Institutionalized support for transdisciplinary research, education, and innovation on trust in human-AI teams:
An example of the TAIGA Socially-Aware Artificial Intelligence focus area

Loïs Vanhée

TAIGA Umeå's Center for Transdisciplinary Artificial Intelligence for the Good of All and Department of Computing Science, Umeå University, lois.vanhee@umu.se

Research, education, and innovation on trust in human-AI teams inherently involves inter/transdisciplinary considerations, which subsequently raise a wide array of challenges on how to produce such research, from networks and funding, to alternate processes for producing science. Whereas the interest around this topic mostly revolves around scientific productions, considerations about how to organize 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 and it presents the Socially-Aware Artificial Intelligence focus area of TAIGA, Umeå University 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.

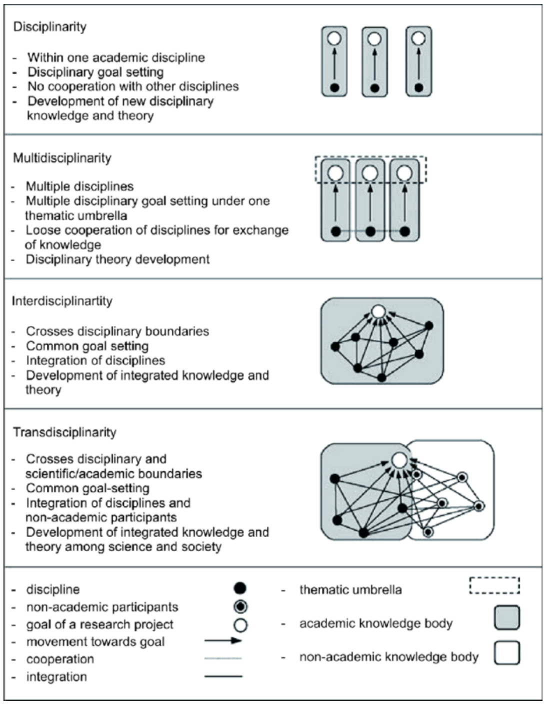
Keywords: transdisciplinarity; interdisciplinary AI; research organization; socially-aware artificial intelligence

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, technical AI, 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 creating trust in Human-AI teams, these research activities are, for a vast majority, carried within organizational structures explicitly identified by their disciplinary starting point: departments, communities, education, sources of funding, career progression ladders, gratification schemes. This disciplinary coloration is also deeply integrated by the researchers themselves who, often unconsciously, carry practices, values, and assumptions that are driven by their disciplinary scientific culture. Incidentally, disciplinary-centered organization creates visible and invisible boundaries that inherently hinder the innovative mindset needed for accounting for trust factors in Human-AI teams. How can research be (re)organized as to enable overcoming these boundaries?

This paper is dedicated to providing 1) an approach to science organization for overcoming these disciplinary boundaries through the example of the Socially-Aware Artificial Intelligence (SAI) Focus Area of TAIGA, the Center for Transdisciplinary AI of Umeå University, and 2) to show how such a structure can be put in practice for initiating new internal and external research collaborations on trust in human-AI teams by enabling the identification of research potentials (i.e. both research lines and motivated researchers) across the university as well as by laying out of financial and organizational structures for supporting the initiation of such collaborations.



Disciplinarity

- Within one academic discipline
- Disciplinary goal setting
- No cooperation with other disciplines
- Development of new disciplinary knowledge and theory

Multidisciplinarity

- Multiple disciplines
- Multiple disciplinary goal setting under one thematic umbrella
- Loose cooperation of disciplines for exchange of knowledge
- Disciplinary theory development

Interdisciplinartity

- Crosses disciplinary boundaries
- Common goal setting
- Integration of disciplines
- Development of integrated knowledge and theory

Transdisciplinarity

- Crosses disciplinary and scientific/academic boundaries
- Common goal-setting
- Integration of disciplines and non-academic participants
- Development of integrated knowledge and theory among science and society

- discipline ●
- non-academic participants ◉
- goal of a research project ○
- movement towards goal
- cooperation
- integration
- thematic umbrella
- academic knowledge body
- non-academic knowledge body

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

2 Organizing transdisciplinary research: Centers and Focus Areas

TAIGA [2], the Center for Transdisciplinary Artificial Intelligence of Umeå University, and its socially-aware artificial intelligence focus area, is organized as an answer to the limitation raised by classic disciplinary research organization. Briefly, TAIGA is a center dedicated to create, enable, develop, and foster transdisciplinary AI research, education, innovation, and social impact at the university level, taking as a starting point that 1) AI is considered as a transversal object of study that can be studied along many frames among which computer science is one of many facets; and 2) embedding AI research as a means for the social impact. Besides a coordinating core, TAIGA is organized along eight focus areas that arise from eight different social challenges, 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).

