HumanE AI Net:

The HumanE AI Network

Grant Agreement Number: 952026
Project Acronym: HumanE AI Net

Project Dates: 2020-09-01 to 2023-08-31
Project Duration: 36 months

D1.1: First Year Microproject Results from Workpackage 1, 2 and 3

Author(s): Sahan Bulathwela, John Shawe-Taylor, George Kampis
Contributing partners: UCL, DFKI
Date: September 13, 2021
Approved by: Paul Lukowicz
Type: Report (R)
Status: Final
Contact: m.bulathwela@ucl.ac.uk

Dissemination Level		
PU	Public	x

DISCLAIMER

DOCUMENT INFO

0.1 Authors

Authors	Institution	e-mail
Sahan Bulathwela (SB)	UCL	m.bulathwela@ucl.ac.uk
John Shawe-Taylor (JST)	UCL	j.shawe-taylor@ucl.ac.uk
George Kampis (GK)	DFKI	George.Kampis@dfki.de

0.2 Document History

Revision		
Date	**Lead Author(s)**	**Comments**
30.09.2020	GK	Empty template
10.09.2021	SB and JST	Initial and Second draft
13.09.2021	GK	Prettifying
22.09.2021	SB	Added Minor Modifications
14.10.2021	JST	Removed sections for D4.1

Table of Contents

0. Introduction

This section introduces the concept of a microproject (MP) and how it has evolved since the original work programme. We will also give some statistics on the development of MPs within the project, concluding with an outline of how we anticipate that the MP activity may develop during the remainder of the network.
Following this introduction, there are separate sections for work packages (WPs) 1, 2 and 3 that cover the MPs that are mainly focused on that WP. Following an overview of the MP activity in that WP, descriptions of the individual MPs are included together with any deliverables that have been completed for those that are either finished or have been running more than 50% of their anticipated time.

0.1 What is a micro-project?[1]

A micro project is a cooperation of two or more partners over a period of typically 1-6 months aimed at producing a tangible outcome (paper, data set, demo, tutorial etc.) to be made available to the community through the AI4EU platform and appropriately promoted in the community. Key hard requirements are:

1. two or more HumaneAI Net partners working together over a period of 1-6 months
2. cooperation to be documented eg. through joint authorship of the paper.
3. topic clearly tied to one or more tasks as described in the proposal (can be tasks from different WPs)
4. tangible outcome (paper, data set, toolset, demo, etc.)
5. outcome made available through the AI4EU platform or appropriate dissemination
6. a short presentation (5-15 mins) recorded at the end of the project to be made available through appropriate channels, including the project YouTube channel)

Originally there was also a very hard travel requirement so that, for the duration of the micro project, all researchers working on the micro-project are to work together at the same site (thus one organisation is the host, while the other participating researchers travel to the host site for the duration of the micro-project). However, due to the covid situation the travel requirement has been temporarily suspended.

0.2 Microprojects and financing

MPs were the vehicle to promote intensive collaboration between partners, leverage synergies between their groups and "spread the resulting knowledge" to the broader European AI community. Resources were allocated to fund this activity in the sense that most PMs in WPs 1-6 must be spent through micro-projects. We also have additional funds managed by the coordinator for which partners with good micro-project ideas can apply. This includes funds for supporting external (non-HumanE AI Net) as participants in micro-projects.
However, during the first year we have concentrated exclusively on internal MPs that make use of 'pre-assigned funds', ensuring that the approval method can be light weight and that activity can begin as quickly as possible. MP submissions were reviewed and approved by

[1] The following two sections are shared with Deliverable 4.1.

WP leaders of the WPs that the MP fell under (this could be typically a principal WP plus one or more additional secondary WPs). Approvals were expected and, in most cases, given within a week, to ensure that delays in initiating the work were minimized.
This deliverable contains reports on the MPs that were initiated and, in many cases, completed during the first year under his procedure.
Our assessment is that the approach has been very successful in initiating the collaborative activity described in the work programme with many impressive outputs delivered. In the context of the COVID disruption, we believe that this has been a very positive outcome.
There are two outstanding issues that must be addressed during the remainder of the network:

- the need to ensure that the workprogramme is fully addressed in the sense that the checks of the WP leaders ensured MPs were addressing the work programme but given the bottom-up procedure they were not able to guarantee that all aspects would be covered by the set of proposed MPs.
- The allocation of additional resources beyond the 'pre-assigned' budget including resources to fund the involvement of additional partners needs to be made.

We plan to address these two issues by reviewing the progress of the overall research against the work programme and identify missing elements as well as particularly promising emerging directions. These will then be used to define calls for new MP proposals addressing these topics, but at the same time leaving open the option for proposals addressing different topics. The proposers will also be able to request additional funding either for internal or new external partners. An appropriate reviewing procedure is under development to ensure that the proposals are objectively and adequately reviewed and decisions about funding can be reached that will ensure gaps in our coverage of the work programme are filled and opportunities for particularly promising work can be supported. The timeline for completing this process is in time for new MPs to initiate early in the new year.

