ANNUAL REPORT

The Hybrid Intelligence Centre - 2021

Augmenting Human Intellect

Contents

01 Executive Summary

02 Highlights

03 Internal Collaborations

04 The Organisation

05 Education & Training

06 Finance

07 External Outreach

08 International Conference

09 Ambitions Year 3

10 PhD/PD Projects

Executive Summary

The year 2021 was the 2nd full year of the Hybrid Intelligence centre. ..

Most of our PhD students and postdocs have now been active on their projects between 1 and 1.5 years, and we have spent around 12% of our 10 year budget. Despite much of 2021 still having been spent under COVID restrictions, we have made substantial progress on each of our 32 individual projects, on consortium-wide topics that cut across these individual projects, and on outreach to the wider community.

Individual projects Our PhD students and postdocs have made progress on their projectst through sharpening and operationalising their research questions, writing literature surveys, defining experiments to be performed, and obtaining some first theoretical and empirical results. Remarkably, these early-stage projects have already resulted in 22 publications in international venues. Our first F2F meeting in September was entirely dedicated to highlighting these individual projects through PhD pitches and poster presentations.

Consortium-wide Our Research Lines working groups (on Collaborative, Adaptive, Responsible and Explainable HI, CARE) continued their regular meetings throughout the year (see section 3). In our cross-thematic Taskforce, and in our online plenary meetings, we have been addressing the consortium-wide questions that cut across the individual projects: our March meeting was dedicated to application areas of Hybrid Intelligence, our July meeting was dedicated to metrics for Hybrid Intelligence, and the November meeting was spent developing a shared terminology for Hybrid Intelligence, which lead to a working group aiming to further develop and publish such a terminology

Community-wide Our Research Agenda (IEEE Computer, 2021) continues to influence the community with multiple new citations each month, and was elected “paper of the year” by the IEEE Computer journal. We continued to provide educational material to the community on our website and we organised two courses for the Dutch national graduate school for Information and Knowledge Systems (SIKS), on neuro-symbolic AI, and on Explainable AI.

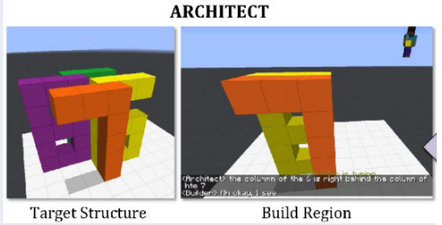
A major new development has been that the University of Twente has joined the Hybrid Intelligence Centre in the autumn, with a team of 3 staff members and 2 PhD students participating in our activities and contributing to the Dutch community.

As an important part of our effort to contribute to the international community, we are hosting a new conference series on Hybrid Human-AI, HHAI 2022, with the first edition hosted at the VU campus in Amsterdam on 13-17 June 2022.

In collaboration with TNO, the Dutch national organisation for applied research, we have actively participated in the Dutch “AINed programme”. As a result, the topic “Hybrid AI systems” has become one of the six focus areas for this major injection of funding in AI by the Dutch government.

Highlights

We published 22 papers in a variety of international venues



A team of four PhD students won the IGLU competition (Interactive Grounded Language Understanding in a Collaborative Environment) at NeurIPS, a challenge for multi-agents collaborating on a construction task in a synthetic environment.

We have initiated a new international conference on Hybrid Human AI Collaboration (HHAI 2022), to be held 13-17 June 2022.

We received the paper of the year award from the IEEE Computer Journal for our paper "A Research Agenda for Hybrid Intelligence

IEEE COMPUTER SOCIETY

2020 Best Paper Award

02

Two of our teams jointly defined a large-scale experiment (n=300) on partner choice in human collaborations, supported with 14k€ funding from the Hybrid Intelligence Centre.

03 Internal Collaborations

Collaborative Hybrid Intelligence

Hayley Hung and Koen Hindriks

Achievements We have had collaborative line research meetings roughly every 6 weeks in a virtual format. Each of these meetings aimed to start to establish common ground between the participants by asking each of the PhD students to present their work. We managed to have all the PhD students present, including also students who are also guests with this research line. This regular meeting highlighted how diverse the research line is. We have only just started to touch the tip of the iceberg in terms of understanding each other's perspectives. Lively discussions were had after each presentation and suggests that we have a good format for continuing to understand each other. In an effort to understand each other further, we also submitted the HAI Workshop on the representation, sharing and evaluation of multimodal agent interaction at the first Hybrid Human Artificial Intelligence (HHAI) conference.

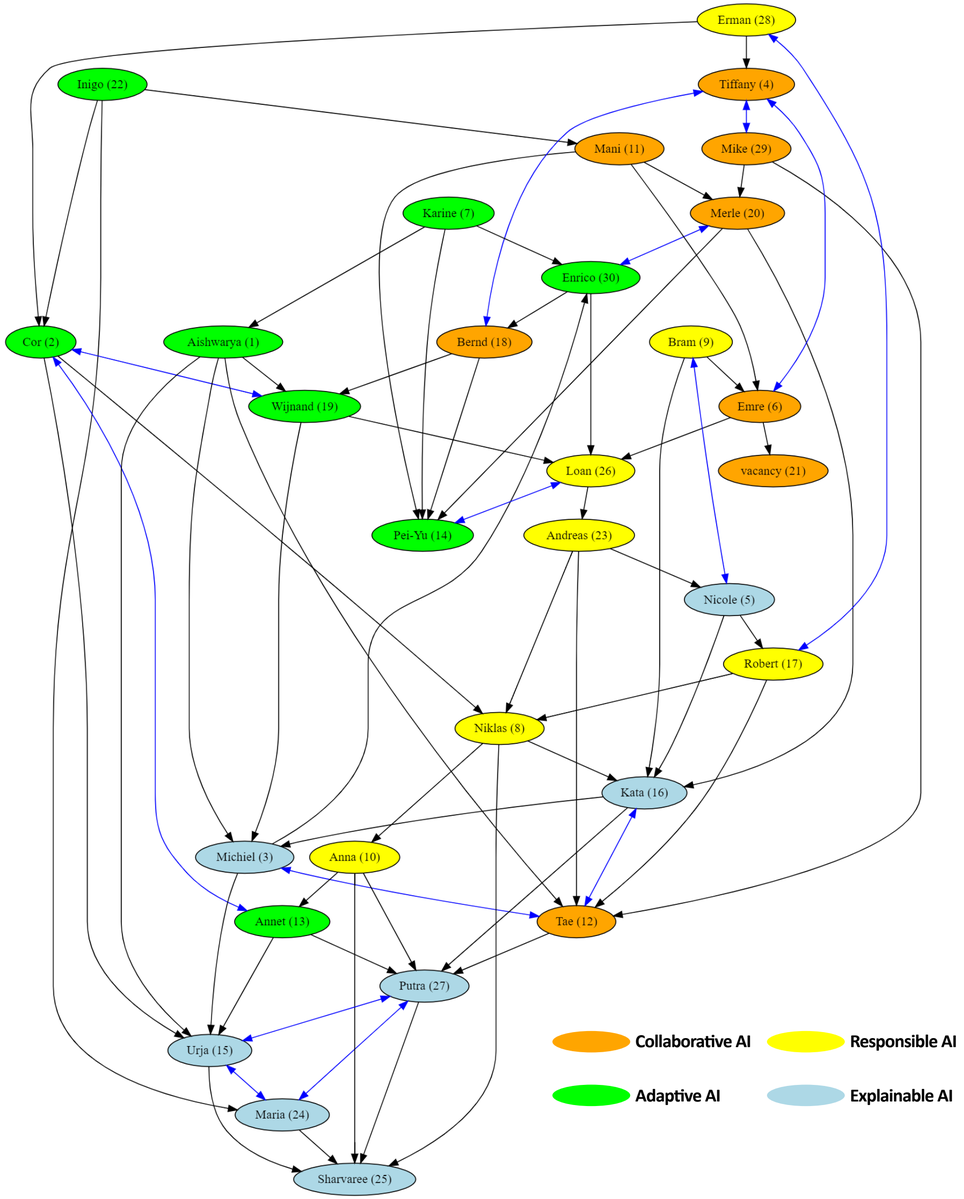
Plans We will continue to meet virtually every 6 weeks. We will start to integrate more presentations from faculty members so as to continue the perspective sharing. We will also continue to have PhD students present their progress. We will use the in-person consortium meetings as an opportunity to meet in person every few months.