The SAI focus area[1], originating from the computing sciences department and taking a perspective of AI methods development as a starting point for the investigation, is the most related to trust factors in Human-AI teaming. The focus of the SAI focus area lies in the study of AI methods that allow for accounting for and adapting to human social behavior (i.e. human-centered SAI) and/or accurately simulating and replicating human social behavior (i.e. human-like SAI). Beyond technical concerns, the SAI focus area is dedicated to developing SAI methods as well as enabling a discussion across all groups of the university that are or can be interested by using SAI methods (e.g. psychology, pedagogy, medicine, arts) as well as study and govern the ramifications of developing and using SAI methods (e.g. sociology, organizational science, business, law) and sectors and actors beyond the university (e.g. businesses, regional authorities, other centers). As a general strategy for achieving this mission, the SAI focus area is dedicated to identifying, enabling, and supporting

[1] https://sites.google.com/view/taiga-socialai/home

the development of all the SAI potential niches available to Umeå University across all the missions raised by the university, including research, education, innovation, social impact, communication and reaching out to as many disciplines as possible.
This mission is carried out through an array of activities, including:

- *internal interactive activities*:
  - internal events (e.g. seminars, workshops, pitch events, networking events);
  - internal diffusion of available SAI methods (e.g. Natural Language Processing)
  - workshops for collection and diffusion of SAI-related interests (e.g. regulation of SAI) in visits tailored to the interests of the disciplinary departments of the university;
- *systematic structuring of internal SAI research:*
  - 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);
  - researchers-focused structures: compilation of expert lists, their topics and their former research productions [3], operating teams of researchers
  - systematic collection of past and future research interests
- *initiation of pedagogical activities:*
  - educational hackathons on AI for social good
  - preliminary steps for transdisciplinary SAI education at the university level
- *support for funding acquisition:*
  - promotion of relevant sources of funding to the interested community;
  - organization of pitch events
  - offer of starting grants (seed funding) for enabling larger grant-writing;
- *academic outreach* (national and international dissemination and outreach)
  - international workshops on SAI topics (e.g. Interdisciplinary Design of Emotion Sensitive Agents Workshop at the Autonomous Agents and Multi Agent Systems Conference 2023 [4])
  - international special interest groups on SAI-related topics (e.g. AI for crisis response [5])
  - special tracks on SAI conferences (e.g. human-like deliberation and deliberation during crises at the Social Simulation Conference 2023 [6] [7]);
- *public outreach*
  - organization and participation in general public seminars (e.g. the frAIday seminar series [8], a weekly seminar that regularly reaches more than a hundred of participants);
  - 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 SAI focus area revolves around creating qualitative spaces for long-term collaboration rather than around "short-term" classic criteria such as number of papers per year (which are better fitted once groups are operating). This approach manifests in the SAI focus area operations. For example, the yearly "seed" funding call, dedicated to offer a starting small 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 staring up of collaborations rather than the other way around.

3 SOCIAL-AI FOCUS AREA FOR ENABLING TRUST IN HUMAN-AI TEAMS RESEARCH

The aforementioned missions can be put in motion towards best enabling internal and external collaboration on trust factors in Human-AI teams research. Taking the lenses of organizing scientific production, assessing, and streamlining the internal potentials for an academic institution to carry out a given line of research are a highly many factorial considerations: not only that one should assess the relevance of the research carried by the institution, but also whether and how researchers (and subsequently, personal time, skills, interests, resources) can be mobilized in new collaborations and many others, such as infrastructural capacity, financial capacity, legal capacity.

From a research-oriented perspective, a review of the SAI focus area-relevant research 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, brings 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 it entails when developing Human-AI systems [9] [10]. 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 the involved humans [11], the integration of norms the system should best try to comply to when interacting with humans [5], the development of intuitive man-machine interaction media in which humans can advise the system on (un)desirable courses of actions [13] [14] and more general frameworks on man-robot teams [15]. A third line of research considers broader frames, bringing forward trustworthy AI considerations and in particular on the topic of transparency and explainability [16] [17] [18].

From a researchers-oriented 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 are focusing on the trustworthy/explainable AI track, either from a technical perspective or a sociological perspective; and one profile is 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 bodiless AI and embodied (robotic) systems.