1. Work Package 1: Human-in-the-Loop Machine Learning, Reasoning and Planning

1.1 Overall Summary for Work Package 1

The work package 'Learning, Reasoning and Planning with Human in the Loop' is concerned with the development of core technologies enabling more inclusive AI solutions that allow both understanding and interaction with humans. Such developments will underpin much of the work in other work packages that implement and demonstrate how humans can be empowered by AI in ways that enrich rather than sideline. The set of microprojects (MPs) that have been proposed cover the wide spectrum of tasks envisaged in the work package. The first of these was the linking of symbolic and sub-symbolic learning evident both explicitly and implicitly in several MPs, hence providing evidence for the value of this integration at a variety of levels. The second task of learning with and about narratives is also an explicit theme in more than one MP, as well as appearing as a thread in other MPs. Again, there is evidence that the 'narratives' perspective can motivate the development of new representations as well as guide the design of systems that are potentially more understandable to humans. The third task was the development of continuous and incremental learning in joint human/AI systems. This is a thread in some MPs such as the building consensus and educational recommenders. It is, however, desirable that this theme be further developed in the next stage of the project. Auto ML, the fourth task, has been given explicit attention in MPs and will be developed further in the integration phase of the network's activities. The final technical task of 'quantifying model uncertainty has received attention in the context of automated driving, but it would be desirable to expand the work of this task in subsequent MPs. The final task is the consolidation and coordination of the research agenda which has been initiated following the results of the initial batch of MPs. This will produce a coherent program of research for the second phase of the project with calls for MPs in areas where a need for expanded or greater attention has been identified.

1.2 Completed Micro Projects

The following microprojects address the challenges raised in WP1.

1. A tale of two consensus. Building consensus in collaborative and self-interested scenarios.

Proposal Submission Date: January 11, 2020
Actual Start Date: January 11, 2020
Expected Duration: 4 Months
Actual Duration: 7 Months

Many citizen science projects have a crowdsourcing component where several different citizen scientists are requested to fulfill a micro task (such as tagging an image as either relevant or irrelevant for the evaluation of damage in a natural disaster or identifying a specimen into its taxonomy). How do we create a consensus between the different opinions/votes? Currently, most of the time simple majority voting is used. We argue that alternative voting schemas (taking into account the errors performed by each annotator) could severely reduce the number of citizen scientists required. This is a clear example of continuous human-in-the-loop machine learning with the machine creating a model of the humans that it has to interact with.

We propose to study consensus building under two different hypotheses: truthful annotators (as a model for most voluntary citizen science projects) and self-interested annotators (as a model for paid crowdsourcing projects).

Expected Outputs:

- Software and documentation for the two new consensus models into the crowd analysis framework.
- New consensus models case study in a citizen science project.
- Algorithm for numerical simulations useful to evaluate the efficacy of the consensus models considered in crowd analysis.
- Report of the results of simulations, with suggestions to improve the consensus models.

Actual Outputs:

- Crowdanalysis, (program/code),
  URL: https://pypi.org/project/crowdnalysis/
- Cerquides, J.; Mülâyim, M.O.; Hernández-González, J.; Ravi Shankar, A.; Fernandez-Marquez, J.L. A Conceptual Probabilistic Framework for Annotation Aggregation of Citizen Science Data. Mathematics 2021, 9, 875., (publication),
  URL: https://doi.org/10.3390/math9080875
- Sánchez-López, B.; Cerquides, J. On the Convergence of Stochastic Process Convergence Proofs. Mathematics 2021, 9, 1470., (publication),
  URL: https://doi.org/10.3390/math9131470
- Hernández-González, J.; Cerquides, J. A Robust Solution to Variational Importance Sampling of Minimum Variance. Entropy 2020, 22, 1405., (publication),
  URL: https://doi.org/10.3390/e22121405
- Jesus Cerquides (2021). Parametrization invariant interpretation of priors and posteriors. arXiv:2105.08304 [cs, math, stat]., (publication),
  URL: https://doi.org/http://arxiv.org/abs/2105.08304

Connection of Results to Work Package Objectives:
Crowdsourcing can be applied to quickly obtain accurate information in different domains, including disaster management scenarios. This requires the computation of the consensus among the different annotators. Probabilistic graphical models can be used to build interpretable consensus models. These models can answer questions such as “Who is the more competent annotator for this task?”, or “How many annotators do I need for this task?” which provide a clear example of machine learning with human-in-the-loop, and fully related to T1.3 Continuous & incremental learning in joint human/AI systems in WP2.
The evolutionary results will help to understand the best way to proceed in the research, suggesting new theoretical and experimental studies to address the topic. Therefore, they make it possible to evaluate the interplay between human acting and AI learning in crowdsourcing tasks, connected with T3.2 Human-AI Interaction/collaboration paradigms

2. Educational Recommenders with Narratives

Proposal Submission Date: November 01, 2020
Actual Start Date: November 02, 2020
Expected Duration: 6 Months

Actual Duration: 9 Months

Through this work, we explore novel and advanced learner representation models aimed at exploiting learning trajectories to build a transparent, personalised and efficient automatic learning tutor through resource recommendations. We elaborate on the different types of publicly available data sources that can be used to build an accurate trajectory graph of how knowledge should be taught to learners to fulfil their learning goals effectively. Our aim is to capture and utilise the inferred learner state and the understanding the model has about sensible learning trajectories to generate personalised narratives that will allow the system to rationalise the educational recommendations provided to individual learners. Since an educational path consists heavily of building/following a narrative, a properly constructed narrative structure and representation is paramount to the problem of building successful and transparent educational recommenders.

