) and making this information available to other researchers [3]; second an ambition-centered approach involving a systematic collection of current and future research interests through forms and networking events.

A3: Development of Specialized Educational Activities through stimulating the creation of transdisciplinary networks for enabling sufficient skill capacity on the teacher's side and interest on the student side as well as for acquiring mastery and legitimacy at the university level for enabling change to occur administratively, as well as demonstrations of feasibility through practical activities such as pedagogic hackathons on AI for social good[1].

A4: Supporting Funding and Development of new research activities through the dissemination of relevant sources of funding to the interested community, the organization of pitch and case-specific or problem-specific events (e.g. AI for sustainability, AI for local urban planning), and offering starting grants (seed funding) for providing the time to engage in the preliminary research required for larger grant-writing.

A5: Academic outreach, both nationally and internationally through the organization of international workshops on SAI topics (e.g. Interdisciplinary Design of Emotion-sensitive Agents at the Autonomous Agents and Multi Agent Systems conference [8]); international special interest groups on SAI-related topics (e.g. AI for crisis response [9]); special tracks on SAI conferences (e.g. human-like deliberation and deliberation during crises at the Social Simulation conference [10, 11]);

A6: Public Outreach through the organization and participation of seminars open to the general public (e.g. the frAIday seminar series [12], a weekly seminar that regularly reaches more

[1] https://www.umu.se/en/research/projects/autograide---automated-grading-of-ai/

than a hundred of participants from the university and beyond); organization and participation in events involving innovation and society actors (e.g. dedicated encounters with stakeholders, contribution to TAIGA conferences).

As an overall modus operandi, the resources of the SAI focus area are dedicated to cater for qualitative spaces for long-term collaboration rather than for short-term reporting (e.g. number of papers per year), which is expressed in at the implementation of the SAI focus area. For example, the yearly seed funding call, dedicated to offer a small seminal budget (funding about a month of research time plus a conference and/or technical resources) is organized as to best support open-minded, creative, high-risk high-rewards collaborations either internal to the university or external, as long as the university is involved. Within this frame, the format of the application is purposefully constrained to be short (600 words), focusing on how the project can develop SAI potentials for the university rather than on (e.g. budgeting) details, while keeping relatively low demands in terms of reporting: the funding is serving the initiation of collaborations rather than the other way around.

3. Case Study: Enabling Trust and Human-AI Teams Research

The aforementioned activities of the Social AI focus area allow for an actor external to Umeå University (a research institution, a research group, an independent research) to effectively identify prospects for collaborations with Umeå research actors in general. This section is dedicated to specifying this identification for best enabling internal and external collaboration on trust factors in human-AI teams research. The carried activities allow covering manyfactorial considerations, including assessing the relevance of the research carried by the institution, whether and how researchers (and subsequently, personal time, skills, interests, resources) can be mobilized in new collaborations, and practical concerns for achieving collaboration in practice (e.g. access to basic resources for initiating collaborations).

From a research perspective, a review of the SAI scientific activities brings into light three lines of research carried out by Umeå University actors that touch upon trust in human-AI teams research. A first line of research, directly related to the topic, bring forward interdisciplinary studies of the impact of cultural factors in trust in human-AI systems: how culture impacts the psychological processes involved in trust building and what does it entail when developing human-AI systems [13, 14]. A second line of research captures human-AI teaming through the lenses of adjustable autonomy and covers psychological factors, such as accounting for the cognitive availability and demands made to involved humans [15], the integration of norms the system should best try to comply to when interacting with humans [16], the development of intuitive man-machine interaction media in which humans can advise the system on (un)desirable courses of actions [17, 18] and more general frameworks on man-robot teams [19]. A third line of research considers broader frames, bringing forward trustworthy AI considerations and in particular on the topic of transparency and explainability [20, 21, 22]. From a researchers' perspective, an array of profiles stand out. Two profiles are engaged in the framing of adjustable autonomy along technical lenses with some involvement on the trustworthy AI track from a technical AI perspective; a set of profiles focusing on the trustworthy/explainable AI track, either from a technical perspective or a sociological perspective; and one profile engaged with

psychological and interactional factors in trust in human-AI teaming (culture, cognitive availability, anxiety) with a more interdisciplinary approach. Most of the involved profiles have worked with both body-less AI and embodied (robotic) systems.

From a resource perspective, the analysis brings forward the availability of an array of AI methods that can be relevant to the topic (e.g. robots, neural networks, natural language processing), albeit most remain to be further articulated in the context of trust in human-AI teaming. This analysis, based on the systematic mapping of already developed research lines and involved researchers, acts as an effective approach for helping internal or external actor to consider prospective collaborations along the line of trust in human-AI teams with internal researchers, by bringing into light an array of feasible research directions and interested collaborators. The next step consists in exploiting other means for engaging with interested teams, either through direct contact with relevant researchers, or through engaging in SAI focus area's activities, such as seminars and networking events, or engaging with the SAI focus area coordinator.

4. Conclusion and discussion

This paper brings into light an approach to research organization dedicated to creating and fostering transdisciplinary research, innovation, education, and social impact at the level of a university on trust in human-AI teaming, by describing one of such structure, namely the Socially-Aware AI (SAI) focus area from TAIGA, the center for transdisciplinary artificial intelligence of Umeå University. Through presenting the approaches and missions of the SAI focus area and specifying them to the case of trust in human-AI teaming, the paper also shows how such an organizational approach can operate as an effective support for connecting new research ideas, internal and external researchers, and means for such collaborations to be initiated (e.g. seed funding).