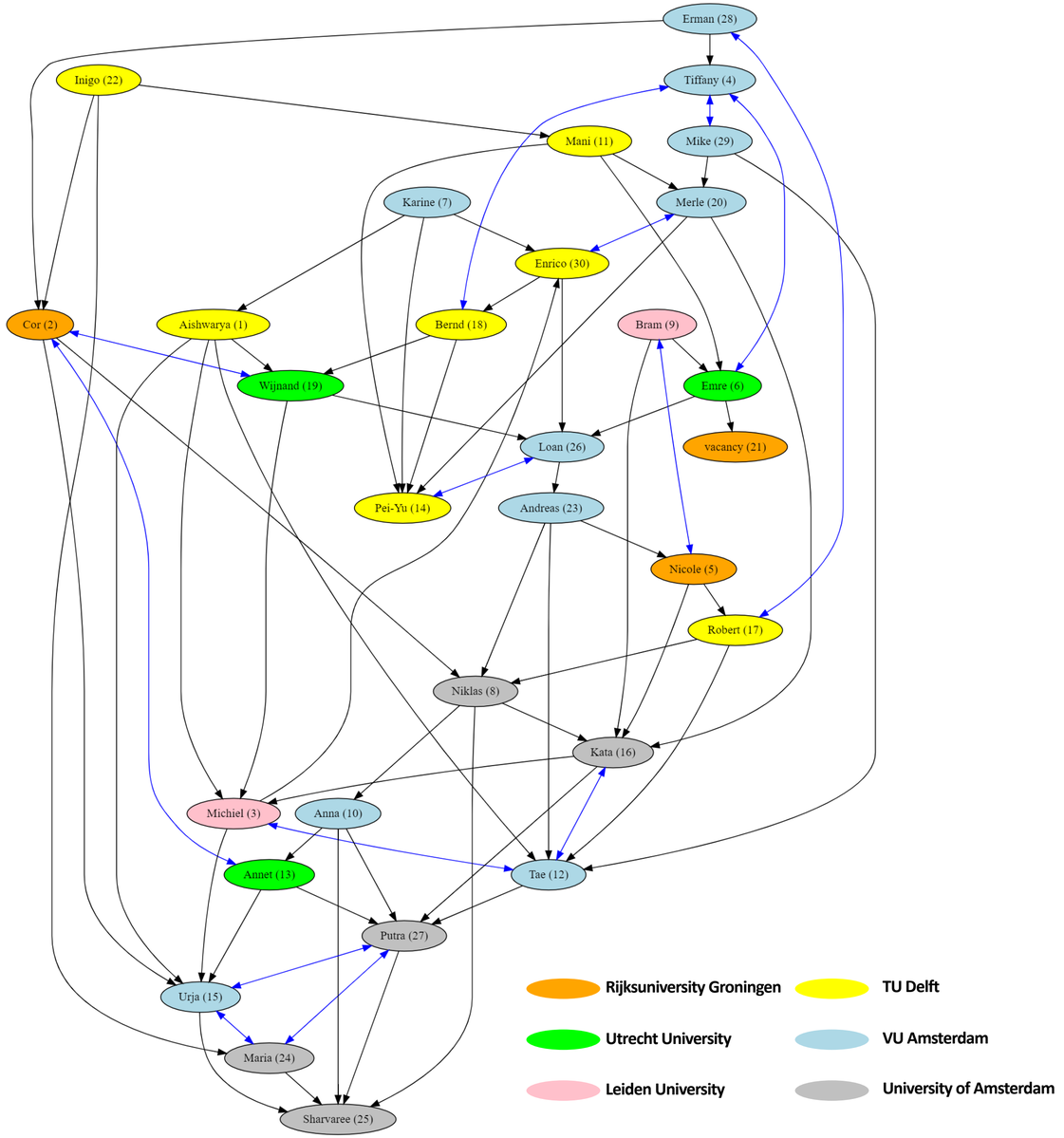
Explainable Hybrid Intelligence

Pradeep Murukannaiah, Piek Vossen, and Antske Fokkens

Achievements The main objectives of the Explainable Hybrid Intelligence (XHI) research line (RL) are: to develop the foundations of explainability and interpretability for hybrid intelligence (HI), and to explore the commonalities among different PhD projects within the RL and across other RLs in order to foster collaboration among the HI consortium members. To meet these objectives, the XHI RL members meet regularly (every six weeks) and discuss topics such as: why explanations are needed in HI systems, what explanations mean in specific applications, and how different projects in the RL facilitate explainability. The key themes that have emerged as answers to the why question are: to improve model performance, to verify whether a model is generalisable, to enhance users' trust in the model, to facilitate collaboration between humans and AI, and to make a model auditable. This shows the variety of topics that are being explored in the XHI RL. In addition to the regular RL meetings, some of the RL members (especially, PhD students) participate in a self-organized study group, where they discuss research papers about explainability. Finally, to foster collaboration across RLs, the XHI RL and the Adaptive HI RL jointly organized a consortium meeting. In this meeting a key activity was to explore the HI-ness of specific applications from the perspectives of adaptivity and explainability. The activity showed both the potential for improving AI applications by transforming them as HI applications as well as the differences in our understanding of HI-ness.

Plans In the upcoming year the XHI RL plans to continue regular RL meetings and the explainability study group. Further, if an opportunity arises, the RL will plan a joint meeting with another RL. Whereas the focus in the previous year has been on the why question, the RL would like to focus also on the what and how questions in this year. Finally, this year, the RL would like to revisit its idea about writing a white paper on a topic of shared interest.





These collaboration graphs between our projects show that the research lines are fruitful platforms for collaboration (left graph),and that collaboration is frequently across university boundaries (right graph)

Responsible Hybrid Intelligence

Bart Verheij and Birna van Riemsdijk

Achievements Research has really started. Aishwarya Suresh, Annet Onnes, Cor Steging, Íñigo Martínez de Rituerto de Troya, Pei-Yu Chen, Urja Khurana and Wijnand van Woerkom have started their work in the 7 current projects in the research line. We have organized several research line meetings in which we are starting to get to know each other, and exploring our various perspectives on the notion of responsibility. The project-wide meeting `Hybrid Intelligence: the good, the bad and the ugly' that we organized in December 2020 has led to a series of 2-pagers on various application settings of HI (Recommendation; Hybrid teams; Healthcare; Social media; Democracy, politics, law). We have distilled first thoughts about technological approaches to good HI, and an initial set of guidelines for what is needed for doing responsible research in the HI field. Moreover, collaboration between HI and the gravitation programme on Ethics of Socially Disruptive Technologies (Esdit) has been initiated. A committee has been formed to stimulate this collaboration, which organized an inspiring and well-attended kickoff meeting before summer. Another HI-Esdit workshop was organized in October on the topic of agency, and a joint panel took place during the 4TU.Ethics bi-annual Conference on October 14th/15th (with Bart Verheij and Pinar Yolum Birbil). A workshop on interdisciplinary collaboration was planned, but had to be canceled due to Covid measures. We hope to organize this in 2022.

Plans While the past year has centered around getting to know each other and the PhD projects in the line, in the coming year we plan to focus more on interaction and discussing the relations between the projects within the line and with projects in other lines. We plan to achieve this via our research line meetings, possibly extended with informal gatherings with the PhD candidates. A SIKS course on Responsible AI will be organized in June 2022, and we will organize a workshop as a follow-up of the 4TU.Ethics panel as part of the HHAI conference programme. We plan to further develop the guidelines for responsible research in HI together with Esdit colleagues, and as part of the HI project meeting planned in September.

Deviations We had one change in the composition of the PhD projects: Nicole Orzan (project 5) moved to the Explainable HI line. A challenge may be that the types of techniques used by PhDs in the line differ greatly among projects, varying from highly technical to highly philosophical. Nevertheless, there is common interest regarding the theme of responsibility, providing sufficient common ground. We will focus discussions on this common aspect, and consider reaching out to other lines to find common interests, for example regarding conversational agents.

Adaptive Hybrid Intelligence

Frans Oliehoek and Herke van Hoof

In HI settings, artificial agents and human agents work together, possibly on multiple tasks, in complex environments. Such settings are seldomly static: team composition and tasks can change, preferences can shift and external conditions (e.g., available resources, and environment) can vary over time. Thus, the competences in terms of theory of mind, collaboration, task execution, et cetera cannot be fixed before deployment. The agents will have to adapt and learn in operation. Moreover, members of the team get to know each other better over time. As such, the ability of HI systems to adapt or learn is a prerequisite not only for them to perform well, but for them to function at all. To accomplish such adaptivity, the agents need to deploy machine learning techniques to learn from data, from experience and from dialogues with other agents (human or artificial).

A1. Learning through interaction. This task investigates different types of learning from interaction with others, such as competitive learning, learning from demonstration, coaching, and collaborative learning. Different modes of interaction are studied such as through rewards, preference information, or in symbolic form. An example is project 17- (Meta-) learning to give feedback in interactive learning. Over the past year a theoretical model has been developed of reinforcement learning for human-AI interaction. The main theoretical result shows that it is impossible to guarantee effective interaction without relatively strong assumptions about the behavior of other agents. The goal of this work is to provide the theoretical foundations for algorithms that learn online how to work with and communicate with other agents.