Expected Outputs:

- Paper on development of narrative representations for learning
- Enhancements of the X5Learn portal for accessing Open Educational Resources (OER)
- Visualisation software for landscapes of learning
- User evaluations of software

Actual Outputs:

- Maria Perez-Ortiz, Claire Dormann, Yvonne Rogers, Sahan Bulathwela, Stefan Kreitmayer, Emine Yilmaz, Richard Noss, and John Shawe-Taylor. 2021. X5Learn: A Personalised Learning Companion at the Intersection of AI and HCI. In 26th International Conference on Intelligent User Interfaces (IUI '21). Association for Computing Machinery, New York, NY, USA, 70–74., (publication), URL: https://dl.acm.org/doi/10.1145/3397482.3450721
- Understanding the Relevance Scores in Cross-Lingual Document Retrieval, (publication), URL: Submitted to International Journal for Information Processing & Management
- Towards Semantically Aware Educational Recommenders, (publication), URL: To be Submitted.
- TrueLearn INK: Accounting for Interests, Novelty and Knowledge of Online Lifelong Learners, (publication), URL: Submitted to EAAI'22
- PEEK: A Large Dataset of Learner Engagement with Educational Videos, (publication), URL: https://arxiv.org/abs/2109.03154
- PEEK Dataset, (dataset), URL: https://github.com/sahanbull/PEEK-Dataset
- TrueLearn Model, (program/code), URL: https://github.com/sahanbull/TrueLearn
- Semantic Networks for Narratives, (program/code), URL: https://github.com/danlou/mp_narrative

Connection of Results to Work Package Objectives:
The main contributions are towards WP1 and WP3. In terms of WP1, improvements made to TrueLearn contribute towards building a much richer representation of the learner in the educational recommender. This builds towards linking symbolic and sub-symbolic AI where the representation is understood by AI while it can be easily translated to a humanly intuitive narrative. The proposed online learning scheme connects with continual learning where the model constantly updates itself on a lifelong basis. EMD and semantic network enrichment also further contribute to human-in-the-loop machine learning by empowering the human user in the recommendation process with potential narratives of learning trajectories. The results also connect to ideas in WP3 as the recommender's foundations are built on user modelling and exploiting user interaction history. Having already built a rich representation, the results present opportunities to further contribute to identifying useful visualisations and human-AI collaboration frameworks to utilising the learner model in-the-wild.

Deviations from the Initial Plan:
Initial findings using Poincaré / hyperbolic embeddings to derive a concept hierarchy using learner interaction data were not producing interesting results leading to a decision to abandon pursuing this idea further.

3. Collection of datasets tailored for HumanE-AI multimodal perception and modelling

Proposal Submission Date: January 01, 2021
Actual Start Date: January 11, 2021
Expected Duration: 4 Months
Actual Duration: 5 Months

HumanE-AI research needs data to advance. Often, researcher struggle to progress for the lack of data. At the same time, collecting a rich and accurate dataset is no easy task. Therefore, we propose to share through the AI4EU platform the datasets already collected so far by different research groups. The datasets will be curated to be ready-to-use for researchers.
Possible extension and variation of such datasets will also be generated using artificial techniques and published on the platform.
A performance baseline will be provided for each dataset, in form of publication reference, developed model or written documentation.
The relevant legal framework will be investigated with specific attention to privacy and data protection, as to highlight limitations and challenges for the use and extension of existing datasets as well as future data collection on the subject of multimodal data collection for perception modelling.

Expected Outputs:
- Publication of OPPORTUNITY dataset (and other datasets if time available) on the AI4EU platform. [lead: UoS, contributor: DFKI]
- Publication of baseline performance pipeline for OPPORTUNITY dataset (and other datasets if time available) on AI4EU platform. [lead: UoS, contributor: DFKI]
- Investigation of data loader and pipeline integration on AI4EU experiment to load HAR dataset and pre-existent pipelines, with a focus on the opportunity dataset (and other datasets if time available) [lead: UoS, contributor: DFKI]

- Generation of variation [lead: DFKI]
- Survey publications describing datasets and performance baseline [lead: DFKI, contributor: UoS]

Actual Outputs:

- Opportunity++, (dataset)
- CapacitiveGym, (dataset)
- HCI FreeHand, (dataset)
- SkodaMini, (dataset)
- Wearlab BeachVolleyball, (dataset),
  URL: https://ieee-dataport.org/open-access/wearlab-beach-volleyball-serves-and-games

Connection of Results to Work Package Objectives:
Multi-modal perception and modeling need data to progress but recording a new rich and accurate dataset allowing for comparative evaluations by the scientific community is no easy task. Therefore, we gathered rich datasets for multimodal perception and modelling of human activities and gestures. We curated the dataset in order to make them easy to use for research thanks to clear documentation and file formats.
The highlight of this microproject is the OPPORTUNITY++ dataset of activities of daily living, a multi-modal extension of the well-established OPPORTUNITY dataset. We enhanced this dataset which contains wearable sensor data, with previously unreleased data, including video and motion tracking data, which make OPPORTUNITY++ a truly multi-modal dataset with wider appeal, such as to the computer vision community.
In addition, we released other well established activity datasets (HCI FreeHand and SkodaMini dataset) as well as datasets involving novel sensor modalities (CapacitiveGym) and skill-assessment dataset (Wearlab BeachVolleyball)

4. Reasoning on Contextual Hierarchies via Answer Set Programming with Algebraic Measures

Proposal Submission Date: January 01, 2021
Actual Start Date: January 11, 2021
Expected Duration: 3 Months
Actual Duration: 4 Months

This project continues the collaboration between FBK and TUW about defeasible knowledge in description logics in the Contextualized Knowledge Repository (CKR) framework.

In applications, knowledge can hold by default and be overridden in more specific contexts. For example, in a tourism event recommendation system, events can appear as suggested to a class of tourists in a general context: in the more specific context of a particular tourist, preferences can be refined to more precise interests, which may override those at higher contexts.

Goal of this project is to enhance the answer set programming (ASP) based realization of CKR to deal with complex context hierarchies: we use an ASP extension recently proposed by TUW, ASP with algebraic measures, which allows for reasoning on orderings induced by

the organization of defeasible knowledge. This collaboration will provide a prototype for reasoning over CKR hierarchies, but also an application for ASP with algebraic measures.