From a resource-oriented 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 actors to consider prospective collaborations along the line of trust in H uman-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, who acts as a facilitator for such collaborations.

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 the topic of trust in Human-AI teaming, by describing one of such structures, namely the Socially-Aware AI focus area organized under 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 touched upon at 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 academics 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). In the case of trust in Human-AI teams, the missions covered by the SAI focus area described in this paper allow effectively identifying prospectively connectable existing research lines (namely, interdisciplinary psychology-grounded perspective on trust building; technical perspectives on trustworthy AI; technical perspectives on adjustable autonomy), as well as providing fast tracks for turning potentially interested internal or external groups into working collaborations (e.g. fundings, visibility, meetings). The options being laid out, now is the window of opportunity for internal and external actors from all disciplines to step in and enable the scaling up of this research.

REFERENCES

[1] G. Tress, B. Tress and G. Fry, “Clarifying Integrative Research Concepts in Landscape Ecology,” *Landscape Ecology,* vol. 20, pp. 479-493, 2004.

[2] Umeå University, “Umeå University's center for transdisciplinary for the good of all (TAIGA),” Umeå University, 01 10 2022. [Online]. Available: https://www.umu.se/centrum-for-transdisciplinar-ai/. [Accessed 14 11 2023].

[3] L. Vanhée, “Social AI research webpage,” TAIGA, 14 11 2023. [Online]. Available: https://sites.google.com/view/taiga-socialai/research. [Accessed 14 11 2023].

[4] M. Borit and L. Vanhée, “Interdisciplinary Design of Emotion-aware Agents (IDEA) workshop,” 01 06 2023. [Online]. Available: https://en.uit.no/project/idea. [Accessed 01 06 2023].

[5] C. Kammler, F. Giardini and L. Vanhée, “Building ResIllienCe with Social Simulations (BRICSS) special interest group,” 01 04 2023. [Online]. Available: https://sites.google.com/view/bricss/home. [Accessed 14 11 2023].

[6] European Social Simulation Association, “Simulating in Crises special track at the Social Simulation Conference 2023,” European Social Simulation Association, 01 07 2023. [Online]. Available: https://ssc23-sphsu.online/simulating-in-crises/?et_fb=1&PageSpeed=off. [Accessed 14 11 2023].

[7] European Social Simulation Association, “Sense and Sensibility special track at the Social Simulation Conference 2023,” European Social Simulation Association, 01 07 2023. [Online]. Available: https://ssc23-sphsu.online/sense-sensibility/?et_fb=1&PageSpeed=off. [Accessed 14 11 2023].

[8] Umeå University, “FrAIday webpage,” Umeå University, 14 11 2023. [Online]. Available: https://www.umu.se/en/research/our-research/features-and-news/artificial-intelligence/fraiday/. [Accessed 14 11 2023].

[9] M. Borit, L. Vanhée and P. Olsen, “Understanding the impact of culture on cognitive trust-building processes: How to increase the social influence of virtual autonomous agents,” in *Trust in Agent Societies*, Paris, 2014.

[10] M. Borit, L. Vanhée and P. Olsen, “Towards enhancing trustworthiness of socially interactive and culture aware robots,” in *International Workshop on Trust in Agent Societies,*, Paris, 2014.

[11] L. Vanhée, L. Jeanpierre and A. I. Mouaddib, “Optimizing Requests for Support in Context-Restricted Autonomy,” in *International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[12] L. Methnani, “Embracing AWKWARD! A Hybrid Architecture for Adjustable Socially-Aware Agents,” Umeå University, 2022.

[13] L. Vanhée, L. Jeanpierre and A. I. Mouaddib, “Augmenting Markov decision processes with advising,” AAAI Conference on Artificial Intelligence, 2019.

[14] L. Vanhée, L. Jeanpierre and 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.

[15] M. Chiou, S. Booth, B. Lacerda, A. Theodorou and S. Rothfuß, “Variable Autonomy for Human-Robot Teaming (VAT),” in *International Conference on Human-Robot Interaction*, 2023.

[16] R. H. Wortham and A. Theodorou, “Robot transparency, trust and utility.,” in *Connection Science*, 2017.

[17] D. Calvaresi, Y. Mualla, A. Najjar and S. Galland, “Explainable multi-agent systems through blockchain technology.,” in *Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, 2019.

[18] S. Knapič, A. Malhi, R. Saluja and K. Främling, “Explainable artificial intelligence for human decision support system in the medical domain,” in *Machine Learning and Knowledge Extraction*, 2021.