A2. Learning how to interact. This task investigates how agents can learn to communicate effectively with humans through different modalities, including adapting to an individual or group and maintaining a representation of the knowledge of interaction partners. An example is project 27- Knowledge Representation Formalisms for Hybrid Intelligence. The current focus is on using rule-based languages and argumentation frameworks to deal with inconsistent information by non-representation of knowledge and by the user's motivation. The framework can allow HI systems to adapt dynamic knowledge based on the user's changing preferences; and allows giving explanations to users. Currently a paper is in preparation to be submitted to the HI conference 2022.

Another example is project 9- AutoAI for Dynamic Data. Work is in progress on constructing agents that learn how to negotiate.with other agents under a fixed negotiation protocol. A key aspect that was pursued last year is to let the agent learn about other agents and use this information to adapt its policy in search for better results. Results are published at the Cooperative AI workshop @ NeurIPS and at AAMAS 2022.

A3. Incremental learning in changing conditions. This task studies how techniques such as AutoML and meta-learning can be used to adapt in a 'few-shot' fashion to changes in the composition of groups, the task, the environment, preferences, and resources. An example is project 10- Continual Learning and Deep Generative Modeling for Adaptive Systems. During the past year work has focused on the robustness of VAEs to minor shifts in data distribution. We have published a paper (ICLR RobustML workshop) where we focus on adversarial attacks: their construction and evaluation metrics. Currently, we are working on a method which can be used to defend VAEs from the attacks and, therefore, improve the model robustness.

A4. Integrating learning, reasoning and planning. This tasks studies how deep learning techniques can be coupled with symbolic representation, reasoning, and planning approaches in order to benefit from symbolic prior knowledge and to respect symbolic constraints (e.g. societal, ethical, legal, safety); as well as to improve transferability, explainability, and predictability of learned models. An example is project 8- Hybrid approaches to learn explainable sequential behaviours. Work has proceeded on integrating knowledge present in a knowledge graph for improved sample efficiency and generalization performance in knowledge-intensive domains such as commonsense games. Specifically, the subclass relation has been investigated and a state abstraction has been formulated, such that an agent could learn behaviour that holds for all objects in a class, even unseen ones. An example is project 23 - Causal Reinforcement Learning and Reasoning. A paper has been submitted on causal discovery with meta-learning in RL which is currently under review. The current focus is on a literature study on model-based RL to improve this approach. This research is important to build models which are robust against causally localised changes in the environment such as the ones caused by humans.

03

Plans In the past year we have actively explored the question of how to measure progress of the adaptiveness of HI systems. This is an ongoing effort. Additionally, we are intending to continue our efforts in identifying how the individual project contribute to the overall goals of HI, thus also identifying possible synergies and opportunities for collaboration.

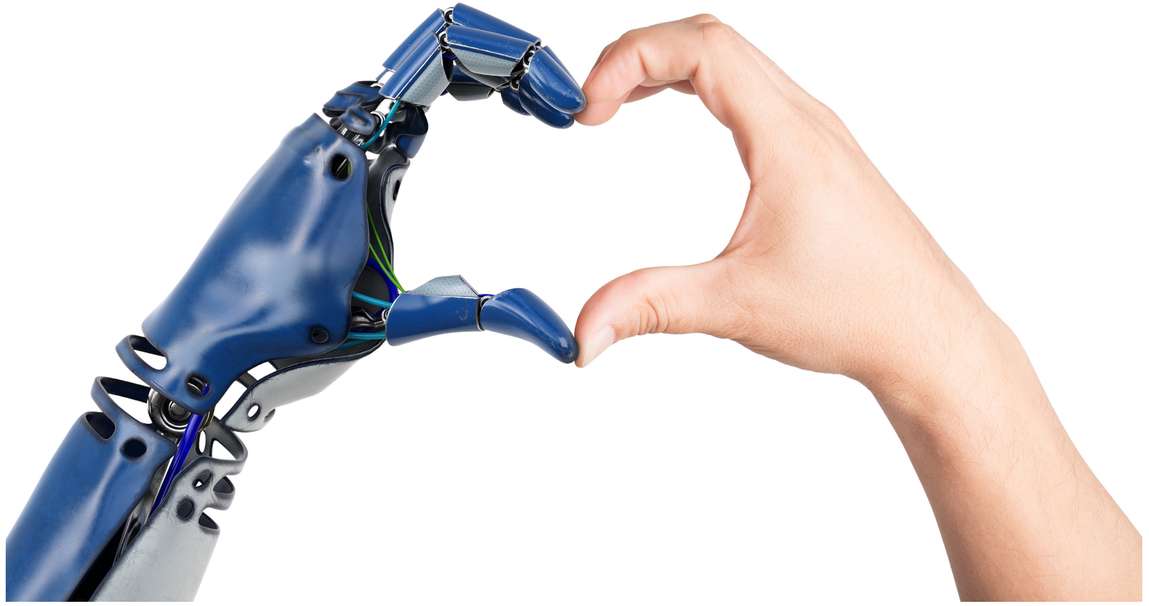
Research line meetings:

Collaborative This RL is meeting every 6 weeks to establish common ground and share project progress the attendance in the group is high.

Explainable Regular meetings are scheduled every 6 weeks and are theme-driven

Responsible Getting to know each other, brief introductions by all PhD students and postdocs discussion the meaning of Responsibility and the organization of RL line is done in quaterly meetings

Adaptive In the April meeting we had short presentations by all of the projects, and discussed the organization of the research line. People indicated that they are hoping to really foster collaboration, and as such prefer physical meetings if at all possible. In the September consortium meeting, we got the chance to meet up physically at VU, and participated in the poster session. The November research line meeting focused purely on the progress of the adaptive research line. We had four presentations by projects, and an invited talk. While the number of participants was a bit lower than we hoped due to last minute cancellations (possibly also COVID related), the participation was very good, and it is clear that the projects are starting to make progress.



Evaluation Taskforce

Wendy Aartsen, Davide Grossi, Catholijn Jonker, Pradeep Murukannaiah, Catharine Oertel, PInar Yolum,

The Taskforce of the Hybrid Intelligence Centre (HIC) has adopted the challenge to make HIC the place to experience how a human and a machine add up to more than two.

The Hybrid Intelligence consortium has the ambition of common grounding on knowledge of HI-ness measures/ metrics/methods through defining constructs that create the HI discipline by:

- Connecting disciplines & research lines (hackathons, roadmaps, prototypes)
- Providing environments and use cases to experience HI-ness
- Sharing research and artifacts (data sets, benchmarks, experiments, methods)

Demonstrators: the TF started as a sequel to the quest for demonstrators that would serve the consortium as case studies to test and showcase our scientific findings. The demonstrator sessions of 2020 resulted in a number of potential scenarios. As a follow-up we asked volunteers from within the consortium to our first TF meeting, organized online and with a miro-board to create and maintain our ideas. However, it quickly turned out that the group (and in fact the consortium as a whole) still struggled with what concepts and criteria would best describe hybrid intelligence. The MT decided to start afresh with some senior researchers to lead the TF by preparing activities with the consortium to determine those concepts and criteria and by doing research into what methods are suitable to develop HI-systems.

The taskforce leaders have since then accomplished the following:

- Determine a vision, mission, and research agenda for the TF, underwritten by the EB
- Embedded the TF into the organizational structure of HI
- Found approval for hiring a Postdoc to boost the research of the TF
- Developed a plan for the upcoming consortium meetings to strengthen the research cohesion across the research lines while working on the TF mission and research

Taskforce meetings: Bimonthly meetings to set up of TF activities, brainstorming about ideas

Plans Consortium meeting February 2022: Knowledge technology exercise to establish the start of a HI vocabulary to be used as a shared ontology for HI research and establishing connections between the individual HI projects.

Consortium meeting March 2022: test HI criteria against projects within the consortium, gather ideas for improving the HI-ness (the measure of levels of a system showing forms of Hybrid Intelligence), which can serve as the basis for HI development methodologies.

With the PD review of suitable methods of developing HI systems and/or improving HI-ness of existing systems. This effort is meant to result in a white paper.

At the Conference HHAI: support the organisation of a Hackaton on HI systems.

Consortium meetings autumn winter 2022: continue with strengthening the cohesion across RLs by investigating deeper into the way individual projects can help each other.

Themes of the consortium plenary meetings:

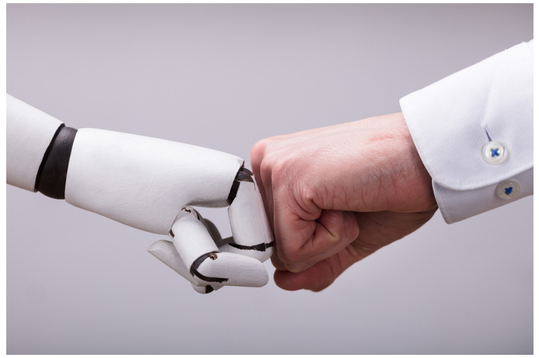
March: making storyboards of HI applications

July: How to measure Hybrid Intelligence?