Expected Outputs:

- Prototype implementation: realization of reasoning service for query answering over CKR with contextual hierarchies. The prototype will be made available in AI4EU platform.
- Report on formalization: technical report and paper submission containing the defining the formal aspects of model selection for contextual hierarchies via ASP with Algebraic Measures and some initial evaluations in the prototype.

Actual Outputs:

- Technical report, (publication),
  URL: https://tinyurl.com/FBK-TUW-MP-Report
- Prototype, (program/code),
  URL: https://github.com/dkmfbk/ckrew/releases/tag/v.1.6

Connection of Results to Work Package Objectives:
The results of the MP are relevant for AI as they show a combination of non-monotonic contextualized DLs in the CKR framework and Logic Programming with numerical measures in weighted LARS.

With respect to the HumaneAI vision, the resulting framework provides a tool for representing, e.g., complex social structures and the contextualization of information relative to such social organizations. With respect to the WP1 objectives, the work combines different AI areas, and follows the direction of joining symbolic and numeric knowledge representation and reasoning methods with notions of uncertainty.

5. Causality and Explainability in Temporal Data

Proposal Submission Date: January 01, 2021
Actual Start Date: February 01, 2021
Expected Duration: 6 Months
Actual Duration: 6 Months

Nowadays ML models are used in decision-making processes in real-world problems, by learning a function that maps the observed features with the decision outcomes. However, these models usually do not convey causal information about the association in observational data, thus not being easily understandable for the average user, therefore not being possible to retrace the models' steps, nor rely on its reasoning. Hence, it is natural to investigate more explainable methodologies, such as causal discovery approaches, since they apply processes that mimic human reasoning. For this reason, we propose the usage of such methodologies to create more explicable models that replicate human thinking, and that are easier for the average user to understand. More specifically, we suggest its application in methods such as decision trees and random forest, since by themselves are highly explainable correlation-based methods.

Expected Outputs:
1. 1 Conference Paper
2. 1 Prototype
3. Dataset Repository

Actual Outputs:
- Methods and Tools for Causal Discovery and Causal Inference, (publication)
- repository of datasets, software, and papers related to causal discovery and causal inference research, (dataset),
  URL: https://github.com/AnaRitaNogueira/Methods-and-Tools-for-Causal-Discovery-and-Causal-Inference

Connection of Results to Work Package Objectives:
The HumanE-AI project thinks a society of increasing interactions between humans and artificial agents. All around the project, causal models are relevant for plausible models of human behavior, man-machine explanations, and upgrading machine-learning algorithms with causal-inference mechanisms.

The output of the micro-project presents a deep study about causal discovery and causal inference. Moreover, the GitHub repository of datasets, papers, and code will be an excellent source of resources for those want to study the topic.

6. Linking language and semantic memory for building narratives

Proposal Submission Date: January 02, 2021
Actual Start Date: January 01, 2021
Expected Duration: 3 Months
Actual Duration: 3 Months

IRL, developed by Luc Steels and collaborators, is a parsing technique that captures the semantics of a natural language expression as a network of logical constraints. Determining the meaning of a sentence then amounts to finding a consistent assignment of variables that satisfies these constraints.

Typically, such meaning can only be determined (i.e., such constraints can only be resolved) by using the context ("narrative") in which the sentence is to be interpreted. The central hypothesis of this project is that modern large-scale knowledge graphs are a promising source of such contextual information to help resolve the correct interpretation of a given sentence.

We will develop an interface between an existing IRL implementation and an existing knowledge-graph reasoning engine to test this hypothesis. Evaluation will be done on a corpus of sentences from social-historical scientific narratives against corresponding knowledge graphs with social-historical data.

Expected Outputs:
- Software: an interface between nat.lang. parsing software (IRL) and reasoning software (knowledge graphs)

Actual Outputs:

- Web-Services library, (program/code),
  URL: https://github.com/SonyCSLParis/Catasto

Connection of Results to Work Package Objectives:
Natural language processing and understanding in machines often relies on statistical pattern recognition. What is missing here is the ability of a machine to describe in a human understandable way how it came to a certain interpretation. This would allow humans to take part in a machine's reasoning process, thereby facilitating human-computer interaction and collaboration.
By using IRL, the interpretation of an utterance is transparently expanded, and ambiguous entities are resolved until a single interpretation is found. At the same time, large datasets with semantic knowledge about the world exist in open repositories on the web. These repositories could be used in a similar way as we humans use our semantic memory, to disambiguate entities that cannot be resolved using the context of a dialogue alone.

7. AI Integration Languages: a Case Study on Constrained Machine Learning

Proposal Submission Date: March 01, 2021
Actual Start Date: March 01, 2021
Expected Duration: 4 Months
Actual Duration: 5 Months

Methods for injecting constraints in Machine Learning (ML) can help bridging the gap between symbolic and sub-symbolic models, and address fairness and safety issues in data-driven AI systems. The recently proposed "Moving Targets" approach achieves this via a decomposition, where a classical ML model deals with the data and a separate constraint solver with the constraints.

Different applications call for different constraints, solvers, and ML models: this flexibility is a strength of the approach, but it makes it also difficult to set up and analyze.

Therefore, this project will rely on the AI Domain Definition Language (AIDDL) framework to obtain a flexible implementation of the approach, making it simpler to use and allowing the exploration of more case studies, different constraint solvers, and algorithmic variants. We will use this implementation to investigate various new constraint types integrated with the Moving Targets approach (e.g., SMT, MINLP, CP).