The SAI focus area is dedicated to providing the structures for alleviating the key pitfalls of classic, discipline-centered organization of most universities by taking a transversal university-wide stance on SAI research, education, and innovation: SAI methods are not (only) computer science methods but an object of study of a wide array of disciplines. By its systematic identification of former and ongoing research results and active researchers and through its proactive approach to initiating and supporting collaborations, the SAI focus area enables for new research tracks and collaborations to be added organically to existing objects of study already mastered by (a part of) the university. This approach, which dis-encloses objects of study from specific disciplines, allows for other disciplines to engage in the scientific debate surrounding these objects of study as well as for scientists in seeking help from other disciplines to be provided with the right contacts for this help, if reachable, to be offered (e.g. computer scientists seeking supports from psychologists for validating their models).

While the SAI focus area remains too young for the potentials it creates to have yet matured in finished research and funded projects (a process commonly scaling in years for interdisciplinary research), early interventions have been positively received by a significant portion of the local SAI community and by reached SAI researchers. Early contacts between computer scientists, sociologists, philosophers, and law in workshops and hackathons facilitated by the SAI focus area

have already demonstrated the potential transformative approach on the research processes (e.g. how models are built, assumptions), research outputs (e.g. qualities of the produced models) and even research purposes (e.g. questioning the fitness of the models for society). As to exemplify the case of trust in human-AI teams, the SAI approach put in motion in this paper brought into light the potentials for collaboration along three research lines, from highly interdisciplinary psychology-grounded perspective on trust, to technical perspectives on trustworthiness and adjustable autonomy as well as fast tracks for these lines to be turned into working collaborations. The options being laid out, now is the window of opportunity for internal and external actors from all disciplines to step in enable the scaling up of this research.

5. Acknowledgement

The author acknowledges the support of TAIGA, Umeå's center for transdisciplinary artificial intelligence for having supported the production of this paper.

References

[1] G. Tress, B. Tress, G. Fry, Clarifying integrative research concepts in landscape ecology, Landscape ecology 20 (2005) 479–493.
[2] F. Siedlok, P. Hibbert, The organization of interdisciplinary research: modes, drivers and barriers, International Journal of Management Reviews 16 (2014) 194–210.
[3] L. Vanhée, Social ai research webpage, 2023. URL: https://sites.google.com/view/taiga-socialai/research.
[4] U. University, Umeå university's center for transdisciplinary for the good of all (taiga), 2023. URL: https://www.umu.se/centrum-for-transdisciplinar-ai/.
[5] K. Båth, E. Mårell-Olsson, J. Wedman, K. Danielsson, K. Grill, Humanities education in the age of chatgpt: risks and opportunities, in: Higher Education in the Age of ChatGPT: risks and opportunities, Online and in Umeå, Sweden, May 23, 2023, 2023.
[6] K. Kjell, P. Johnsson, S. Sikström, Freely generated word responses analyzed with artificial intelligence predict self-reported symptoms of depression, anxiety, and worry, Frontiers in Psychology 12 (2021) 602581.
[7] S. Lindgren, Critical theory of AI, John Wiley & Sons, 2023.
[8] L. Vanhée, Interdisciplinary design of emotion-aware agents (idea) workshop, 2023. URL: https://en.uit.no/project/idea.
[9] F. G. C. Kammler, L. Vanhée, Building resillience with social simulations (bricss) special interest group, 2023. URL: https://sites.google.com/view/bricss/home.
[10] E. S. S. Association, Simulating in crises special track at the social simulation conference 2023, 2023. URL: https://ssc23-sphsu.online/simulating-in-crises/.
[11] E. S. S. Association, Sense and sensibility special track at the social simulation conference 2023, 2023. URL: https://ssc23-sphsu.online/sense-sensibility/.
[12] U. University, Fraiday webpage, 2023. URL: https://www.umu.se/en/research/our-research/features-and-news/artificial-intelligence/fraiday/.

[13] M. Borit, L. Vanhée, P. Olsen, Understanding the impact of culture on cognitive trust-building processes: How to increase the social influence of virtual autonomous agents., in: TRUST@ AAMAS, 2014, pp. 100–111.
[14] M. Borit, L. Vanhée, P. Olsen, Towards enhancing trustworthiness of socially interactive and culture aware robots (2014).
[15] L. Vanhée, L. Jeanpierre, A.-I. Mouaddib, Optimizing requests for support in context-restricted autonomy, in: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2021, pp. 6434–6440.
[16] L. Vanhee, H. Aldewereld, F. Dignum, Implementing norms?, in: 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, volume 3, IEEE, 2011, pp. 13–16.
[17] L. Vanhée, L. Jeanpierre, A.-I. Mouaddib, Augmenting markov decision processes with advising, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 2531–2538.
[18] L. Vanhée, L. Jeanpierre, A.-I. Mouaddib, Augmenter les processus de décision via des conseils, in: Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JFPDA), 2019.
[19] M. Chiou, S. Booth, B. Lacerda, A. Theodorou, S. Rothfuß, Variable autonomy for human-robot teaming (vat), in: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, 2023, pp. 932–934.
[20] R. H. Wortham, A. Theodorou, Robot transparency, trust and utility, Connection Science 29 (2017) 242–248.
[21] D. Calvaresi, Y. Mualla, A. Najjar, S. Galland, M. Schumacher, Explainable multi-agent systems through blockchain technology, in: Explainable, Transparent Autonomous Agents and Multi-Agent Systems: First International Workshop, EXTRAAMAS 2019, Montreal, QC, Canada, May 13–14, 2019, Revised Selected Papers 1, Springer, 2019, pp. 41–58.
[22] S. Knapič, A. Malhi, R. Saluja, K. Främling, Explainable artificial intelligence for human decision support system in the medical domain, Machine Learning and Knowledge Extraction 3 (2021) 740–770.