Sep: PhD's in the spotlight: PhD pitches, posters

Consortium

03



Governance

In 2021, the team was completed and most PhDs are now at the end of their first year (for some recruitment was earlier 1,5 years and for some later 0,5 year). The Management Team meets weekly, and prepares the monthly meetings with the Executive Board (EB), as well as the yearly Scientific Advisory Board meeting and the Governing Board meeting, held on the15th of April 2021 and 23rd of June 2021 respectively.

The MT team is meeting the RL cooridnators each month to prevent silo formation and learn from their experiences while working together. We should keep rethinking our unique identity that contrasts with other competing projects Keep focus on the foundational and long-term sceince perspective like: “what does HI mean?” and “what are the limits?”

The taskforce and the MT team are currently facing the challenge of bringing together all the consortium projects into a more applied setting like “shared experimental testbeds”, using instruments like Hackathons and involving human participants in the experimental settings.

Building the community

We consider it our mission to build an active national Hybrid Intelligence community that extends beyond the partners currently involved in the Hybrid Intelligence Centre. We have extended our consortium with the participantion of the TU Twente as associated partner. The team joins with an interest in virtual reality, user interfaces, socially intelligent computing, conversational agents,among others.

UNIVERSITEIT TWENTE.

The consortium agreement was signed expressing the willingness to have PhDs, postdocs and staff join in the HI programme by joining the consortium meetings, RL meetings, taskforce as shared experimental testbed, reading club and have two PhD/PD projects co-supervised.

Collaboration with ESDiT

We have engaged with a sister consortium funded by the Gravitation, the programme on Ethics of Disruptive Technologies (ESDT). Apart from jointly co-supervising one of our PhD students, the HI-ESDiT collaboration kicked-of on the 18th of June 2021 with a 2 hour online workshop. The 31 participants were equally distributed over HI and ESDiT.

A shared Slack channel has been installed for the HI-ESDit Collaboration, where we announce upcoming events, presentations and workshops with shared themes of interest, like:.

1. Workshop on agency, October 11th 2021
2. Panel 4TU.Ethics conference on AI, human values and the law, October 14th/15th 2021
3. Workshop vulnerability in HCI, December 2nd 2021
4. Presentation in ethics working group of NLAIC, December 14th 2021 by Bart Verheij and Birna van Riemsdijk
5. Start of the STOA report on ethics of AI and HI

Staff leaving

Due to career moves some of the consortium members have left

1. Prof. Max Welling has left the HI EB, when he accepted a career move to Microsoft Research

Staff arriving

Two new members joined the HI EB:

1. Prof. PInar Yolum (UU)
2. Prof. Joost Batenburg (ULEI)

We have attracted new members from within participating organisations and newly joining organisations:

3. Dr. Shenghui Wang (TU Twente)
4. Prof. Dirk Heylen (TU Twente)
5. Prof. Aske Plaat, Dr. Max van Duijn, Dr. Daan Pelt and Dr. Joost Broekens (ULEI)
6. Dr. Andrew Yates (UvA)
7. Dr. Shihan Wang (UU)

The Organisation 04

SWOT Analysis

04

STRENGTHS

1. **Diversity of our consortium** in gender, generation and nationality
2. **Strong results:** high numbers of (award winning) papers
3. **Due to a strong and attractive proposition the consortium grows** with new collaborations, partners and individuals
4. Strong **governance structure** with an ambitious Executive Board, committed Educational Director, Management Team and RLC's, with enough redundancy for resilience against staff change.
5. High ambition towards **coherence and commitment** between and to the individual projects
6. Organisation of the first **International Conference on Hybrid Human-Artificial Intelligence** HHAI
7. **Strong educational programme**, with a dedicated educational director and in collaboration with the national **Graduate School**

WEAKNESSES

1. **Lack of coherence** between 30 individual projects and overall project aims
2. Slow internal onboarding and challenging internal communication
3. **International outreach programme** has not been defined yet
4. **Managerial capacity** needs to ramp up and follow the consortium growth
5. A **small fraction** of our staff (mid-career) are not (yet) well connected to the main research goals (approx. 15%)

OPPORTUNITIES

1. **AI climate**, Hybrid Intelligence is currently a **hot topic** worldwide (decision makers, scientists, journalists): greater visibility, easier recruiting, opportunities for growth
2. **Building a larger community**, collaboration with other (current and future) **non-academic** partners
3. Increase international **outreach** and apply for additional **funding**
4. Initiate an **International programme for HI Education & Teaching**

THREATS

1. **Reduced unique selling point (USP)** because of the growing international HI community
2. **Strong competition** due to global increase in related initiatives
3. **War on talent,** staff moving to other institutions in a very competitive labour market, difficulties in recruiting new talent

Actions needed

1. Bring management capacity back in line with size of the consortium
2. Use new funding round to improve coherence and connectedness of mid-careers
3. Apply for funding for international training & teaching programme

Education & Training

05

Course development

Much of the PhD training in AI in The Netherlands is organised by SIKS, the national School for Information and Knowledge Systems. We organised 2 out of the 11 SIKS courses in 2021, with one course on AI systems that combine learning and reasoning (also known as neuro-symbolic systems), and one course on Explainable AI. Both courses were attended by many of our own PhD students, but also by a substantial number of PhD students from outside the Hybrid Intelligence Centre, including students from universities not affiliated with our Hybrid Intelligence Centre.

PhD presentations

We continued the successful format of the bi-weekly online “reading club” for PhD students. With an increasing number of our PhD students now making progress in their own research, the format has shifted into a mixture of reading papers from the literature and students pitching their own work to each other. This is both a useful training experience for them, as well as a great launching pad for collaborations between our PhD students. For the same reason, all of our PhD students and postdocs presented a poster at a live poster-market during our first-ever in-person meeting in September.

PhD progress monitoring

All PhD students produced a 1-page report at the end of 202 in a standard format (“what are your research questions?”; “how is your progress?”, “how does your work fit in the overall HI program?”). Extracts from these 1-page reports are included as the final chapter in this report. As is customary at Dutch universities, all PhD’s who have been employed for a full year have had a progress-advice meeting with their 1st (internal) and 2nd (external) supervisors (every PhD student in the Hybrid Intelligence Center has a 2nd supervisor from another participating university).

Preparing for internships

To facilitate internships for our PhD students in their later years, we have signed letters of intent for this purpose with a consortium at RWTH Aachen University working on HI for Management (prof. van der Aalst), with a consortium from the Universities of Stuttgart and Augsburg working on Intelligent Human-Computer Mediation (profs. Staab and André), and with the European Humane AI Net project (lead by prof. Lukowicz). Of course our sister Zwaartekracht initiative ESDIT on the Ethics of Socially Disruptive technologies is also a partner for future internships. We are currently building similar relationships with partners from industry.