Expected Outputs:

- Stand-alone moving targets system distributed via the AI4EU platform
- Interactive tutorial to be available on the AI4EU platform
- Scientific paper discussing the outcome of our evaluation and the resulting system

Actual Outputs:

- Example Jupyter Notebooks (3 data sets), (other),
  URL: https://gitsvn-nt.oru.se/uwe.kockemann/moving-targets
- Experiments Jupyter Notebooks (3 data sets), (other),

URL: https://gitsvn-nt.oru.se/uwe.kockemann/moving-targets
- Moving targets tutorial, (other),
  URL: https://gitsvn-nt.oru.se/uwe.kockemann/moving-targets
- Python library: Moving targets via AIDDL, (program/code),
  URL: https://gitsvn-nt.oru.se/uwe.kockemann/moving-targets

Connection of Results to Work Package Objectives:
T1.1 (Linking Symbolic and Sub-symbolic Learning)
Moving targets provides a convenient approach to enforce constraint satisfaction in sub-symbolic ML methods, within the limits of model bias. Our AIDDL integration pulls this idea all the way to the modeling level where, e.g., a fairness constraint can be added with a single line.
T1.4 (Compositionality and Auto ML)
The moving targets method, combined with an easy way of modeling constraints via AIDDL may increase trust in fully automated machine learning pipelines.
T2.6 (Dealing with Lack of Training Data)
Training data may be biased in a variety of ways depending on how it was collected. We provide a convenient way to experiment with constraining such data sets and possibly overcome unwanted bias due to lack of data.

8. Online Deep-AUTOML

Proposal Submission Date: January 01, 2021
Actual Start Date: February 01, 2021
Expected Duration: 4 Months
Actual Duration: 6 Months

Online AutoML in environments where the working conditions change over time.
The main goal consists of studying online optimization methods for hyper-parameter tuning. In dynamic environments, the “optimal” hyper-parameters might change over time.
Online AutoML will consist of an exploration phase followed by an exploitation phase.
The exploration phase is looking to find the set of hyper-parameters for the current working condition. The exploitation phase will continuously monitor the learning process to detect degradation in the performance of the system which triggers a new exploitation phase.
We will consider complex problems described by pipelines where each step in the pipeline has its own hyper-parameters. We will consider problems with many hyper-parameters where some of them might be irrelevant. Among the relevant parameters, the complexity of the model architecture (with particular reference to deep networks) is of particular relevance and will be the objective of our study.

Expected Outputs:
- 1 Conference Paper
- 1 Journal Paper
- 1 Prototype software

Actual Outputs:
- Hyper-Parameter Optimization for Latent Spaces in Dynamic Recommender Systems, (publication)

- Self Hyper-parameter tunning, (program/code),
  URL: https://github.com/BrunoMVeloso/ECMLPKDD2021
- Generator for preference data, (dataset),
  URL: https://github.com/BrunoMVeloso/ECMLPKDD2021

Connection of Results to Work Package Objectives:
The main goal of this micro-project is to develop tools that help people use sophisticated machine learning algorithms by helping in the parameterization of these algorithms.
The micro-project involved three groups from 3 different countries:

- INESC TEC, Portugal
- ICAR-CNR, Italy
- University Leiden, Netherlands

Based on this MP, the same groups will propose two new micro-projects.

1.3 Ongoing Micro Projects (About 50% Complete)

The following are the partially completed microprojects that address the challenges raised by WP1.

1. Coping with the variability of human feedback during interactive learning through ensemble reinforcement learning

Proposal Submission Date: January 01, 2021
Actual Start Date: January 01, 2021
Expected Duration: 4 Months
Actual Duration: 4 Months

This project entails robot online behavioral adaptation during interactive learning with humans. Specifically, the robot shall adapt to each human subject's specific way of giving feedback during the interaction. Feedback here includes reward, instruction and demonstration, and can be regrouped under the term "teaching signals". For example, some human subjects prefer a proactive robot while others prefer the robot to wait for their instructions; some only tell the robot when it performs a wrong action, while others reward correct actions, etc. The main outcome will be a new ensemble method of human-robot interaction which can learn models of various human feedback strategies and use them for online tuning of reinforcement learning so that the robot can quickly learn an appropriate behavioral policy. We will first derive an optimal solution to the problem and then compare the empirical performance of ensemble methods to this optimum through a set of numerical simulations.

Expected Outputs:

- Paper in IEEE RO-MAN or ACM/IEEE HRI or ACM CHI

Actual Outputs:

- Journal paper in preparation, (publication),
  URL: in preparation
- Open-source code to be uploaded on GitHub, (program/code),
  URL: in preparation
- Preprint to be made open on HAL, (publication),
  URL: in preparation

Connection of Results to Work Package Objectives:
Humans and robots are doomed to cooperate more and more within the society. This micro-project addresses a major AI challenge to enable robots to adapt on-the-fly to different situations and to different more-or-less naive human users. The solution consists in designing a robot learning algorithm which generalizes to a variety of simple human-robot interaction scenarios. Following the HumanE AI vision, interactive learning puts the human in the loop, prompting human-aware robot behavioral adaptation.
The micro-project directly contributes to one of the objectives of WP1 (T1.3) to enable 'continuous incremental learning in joint human/AI systems' by 'exploiting rich human feedback'. It also directly contributes to one of the objectives of WP3 (T3.3) to enable reflexivity and adaptation in Human AI collaboration.

2. Neural-Symbolic Integration: explainability and reasoning in KENN

Proposal Submission Date: January 03, 2021
Actual Start Date: September 01, 2021
Expected Duration: 3 Months
Actual Duration: 3 Months

This project builds on earlier work by FBK in Trento on KENN [1] and by VUA in Amsterdam [2] and aims to combine the insights of both. The project has 3 aims, depending on difficulty we may achieve one, two or all three.

1. The current version of KENN uses the Gödel t-conorm. We will develop versions of KENN based on other t-conorms (like the product t-conorm and Łukasiewicz), whose properties have been investigated in the earlier work by VUA. This should improve the performance of KENN.

2. We will try to extend the expressivity of the logical constraints in KENN from sets of clauses to implications, again using the earlier theoretical work by VUA. This should increase the reasoning capabilities of KENN.

3. It should be possible to check the exact contribution of each clause to the final predictions of KENN. This will increase explainability of KENN.

[1] https://arxiv.org/pdf/2009.06087.pdf
[2] https://arxiv.org/abs/2006.03472

Expected Outputs:
- paper describing improvements to KENN, published in workshop or conference
- software: new version of KENN

Actual Outputs:
- To be finalised

Connection of Results to Work Package Objectives:
Project has run for less than 50% of its allocated time yet.

3. Uncertainty Handling in Highly Automated Driving: Beyond Data

Proposal Submission Date: April 01, 2021
Actual Start Date: November 01, 2021
Expected Duration: 4 Months
Actual Duration: 6 Months

Involving human knowledge into the learning model can be done in diverse ways, e.g., by learning from expert data in imitation learning or reward engineering in deep reinforcement learning. In many applications, however, the expert data usually covers part of the search space or "normal" behaviors/scenarios. Learning a policy in the autonomous driving application under the limited dataset can make the policy vulnerable to novel or out of distribution (OOD) inputs and, thus, produce overconfident and dangerous actions. In this microproject, we aim to learn a policy based on the expert training data, while allowing the policy to go beyond data by interacting with an environment dynamics model and accounting uncertainty in the state estimation with virtual sensors. To avoid a dramatic shift of distribution, we propose to use the uncertainty of environment dynamics to penalize the policy for states that are different from human behavior.

Expected Outputs:

- Paper https://2021.ieee-iv.org and/or
- Paper http://www.auai.org/uai2020/

Actual Outputs:

- To be finalised

Connection of Results to Work Package Objectives:
Our key idea in this project is to learn a representation with deep models in a way to incorporate rules (e.g., physics equations governing dynamics of the autonomous vehicle) or distributions that can be simply defined by humans in advance. The learned representations from the source domain (the domain whose samples are based on the defined equations/distributions) are then transferred to the target domain with different distributions/rules and the model adapts itself by including target-specific features that can best explain variations of target samples w.r.t. underlying source rules/distributions. In this way, human knowledge is considered implicitly in the feature space.

Deviations from the Initial Plan:
We have refined the initial proposal according to the research focus in related projects that serve as a carrier for this project. We aim to develop a robust and generalized model that can perform well on out-of-distribution or novel data/domains/environments. The key idea is to learn the fundamental feature distribution shared between both source (training) and the target domain (test) while learning and including target-specific features that account for different variations of the shared distribution on the target domain.

4. Discovering Temporal Logic patterns as binary supervised learning

Proposal Submission Date: April 01, 2021
Actual Start Date: February 04, 2021

Expected Duration: 5 Months
Actual Duration: 5 Months

Making sense of data is a main challenge in creating human understandable descriptions of complex situations. When data refer to process executions, techniques exist that discover explicit descriptions in terms of formal models. Many research works envisage the discovery task as a one-class supervised learning job. Work on deviance mining highlighted nonetheless the need to characterise behaviours that exhibit certain characteristics and forbid others (e.g., the slower, less frequent), leading to the quest for a binary supervised learning task.
In this microproject we focus on the discovery of declarative process models, expressed through Linear Time Temporal Logic, as a binary supervised learning task, where the input log reports both positive and negative behaviours. We therefore investigate how valuable information can be extracted and formalised into a “optimal” model, according to user-preferences (e.g., model generality or simplicity). By iteratively including further examples, the user can also refine the discovered models

Expected Outputs:

- Paper to be submitted to relevant journal
- Machine learning tool
- Artificial data set

Actual Outputs:

- LoanApproval1, (dataset),
  URL: https://drive.google.com/drive/folders/15BwG4PJq8iIMh9Sr9dpMXAYBY-qp7QDE?usp=sharing
- LoanApproval2, (dataset),
  URL:https://drive.google.com/drive/folders/1fcJ8itzdMbNOjEAeV6nUEeI5B6__aB_c?usp=sharing
- Discovery Framework, (program/code),
  URL: https://zenodo.org/record/5158528

Connection of Results to Work Package Objectives:
The results of the microproject mainly contribute to WP1 (Human-in-the-Loop Machine Learning, Reasoning and Planning). Indeed, on the one hand, the micro-project aims at leveraging machine learning techniques (sub-symbolic learning) to provide LTL patterns (symbolic representation) of a set of "positive" traces, while excluding the "negative" ones (T1.1). On the other hand, the micro-project is a first step towards including the human in the loop of the discovery of LTL patterns representing all and only the cases the human wants to represent (T1.3). The user could indeed iteratively refine the discovered patterns so as to be sure to include all the cases she is interested to include, while excluding all those cases that she wants to exclude.