Realisation Hybrid Intelligence 2021

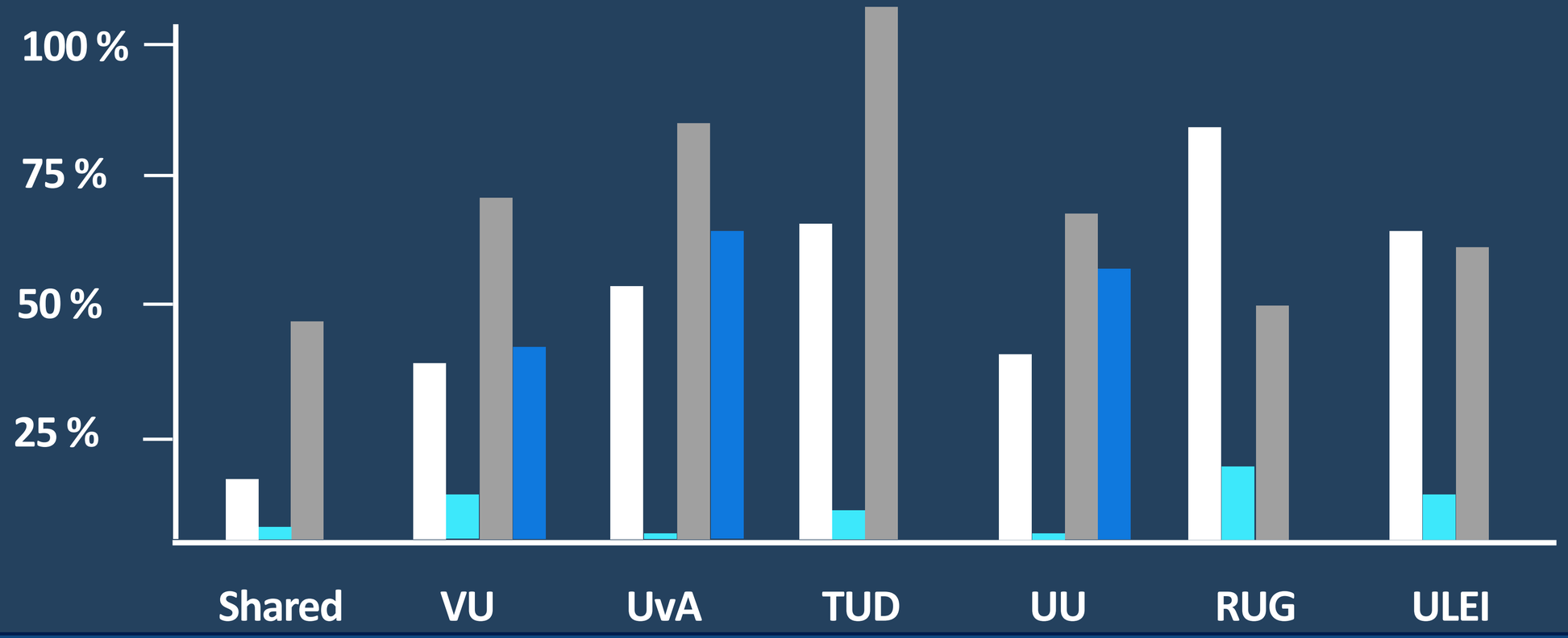
Due to the COVID pandemic the Hybrid Intelligence consortium has mainly operated online for the first two years. As a consequence the central resources reserved for meeting costs have not been invested as reflected in the underspending. This effects the materials costs of individual partners in a similar way, as travelling between locations has been heavily reduced.

Overall the expenses on personnel costs (NWO contribution, in kind contribution or in cash contribution) is delayed as the first year has been focussed on getting the team complete and recruiting the first batch of PhD students and postdocs.

Investments will be done on the recruitment of a postdoc at UU that will be a dedicated function aligned to the taskforce with the mission: “how to measure HI” and on a large collaboration-formation experiment.

Finance

06



07 External Outreach

We made substantial efforts to communicate about our scientific work to the general public and to society at large.

The VU produced a high-quality video on “Building A.I.nstein- the artificial scientific collaborator”, a possible moonshot for our project.

We published a Dutch-language interview on the widely read tech-magazine Tweakers.net: “Mens en intelligente machine- Niet vervangen, maar samenwerken” (See here for a version that’s not behind a pay-wall)

We contributed to the Dutch national AI course, a free resource for the Dutch public..

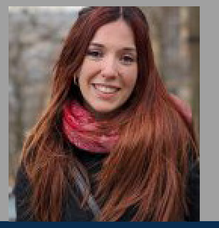
Catholijn Jonker presented for the Justice Committee of the Upper House (“Eerste Kamer”) of the Dutch Parliament.

We contributed to “NWO Insipireert”, an event organised by the Dutch science foundation NWO, aimed at science communication professionals.

We were one of the debate-leaders at the “Evening of science and society”, an event for thought leaders from politics, culture, science and industry, organised by the Dutch ministry of Science and Education.

08 International conference

Stefan Schlobach


Maria Perez-Ortiz

Myrthe Tielman

HUMANE AI NET

Artificial Intelligence
www.elsevier.com/locate/artint

SIKS

TNO innovation for life

NETWORK INSTITUTE

IOS Press

HHAI2022

In order to strengthen the position of Hybrid Intelligence on the international research agenda, we have initiated a new conference series under the name “Hybrid Human-Artificial Intelligence” <https://www.hhai-conference.org/>

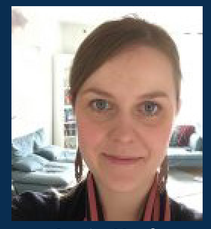
In order to facilitate community building, HHAI 2022 will be a single track conference, preceded by workshops, and allowing not only for regular papers, but also for working papers and extended abstracts, as well as demo’s and posters.

The conference is chaired by Stefan Schlobach (VU), with Maria Perez-Ortiz (UCL) and Myrthe Tielman (TUDelft) as program chairs with proceedings published as Open Access by IOS Press.

The conference is co-organised by the Hybrid Intelligence Centre and the European Humane AI Net initiative. The interest in such a new conference series is witnessed among other things by the over 80 submitted papers organised by the poster and demo Chairs Shenghui Wang (TU Twente), Janine Koch (INRIA), 8 workshops guided by workshop chairs Ana Valdivia (KCL), Roel Dobbe (TU Delft), a hackathon organised by TNO Jasper van der Waa (TNO/TU Delft), as well as the list of sponsors, which include Elsevier, Huawei, IOS press and Sony, among others.

HHAI 2022 will take place 13-17 June 2022 in Amsterdam, local chairs and PR Ilaria Tiddi and Romana Pernisch (VU)

Shenghui Wang


Janin Koch

Ana Valdivia

Roel Dobbe

Jasper van der Waa

Ilaria Tiddi

Romana Pernisch

Ambitions Year 3

09

Individual projects

Our main ambition is the success of all of our 32 PhD students and postdocs. We have achieved all our goals from last year's annual report on our individual projects: all PhDs have a research plan, are embedded in a local team, collaborate in at least one research line, and participate in our PhD community. By the end of 2022, we expect all of our PhD students to have at least 1, preferably 2 publications in substantial venues on their way to their PhD degree. We also expect a number of PhD students to have started an internship at an external organisation.

We will aim for a number of "dovetail" collaborations, where the results of one project (data, algorithms, code, theoretical results, experimental results) will be used as input for the work in another.

Consortium-wide

In line with our plans from last year, we have successfully established four active research lines.

Progress in building shared resources such as benchmark datasets, simulator-environments, exchange formats and protocols, and standardised experiment designs, is slower than we had hoped. In 2022, we will run a large experiment on partner-selection for collaborative tasks, and we aim to initiate at least one more large experiment in another area of interest to the members of the Centre.

We will appoint a centrally funded postdoc dedicated to working on cross-cutting research questions such as criteria for evaluating Hybrid Intelligent AI systems, metrics for measuring the degree to which a system is successfully "hybrid", and a terminology for facilitating the communication in the field of Hybrid Intelligence

Community-wide

In line with our ambition from last year, we have expanded our consortium with one partner (growing from 6 to 7 affiliated universities), and we aim to add at least one more institution to the Hybrid Intelligence Centre in 2022.

We have also achieved our goal of initiating a new international conference, which will take place 13-17 June in Amsterdam. We intend this to be the first event in a continuing series, and we aim to have the 2nd HHAI conference organised outside The Netherlands in 2023.

We have not achieved our goal of acquiring a funded project under the national AINed programme, due to much slower than expected roll-out of AINed. Together with our partners at TNO, we will continue to influence the Dutch AINed programme, aiming for substantial funding for both applied and foundational programmes on Hybrid Intelligent AI Systems.

10

PhD/PD projects

Summaries

Project 1, Aishwarya Suresh

How can AI facilitate Moral Self-Development

Goal: The goal of this project is to make AI systems support (or constraint) more explicit as a step towards empowering better human moral judgments.

Progress: The first year was spent on literature review and refining the research project, resulting in reframing the initial research question into 'How can AI facilitate Moral Education for Moral Progress?', with the following sub-questions: 1) What role can the development of a moral-self play in (any kind of) moral progress? 2) Is AI able to facilitate the development of such a process? 3) How ethical is it to use AI for the development of the moral self?

Project 2, Cor Steging

Aligning Learning and Reasoning for Responsible Hybrid Intelligence

Goal: This project aims at developing mechanisms for the verification and evaluation of hybrid systems that combine manual knowledge-based design and learning from data.

Progress: Based on earlier research, a method is being developed for rationale evaluation that can be used to measure the soundness of the decision-making process of machine learning systems. Additionally, several experiments in the legal domain were performed and it was shown that the insights gained through our rationale evaluation method cannot be achieved through commonly used explainable AI methods.

C. Steging, S. Renooij, and B. Verheij, “Discovering the rationale of decisions: towards a method for aligning learning and reasoning,” in ICAIL ’21: Eighteenth International Conference for Artificial Intelligence and Law, São Paulo Brazil, June 21 – 25, 2021, 2021, p. 235–239. doi:10.1145/3462757.3466059 [Download PDF]

C. Steging, S. Renooij, and B. Verheij, “Discovering the Rationale of Decisions: Experiments on Aligning Learning and Reasoning,” in 4th EXplainable AI in Law Workshop (XAILA 2021), 2021, p. 235–239. [Download PDF]

C. Steging, S. Renooij, and B. Verheij, “Rationale Discovery and Explainable AI,” in Legal Knowledge and Information Systems – JURIX 2021: The Thirty-fourth Annual Conference, Vilnius, Lithuania, 8-10 December 2021, 2021, p. 225–234. doi:10.3233/FAIA210341 [Download PDF]

Project 3, Michiel van der Meer

Mining texts for perspectives in human-machine deliberation

Goal: Perspectives in deliberation are about developing Artificial Intelligence techniques to find the underlying structures of debates and group deliberations to help participants in a debate /deliberation understand why others have a different opinion. Hereto, proposition mining, implication mining and perspective mining will be used.