5. Feasibility analysis of hardware acceleration for AML

Proposal Submission Date: May 01, 2021
Actual Start Date: May 01, 2021
Expected Duration: 5 Months

Actual Duration: 5 Months

Algebraic Machine Learning (AML) offers new opportunities in terms of transparency and control. However, that comes along with many challenges regarding software and hardware implementations. To understand the hardware needs of this new method it is essential to analyze the algorithm and its computational complexity. With this understanding, the final goal of this microproject is to investigate the feasibility of various hardware options particularly in-memory processing hardware acceleration for AML.

Expected Outputs:
- Simulation model for a PIM architecture using AML
- Report

Actual Outputs:
- AML engine prototype using bitarrays, (program/code),
  URL: www.algebraic.ai

Connection of Results to Work Package Objectives:
This work is critical to speed up the calculation of Algebraic Machine Learning models and in so doing contribute to:
1- Bidirectional human-machine communication using formal expressions
2- Possibility to set goals and establish limits via formal constraints
3- Reduced dependency on statistics can help overcome bias
4- Transparency by design
5 -Possibility for decentralized, cooperative distributed machine learning

6. Interactive Reinforcement Learning for Humorous Agents

Proposal Submission Date: May 01, 2021
Actual Start Date: July 01, 2021
Expected Duration: 4 Months
Actual Duration: 6 Months

This project aims at investigating the construction of humor models to enrich conversational agents through the help of interactive reinforcement learning approaches.
Our methodology consists in deploying an online platform where passersby can play a game of matching sentences with humorous comebacks against an agent.
The data collected from these interactions will help to gradually build the humor models of the agent following state of the art Interactive Reinforcement Learning techniques.
We plan to work on this project for 4 months, resulting in an implementation of the platform, a first model for humor-enabled conversational agent and a publication of the obtained results and evaluations.

Expected Outputs:
- Online game for collecting humorous interaction data
- Humor models for conversational agents
- Paper in International Conference of Journal related to AI and AI in Games

Actual Outputs:

- Dataset - 1712 jokes, rated on a scale of 1 to 9 in terms of joke level, originality, positivity, entertainment, whether it makes sense and whether it is family-friendly, (dataset)
- Online Game - A game of matching sentences with humorous comebacks against an agent (like the game Cards Against Humanity), (program/code)

Connection of Results to Work Package Objectives:
The micro-project produced for HumaneAI-net a dataset of annotated associations between black and white cards following the game design of Cards Against Humanity. By doing so, the micro-project led to the creation of a unique dataset of humorous associations between concepts, annotated in terms of different humor styles by the participants of the experiment. The preliminary analysis on how the dataset can be leveraged to build different humor models for conversational agents is particularly relevant for the tasks T3.3 and 3.4 of WP3. Additionally, the micro project aims at exploring how to refine the humor models through an interactive learning approach, particularly relevant for the task T1.3 of WP1.

7. Adaption of ASR for Impaired Speech with minimum resources (AdAIS)

Proposal Submission Date: June 15, 2021
Actual Start Date: June 15, 2021
Expected Duration: 5 Months
Actual Duration: 5 Months

This micro project will study the adaptation of automatic speech recognition (ASR) systems for impaired speech. Specifically, the micro-project will focus on improving ASR systems for speech from subjects with dysarthria and/or stuttering speech impairment types of various degrees. The work will be developed using either German “Lautarchive” data comprising only 130 hours of untranscribed doctor-patient German speech conversations and/or using English TORGO dataset. Applying human-in-the-loop methods we will spot individual errors and regions of low certainty in ASR in order to apply human-originated improvement and clarification in AI decision processes.

Expected Outputs:

- Paper for ICASSP 2021 and/or Interspeech 2022

Actual Outputs:

- (publication)

Connection of Results to Work Package Objectives:
WP1 Learning, Reasoning and Planning with Human in the Loop
T1.1 Linking symbolic and sub-symbolic learning
WP3 Human AI Interaction and Collaboration
T3.1 Foundations of Human-AI interaction and Collaboration
T3.6 Language-based and multilingual interaction
T3.7 Conversational, Collaborative AI
WP6 Applied research with industrial and societal use cases
T6.3 Software platforms and frameworks
T6.5 Health related research agenda and industrial use cases

8. Multimodal Perception and Interaction with Transformers

Proposal Submission Date: April 01, 2023
Actual Start Date: April 01, 2018
Expected Duration: 6 Months
Actual Duration: 6 Months

Transformers and self-attention (Vaswani et al., 2017), have become the dominant approach for natural language processing (NLP) with systems such as BERT (Devlin et al., 2019) and GPT-3 (Brown et al., 2020) rapidly displacing more established RNN and CNN structures with an architecture composed of stacked encoder-decoder modules using self-attention.

This micro-project will provide tools and data sets for experiments and a first initial demonstration of the potential of transformers for multimodal perception and multimodal interactions. We will define research challenges, benchmark data sets and performance metrics for multimodal perception and interaction tasks such as (1) audio-visual narration of scenes, cooking actions and activities, (2) audio-video recordings of lectures and TV programs (3) audio-visual deictic (pointing) gestures, and (4) perception and evocation of engagement, attention, and emotion.

Expected Outputs:

- Benchmark data and performance targets for a phased set of research challenges of increasing difficulty.
- Tools for experiments to explore use of embeddings, encoder-decoders, self-attention architectures and related problems associated with applying transformers to different modalities.
- Concept demonstrations for simple examples of multimodal perception.

Actual Outputs:

- A survey of tools and datasets for multimodal perception with transformers, (dataset)
- A tutorial on the use of transformers for multimodal perception., (other)
- Research challenges for the use of transformers for multimodal perception and interaction., (other)

Connection of Results to Work Package Objectives:
This microproject will aid and encourage the use of a transformers and self-attention for multimodal modal interaction by Humane AI Net researchers, by identifying relevant tools and benchmark data sets, by providing tutorials and training materials for education, and by identifying research challenges for multimodal perception and interaction with Transformers.