Progress: Sofar, a method was developed for hybrid (human + AI) argument extraction from contemporary policy-making deliberations, and a review paper was (co-)produced on the role of AI and computer science in deliberation platforms.

E. Liscio, M. van der Meer, L. C. Siebert, C. M. Jonker, N. Mouter, and P. K. Murukannaiah, “Axies: Identifying and Evaluating Context-Specific Values,” in Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, 2021, p. 799–808. [Download PDF]

E. Liscio, M. van der Meer, C. M. Jonker, and P. K. Murukannaiah, “A Collaborative Platform for Identifying Context-Specific Values,” in Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, 2021, p. 1773–1775.

Project 4, Tiffany Matej Hrkalovic

Hybrid intelligence techniques for initiating collaborative behavior

Goal: This project aims to develop hybrid intelligence that can understand, predict, and potentially aid in initiating collaborative behavior.

Progress: The first year a detailed setup for a dataset was developed, enabling us to research dyadic interactions, collaborations, cooperative behavior and partner selection; the collection and the implementation of the dataset was prepared; and an outline of the research and future publications was made. Besides, a contribution was made to research on how nonverbal social cues can be used for automatic classification of subjective perceptions of situation interdependence.

B. Dudzik, S. Columbus, T. M. Hrkalovic, D. Balliet, and H. Hung, “Recognizing Perceived Interdependence in Face-to-Face Negotiations through Multimodal Analysis of Nonverbal Behavior,” in Proceedings of the 2021 International Conference on Multimodal Interaction, New York, NY, USA: Association for Computing Machinery, 2021, p. 121–130.

Project 5, Nicole Orzan

Computational deliberation

Goal: Deliberation is a key ingredient for successful collective decision-making. In this project we will: 1) investigate the mathematical and computational foundations of deliberative processes in groups of autonomous agents; 2) lay the theoretical groundwork for modelling and analyzing how opinions are exchanged and evolve in groups of autonomous decision-makers; 3) develop principled methods for the design of deliberation procedures that can provably improve the quality of group decisions.

Progress: The focus of the first year was on multi-agent systems where agents learn through reinforcement learning and where the emergence of communication is necessary to complete tasks. Currently, we are studying fully collaborative navigation systems, where the goal is the same for all agents, but the observability of the environment is partial for each agent, while exploiting the new technique of Upside-Down Reinforcement Learning.

Project 6, Emre Erdogan

Computational Theory of Mind for Collaborative Privacy

Goal: The long-term goal of this research is to design and develop a generic, feasible, and human-inspired computational Theory of Mind (ToM).

Progress: As an initial step towards this goal, an abstraction framework for ToM was developed over which we constructed an abstraction heuristic. This was studied in our computational affinity model which captures how humans use affinity in their interactions and can be used in tandem with ToM reasoning when designing.

agent models. Our first results show that agents can benefit from modelling affinity computationally and even more when they combine it with basic ToM reasoning, which in turn provides the motivation for us to develop more sophisticated agents and test their capabilities in more complex settings.

Project 7, Karine Miras

Ethical Robot Stewards & Human-bot cooperations & Evolvable Robot

Goal: By bringing roboticists and ethicists together the project addresses two fundamental issues. Firstly, how to translate ethical requirements into design and technical features. Secondly, we explore the trade-off between adaptivity and responsibility attribution trying to find Pareto-optimal solutions.

Progress: The work in the first year was distributed over three subprojects: 1) Ethical Robot Stewards (ERS), demonstrating how to use the Ethicist As Designer methodology to incorporate ethical values into robot design choices. Additionally, the limitations of the method are analyzed and criticized. 2) Hu-bot cooperation (HBC), studying the cooperation between robots and humans aiming at promoting safety through the use of human help to prevent undesired robot behaviour: a human can interfere in the learning loop to teach the robot. 3) Evolvable Robots (ER), where it is investigated from different perspectives how machines can have their bodies and brains adapting to the environment.

K. Miras, J. Cuijpers, B. Gülhan, and A. Eiben, “The Impact of Early-death on Phenotypically Plastic Robots that Evolve in Changing Environments,” in ALIFE 2021: The 2021 Conference on Artificial Life, 2021. [Download PDF]

K. Miras, “Constrained by Design: Influence of Genetic Encodings on Evolved Traits of Robots,” Frontiers Robotics AI, vol. 8, p. 672379, 2021. doi:10.3389/frobt.2021.672379

Project 8, Niklas Höpner

Hybrid approaches to learn explainable sequential behaviours

Goal: This project combines (deep) reinforcement learning research with prior knowledge from knowledge graphs to learn explainable strategies for complex sequential tasks. The combination of methods for classical planning based on symbolic representations and data-driven approaches is expected to advance the state of the art in the areas of knowledge representation and reinforcement learning.

Progress: The first project called ”Leveraging class abstraction for commonsense reinforcement learning via residual policy gradient” was finalized and submitted to IJCAI 2022. The project looked at enabling an RL agent to leverage commonsense knowledge. A novel method was developed to enable an agent to use state abstraction to learn with fewer samples and have the ability to generalise to unseen objects.

Project 9, Bram Renting

AutoAI for dynamic data

Goal: The goal of this project is to develop automated machine learning methods for changing/non iid data, with applications to standard learning scenarios as well as to automated negotiation.

Progress: The first year focussed on agents that can automatically learn to negotiate, an environment that is unrelated to scheduling problems but forms a decent first step towards the goal. The second step was towards more adaptive negotiation agents that learn and adapt their strategy based on the composition of the group of agents and the problem at hand. Experiments on both topics have been successful and publishing is in progress.

B. M. Renting, H. H. Hoos, and C. M. Jonker, Automated Configuration and Usage of Strategy Portfolios for Bargaining, 2021. [Download PDF]

B. M. Renting, H. H. Hoos, and C. M. Jonker, “Automated Configuration of Negotiation Strategies,” in Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, 2020, p. 1116–1124. [Download PDF]

Project 10, Anna Kuzina

Continual Learning and Deep Generative Modelling for Adaptive Systems

Goal: This project focuses on fundamental research into combining various learning paradigms for building intelligent systems capable of learning in a continuous manner and evaluating uncertainty of the surrounding environment. We aim at designing new models and learning algorithms by combining multiple machine learning methods and developing new ones.

Progress: During this year, we have been working on increasing the robustness of Variational Autoencoders (VAE) to minor shifts in data distribution, which has been published. We are currently working on a method that can be used to reduce the damage of the attacks. Another paper is in preparation. Furthermore, a pre-print on the efficient neural networks parametrization (Romero et al., 2021), which is currently under review for ICLR 2022, was published.

A. Kuzina, M. Welling, and J. M. Tomczak, “Diagnosing Vulnerability of Variational Auto-Encoders to Adversarial Attacks,” in ICLR 2021 Workshop on Robust and Reliable Machine Learning in the Real World, 2021. [Download PDF]

D. W. Romero, A. Kuzina, E. J. Bekkers, J. M. Tomczak, and M. Hoogendoorn, “CKConv: Continuous Kernel Convolution For Sequential Data,” CoRR, vol. abs/2102.02611, 2021. [Download PDF]

Project 11, Mani Tajaddini

A Design Pattern Language for HI teams

Goal: The ambition for this position is to define design patterns for successful configurations of Hybrid Intelligent systems and we aim to develop a corresponding pattern language to express such design patterns in conceptual, (semi-)formal or computational form, and an empirical method to validate the patterns.

Progress: The first year was spent on generating a systematic review paper on design patterns related to human-AI interaction, to map what design patterns are already available and to find areas that need design patterns to be developed, including representation and evaluation methods employed with design patterns so far. In addition, work was done on creating design patterns in needed areas in collaboration with other HI and non-HI projects and researchers primarily in fields like healthcare and adaptive explainable AI.

Project 12, Taewoon Kim

Identity, reference and perspective in personal communication

Goal: This project addresses the phenomena of identity, reference and perspective within communicative scenarios between AI and people in real-world situations. It aims to build personal relationships between people and AI in an adaptive environment to learn how to communicate about shared situations or contexts.

Progress: The first year was started with literature review, and with the set-up of two novel memory systems. Furthermore, the memory systems are being modelled with symbolic knowledge graphs to demonstrate that, combined with machine learning, this hybrid of symbolic and sub-symbolic artificial intelligence is better than only having one. The first experiment is underway.

S. Baez Santamaria, T. Baier, T. Kim, L. Krause, J. Kruijt, and P. Vossen, “EMISSOR: A platform for capturing multimodal interactions as Episodic Memories and Interpretations with Situated Scenario-based Ontological References,” in Proceedings of the 1st Workshop on Multimodal Semantic Representations (MMSR), Groningen, Netherlands (Online), 2021, p. 56–77. [Download PDF]

Project 13, Annet Onnesi

Monitoring and constraining adaptive systems

Goal: This project aims at raising and educating a new generation of artificially intelligent agents: adaptive agents who need to learn to abide by our rules in hybrid intelligent teams. It focuses on fundamental research into integrating interpretable knowledge representation and reasoning with learning in the context of adaptive systems.

Progress: The project started with unravelling all the relevant concepts and the relations between them, leading to brief conceptual analyses of concepts such as adaptive systems and monitoring. Subsequently, more concrete research on models and distributions was done, with a literature review of probability distributions and different measures, specifically distance measures. Finally, work was started on finding a way of expressing formally how a simple model (having less detail) can represent a multitude of more complex models and how to define this set of more complex models.

Project 14, Pei-Yu Cheni

Interactive Machine Reasoning for Responsible HI

Goal: The goal of this project is to realize a support agent that is able to provide support that is in alignment with what users need or want. As opposed to most data-driven approaches, we propose the concept of Alignment Dialogue to achieve it.

Progress: The project started with a literature review on the relevant topics such as user modelling, personalization, and AI alignment, as the basis for drawing up a detailed research plan. In addition, the first study was conducted: an exploratory user study in the form of a focus group where we collected and analyzed end-users’ opinions in an alignment dialogue (Paper in preparation). We were able to establish some foundations centered around alignment dialogues. In addition to the findings, the user study shed light on what is still missing in the picture, which serves as the basis for the research questions in the following years.

Project 15, Urja Khuranai

Predicting generalizability of NLP experiments

Goal: This project aims to address two important questions regarding real-life applicability of NLP methods,

by combining advanced systematic error analyses and formal comparison of textual data and language models: First, what kind of errors does the method make and how problematic are they for the application? Second, how predictive are results obtained on the benchmark sets for the data that will be used in the real-life application?

Progress: We conducted a first experiment to look at robustness from the perspective of stability, which indicated that prediction of generalization should be done on a checkpoint-level. Furthermore, we are working on improving the definition of hate speech in NLP literature by discussing our own criteria.

U. Khurana, E. Nalisnick, and A. Fokkens, "How Emotionally Stable is ALBERT? Testing Robustness with Stochastic Weight Averaging on a Sentiment Analysis Task," in Proceedings of the 2nd Workshop on Evaluation and Comparison of NLP Systems, Punta Cana, Dominican Republic, 2021, p. 16–31. [Download PDF]

Project 16, Kata Naszadii

Context modelling for dialogue systems

Goal: In this project, we will investigate the role of context (agent or environment) on natural dialogue generation. We will explore several research directions such as the detection and representation of knowledge gaps, the generation of contextually appropriate responses, and novel ways to represent and access large contextual information.

Progress: In the first year we have been looking at a fixed-semantics setting, and into properties of learned communication protocols in multi-agent games, in order to spot interesting pragmatic phenomena. Investigations are ongoing in goal-oriented, situated natural language dialog, to analyse where we benefit from modelling agency and being goal oriented, as opposed to just supervised learning of dialogs. Together with other HI students, I took part in the IGLU challenge (https://www.iglu-contest.net/)..

Project 17, Robert Loftin

(Meta-) learning to give feedback in interactive learning

Goal: This project aims to use deep reinforcement learning (RL) to train agents that can reliably collaborate with and adapt to arbitrary human partners.

Progress: So far, our main theoretical contribution has been a result proving that it is impossible to guarantee reliable cooperation with a human partner whose behavior is allowed to adapt arbitrarily to our agent's own behavior. A paper on this impossibility result is currently under review for ICML 2022. We have also done empirical and engineering work related to questions on how we can learn control policies that can reliably adapt to any human partner whose behavior satisfies our predefined set of assumptions. .

Project 18, Bernd Dudzik

Automated social perception for aiding human-human cooperation in dynamic settings

Goal: In this project, we will investigate the role of context (agent or environment) on natural dialogue generation. We will explore several research directions such as the detection and representation of knowledge gaps, the generation of contextually appropriate responses, and novel ways to represent and access large contextual information.

Progress: In the first year, we developed a detailed research plan for the remainder of the project duration. We identified relevant challenges and research questions (conceptual and technical), e.g. with regards to temporal dynamics of cognitive-affective processes, demonstrated that important types of subjective impressions can potentially be estimated by automatic analysis of individuals' nonverbal behavior in face-to-face negotiation settings, and developed a setup and protocol for a large-scale online experiment/data collection about conversational interactions, collaborations, and partner selection.

B. Dudzik and J. Broekens, "A Valid Self-Report is Never Late, Nor is it Early: On Considering the Right Temporal Distance for Assessing Emotional Experience," , 1, 2021.

B. Dudzik, S. Columbus, T. M. Hrkalovic, D. Balliet, and H. Hung, "Recognizing Perceived Interdependence in Face-to-Face Negotiations through Multimodal Analysis of Nonverbal Behavior," in Proceedings of the 2021 International Conference on Multimodal Interaction, New York, NY, USA: Association for Computing Machinery, 2021, p. 121–130. doi:10.1145/3462244.3479935 [Download PDF]

Project 19, Wijnand van Woerkom

Explaining data-driven decisions with legal, ethical or social impact to end-users

Goal: This project aims to explain the outcomes of data-driven machine-learning applications that support decision-making procedures to end-users of such applications, such as lawyers, business people, or ordinary citizens.

Progress: At the start of the project, a literature survey was conducted on the main forms of XAI that we are seeing nowadays such as surrogate model explanations, counterfactual explanations, and case-based reasoning. Subsequently, we have been working on an experiment on the topic of argumentation-based XAI, aiming to critically assess the explanations it generates, particularly with respect to detecting biased decision making, by testing it on a recidivism dataset.

Project 20, Merle Reimann

Enabling co-regulation for long-term engaging semi-structured conversations

Goal: The aim of this project is to develop a conversational agent that is able to establish and maintain a common understanding in co-regulation with its user, by offering the flexibility to deviate from the conventional sequences in structured question-and-answer or instructional dialogues. A second aim is to address the level of user engagement in such highly structured conversations that tend to be repetitive, by enabling the agent to memorize and refer back to relevant pieces of information in the conversation history or earlier conversations with the same user, and by incorporating variation in the agent's utterances.

Progress: After reviewing the literature on Dialogue Management (DM) in Human-Robot Interaction (HRI), we found that the two fields are rather disconnected. Therefore, we decided to write an extensive systematic review about DM in HRI, as a step towards bridging the gap between the two fields. We review the papers in regard to the used DM framework, shortcomings, evaluation methods, and common capabilities. We are currently in the process of writing down our results. Based on the insights gained from the review process and a demo implementation of an interview robot, we were able to make some adjustments to the future focus of the project and refine the research questions.

Project 22, Íñigo Martínez de Rituerto de Troya

Including ethical and legal considerations in the HI design process

Goal: The project focuses on understanding how human factors are crucial for informing the design of AI systems in making sure these are safe in their behavior, in their treatment of people, and allow affected stakeholders to be empowered to hold the system and its managers/operators/developers accountable.

Progress: In the first year, I focused on scoping the research objectives, including defining the knowledge gap and research questions above. The work has resulted in a detailed research proposal and timeline for the next 3 years. Additionally, we have started conversations with two potential project partners and are continuing to explore further collaboration..

Project 23, Andreas Sauter

Causal Reinforcement Learning and Reasoning

Goal: In this project, we will design an adaptive agent which will enable causal inference through learning and reasoning in a reinforcement learning environment. In particular, the agent is expected to combine ontological background knowledge and causal discovery and inference effectively in achieving a given task in a partially observable environment. We will try to obtain theoretical guarantees, and outline its performance empirically.

Progress: The project was initiated recently, and the first period was spent on setting up a training and supervision framework for the PhD, and literature research on related work. Based on earlier work, a journal paper was submitted, describing a reinforcement learning setup that allowed our agent to build a causal model of the (very simple, synthetic) environment. We also identify problems with our model regarding environments that are too simple and the mandatory apriori knowledge of a vast training set of training causal models, which will be the starting point for future research.

Project 24, Maria Heuss

Explainable De-biasing in Learning from Interactions

Goal: Learning to rank is at the heart of modern search and recommender systems. Increasingly, such systems are optimized using historical, logged interaction data, which is known to be biased in a myriad of ways. This project aims to design debiasing methods that are explainable and that can be applied selectively, depending on context, user (or user model), and additional constraints such as fairness.

Progress: In the first year, work was done on understanding and mitigating the effect of outliers in fair ranking, which will be published at WSDM22. In addition, a paper is in preparation on fairness of exposure in light of incomplete exposure estimation. Currently, I am dissecting existing fairness evaluation metrics for ranking systems, indicating that none of these metrics are fit to properly evaluate the fairness of exposure with respect to individual fairness. I am also working on developing an evaluation method that avoids these shortcomings and can be used in a broader context than previous methods.

F. Sarvi, M. Heuss, M. Aliannejadi, S. Schelter, and M. de Rijke, "Understanding and Mitigating the Effect of Outliers in Fair Ranking," in WSDM 2022: The Fifteenth International Conference on Web Search and Data Mining, 2022. Download PDF]

Project 25, Sharvaree Vadgama

Teaching artificial neural networks how to reason

Goal: The aim of this project is to model, study, and visualize decision-making processes in (convolutional) neural networks. In particular, emphasis is put on explainability in critical applications where reliability and trustworthiness of AI systems are crucial, such as in the medical domain. The main theme will be the design of AI systems through recurrent NNs where focus is put on modeling (and visualization of) the dynamics of decision making rather than relying on the static black box mechanics of feed-forward networks.

Progress: To develop methods and tools to understand and interpret model outcomes, we proposed a generative explanation technique using VAEs wherein the output of the model can be explained locally using latent space traversals. This work was done in collaboration with University Medical Center Utrecht. For building interpretable architectures, I am currently focusing on Capsule Networks. We aim to design flexible and efficient capsule networks which are fully equivariant. At the moment we have a full framework and are running experiments on point cloud data and molecule data and developing the code for our method.

M. B. Vessies, S. P. Vadgama, R. R. van de Leur, P. A. F. M. Doevendans, R. J. Hassink, E. Bekkers, and R. van Es, "Interpretable ECG classification via a query-based latent space traversal (qLST)," CoRR, vol. abs/2111.07386, 2021. [Download PDF]

Project 26, Loan Ho

Knowledge Representation Formalisms for Hybrid Intelligence

Goal: Knowledge graphs (KGs) have a significant role in storing and providing access to global knowledge, common and accessible to both human and artificial agents, representing knowledge for different agents in Hybrid Intelligence (HI) settings. Unfortunately, providing suitable formalisms to work with complex, conflicting, dynamic and contextualised knowledge is still a big challenge. This project aims to propose an approach to represent and adapt knowledge dynamically in different scenarios (user preference, available sources, dynamic or contextualized knowledge).

Progress: Until now, we have provided an argumentation framework (AF) to address representation and reasoning in inconsistent KBs. Besides, we have studied a dialogical explanation technique based on argumentation trees to allow interaction between humans and agents. Currently, work on the development of a dialogical explanation technique is in progress. In addition, we are performing an experiment for the AF framework in the Biographical Dictionary domain.

Project 27, Putra Manggala

Functional Priors for Fast Bayesian Adaptation under Human Guidance

Goal: This project aims to use Bayesian priors defined in function space to enable data-efficient and human-guided model adaptation.

Progress: In the first year, we have created a Bayesian framework that models the uncertainties in weak supervision and manages the misspecifications in the labelling processes. It handles multiple sources of weak supervision and uses a set of unsupervised objectives to tune the final predictive distribution.

P. Manggala, H. H. Hoos, and E. Nalisnick, "Bayesian Regression from Multiple Sources of Weak Supervision," in ICML 2021 Machine Learning for Data: Automated Creation, Privacy, Bias, 2021. [Download PDF]

Project 28, Erman Acar

Game-Theoretic Development and Analysis for Hybrid Intelligence

Goal: In this project, we will develop a game-theoretic framework that can capture the cooperation scenarios in heterogeneous groups (a mix of human and artificial agents). Such a study will require the development of necessary game-theoretic concepts; among them, refinement of existing equilibrium concepts, bounded rationality models, and provably good guarantees.

Progress: So far, we have created an environment in which we showed that RL agents can learn to cooperate. Besides, we observed that human-signaling can help agents' coordination. Currently, we are working on testing its transferability. Work on the question of human demonstration is in progress, and the investigation was just started on the benefits of high-level symbolic information for helping tasks that need coordination.

Project 29, Mike Ligthart

Personalised long-term interaction with an application to educational robots

Goal: A key challenge in human-robot interaction is to ensure engagement over time. This is also particularly true for child-robot interaction. In this project we will research how personalisation can be used to address this challenge and motivate children. We will test and evaluate our ideas in the education domain with the

aim to engage children to keep practicing maths problems with a robot. We will create a conversational design for long-term engaging child-robot interaction. We will design a memory of the interaction history and strategies to use that memory to personalise the interaction. The final challenge will be to design a personalised user model to further tailor the interaction to the child's needs and preferences.

Progress: So far we developed four robot behavior components. The first is a narrative dialog concept that offers maths sums in the form of a story-rich conversation with the robot. The second is a memory-based personalization strategy that allows the robot to tailor the story to the child's interests and preferences. The third component is a supportive guidance strategy for when the child does not know the answer. The fourth component is a dynamic maths level adaptation strategy allowing the robot to adjust the complexity of the sums based on the child's performance during the interaction. We are currently preparing a large in-the-wild user study (targeted N=200, 8-10 y.o. school children) where we will evaluate the personalization strategy and the guidance strategy in a 2x2 between-subjects design.

Project 30, Enrico Liscio

Game-Theoretic Development and Analysis for Hybrid Intelligence

Goal: Values are abstract motivations that justify opinions and actions. Understanding values is an essential milestone in achieving beneficial AI, with applications in fields such as autonomous driving and healthcare. The goal of this project is to use natural language processing (NLP) to estimate values directly from discourse.

Progress: First, investigations were done to understand how to identify the values that are relevant to a context, and the results were published in two papers at AAMAS 2021. Next, we focussed on whether transfer learned value rhetoric can be transferred across contexts, and the resulting paper has been submitted to ACL. Work on the learning capacity of value classifiers for context-specific connotations of values in progress, and studies on the best way to annotate values just initiated.

E. Liscio, M. van der Meer, L. C. Siebert, C. M. Jonker, N. Mouter, and P. K. Murukannaiah, "Axies: Identifying and Evaluating Context-Specific Values," in Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, 2021, p. 799–808. [Download PDF]

E. Liscio, M. van der Meer, C. M. Jonker, and P. K. Murukannaiah, "A Collaborative Platform for Identifying Context-Specific Values," in Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, 2021, p. 1773–1775. [Download PDF]

Project 31, Mark Adamik

Common-sense knowledge representation & reasoning for embodied agents

Goal: This PhD project will tackle the problem of enabling intelligent systems to reason over common sense knowledge to achieve their tasks. We will focus on the specific case of embodied agents performing everyday activities such as cooking or cleaning, i.e. tasks that require reasoning over different types of knowledge (we call these knowledge-intensive tasks), and using for this purpose the available common sense knowledge graphs, i.e. structured data sources that contain knowledge about the everyday world. We will propose new methods to allow robots to achieve knowledge-intensive tasks through combining existing structured data (knowledge graphs) with situational personalised knowledge captured through their sensors. We will study the hybrid application of advanced symbolic and subsymbolic AI methods to robotic systems operating in complex real-world scenarios.

Progress: As of now, a simulation environment was created that allows the simplification of some of the problems (such as perception) and a real-world robot was acquired. Furthermore, existing cognitive architectures are examined that allow the efficient interpretation of the multi-modal data acquired by the robot and their translation into actions. Taking these existing works as an inspiration, we are currently working on creating our own architecture that allows the robot to integrate knowledge graphs and reasoning into the decision-making process.

Project 32, Selene Baez

Common-sense knowledge representation & reasoning for embodied agents

Goal: In this Ph.D. we study trust as a relationship between social agents in a multimodal world that involves multifaceted skills and complex contexts. This work aims to create and evaluate a computational model of trust, from an artificial agent's perspective towards trusting humans in collaborative tasks. We propose that an artificial agent should be able to establish a trust relationship with humans, seeing them as sources of information..

Progress: We have decided to focus on whether a collaborating agent is skilled (has domain knowledge) and reliable (communicates this knowledge consistently). We have developed a platform to capture multimodal interactions that yield episodic Knowledge Graphs to represent an agent's accumulated experiences. From these graphs, we can identify which agents bring in new valuable information for a task, and which may have provided inconsistent knowledge over time. We are using this finding to create a computational model that can identify agents that bring in consistent and relevant domain knowledge.