1.4 Ongoing Micro Projects (Just Started or About to Start)

The following are newly started microprojects addressing the challenges raised in WP1.

1. Combining symbolic and sub-symbolic approaches - Improving neural Question-Answering-Systems through Document Analysis for enhanced accuracy and efficiency in Human-AI interaction.

Proposal Submission Date: February 01, 2021
Expected Duration: 6 Months

Knowledge discovery offers numerous challenges and opportunities. In the last decade, a significant number of applications have emerged relying on evidence from the scientific literature. AI methods offer innovative ways of applying knowledge discovery methods in the scientific literature facilitating automated reasoning, discovery and decision making on data. This micro-project will focus on the task of question answering (QA) for the biomedical domain. Our starting point is a neural QA engine developed by ILSP addressing experts' natural language questions by jointly applying document retrieval and snippet extraction on a large collection of PUBMED articles, thus, facilitating medical experts in their work. DFKI will augment this system with a knowledge graph integrating the output of document analysis and segmentation modules. The knowledge graph will be incorporated in the QA system and used for exact answers and more efficient Human-AI interactions. We will primarily focus upon scientific articles on Covid-19 and SARS-CoV-2.

Expected Outputs:

- Paper(s) in a conference or/and journal
- Demonstrator

2. Memory Aware Conversational AI to aid virtual Team-Meetings

Proposal Submission Date: May 03, 2021
Expected Duration: 6 Months

In this micro-project, we propose investigating human recollection of team meetings and how conversational AI could use this information to create better team cohesion in virtual settings.
Specifically, we would like to investigate how a person's emotion, personality, relationship to fellow teammates, goal and position in the meeting influences how they remember the meeting. We want to use this information to create memory aware conversational AI that could leverage such data to increase team cohesion in future meetings.
To achieve this goal, we plan first to record a multi-modal dataset of team meetings in a virtual setting. Second, administrate questionnaires to participants in different time intervals succeeding a session. Third, annotate the corpus. Fourth, carry out an initial corpus analysis to inform the design of memory-aware conversational AI.
This micro-project will contribute to a longer-term effort in building a computational memory model for human-agent interaction.

Expected Outputs:

- A corpus of repeated virtual team meetings (6 sessions spaced, 1 week each)

- manual annotations (people's recollection of the team meeting etc.)
- automatic annotations (e.g., eye-gaze, affect, body posture etc.)
- A paper describing the corpus and insights gained on the design of memory-aware agents from initial analysis

3. A Graph-Based Drift-Aware Data Cloning Process

Proposal Submission Date: September 01, 2021
Expected Duration: 6 Months

The goal is to devise a data generation methodology that, given a data sample, can approximate the stochastic process that generated it. The methodology can be useful in many contexts where we need to share data while preserving user privacy.
There are known literature for data generation based on Bayesian neural networks/hidden Markov models that are restricted to static and propositional data. We focus on time-evolving data and preference data.
We will study essentially two aspects: (1) the generator to produce realistic data, having the same properties of the original one. (2) we want to investigate how to inject drift within the data generation process in a controlled manner. The idea is to model the stochastic process through a dependency graph among random variables so that the drift can be simply modeled by changing the structure of the underlying graph through a morphing process.

Expected Outputs:

- 1 Conference/Journal Paper
- 1 Prototype
- Dataset Samples

2. Work Package 2: Multimodal Perception and Modeling

2.1 Overall Summary for Work Package 2:

The objective of WP2 is to provide a foundation for multimodal perception and modeling needed for natural interaction and collaboration between humans and intelligent systems.
Humans interact with the world using five major senses: sight, hearing, touch, smell, and taste. Almost all interaction with the environment is naturally multimodal, as audio, tactile or paralinguistic cues provide confirmation for physical actions and spoken language interaction. To interact and collaborate with people, intelligent systems must be able to perceive and model humans, human actions, and behaviors, human attention and awareness, human emotions, human language and human social interaction, as well as real-world human environments. Multimodal interaction seeks to fully exploit these parallel channels for perception and action to provide robust, natural interaction and collaboration with intelligent systems.
WP2 builds on recent advances in multimodal perception and modeling. Specifically, we define research challenges for natural multimodal interaction, document and augment the existing tools and data sets needed to meet such challenges and demonstrate the use of recent advances in machine learning for building systems.
Work on multimodal perception and modeling is organized around the following activities: Learning of multimodal models grounded in physical reality (T2.1), Multimodal perception and modeling of actions, activities and tasks (T2.2), Multimodal perception of awareness, emotions, and attitudes (T2.3), Perception of Social Signals and Social Dynamics (T2.4), Distributed Collaborative Perception and Modeling (T2.5), Dealing with lack of labeled training data (T2.6), and Assembling benchmark datasets (T2.7).

2.2 Micro Projects related to WP2

During the first 12 months, seven micro-projects have addressed the challenges raised in multimodal perception and modeling.

1. Collection of datasets tailored for HumanE-AI multimodal perception and modelling
2. Causality and Explainability in Temporal Data
3. Machine supervision of human activity: The example of rehabilitation exercises
4. Prediction of static and perturbed reach goals from movement kinematics
5. AI Integration Languages: a Case Study on Constrained Machine Learning
6. Multimodal Perception and Interaction with Transformers
7. Prediction of static and perturbed reach goals from movement kinematics

The first 5 of these microprojects have recently completed. The final two projects are currently underway. All of these microprojects have been described by 2-minute video presentations in June 2021.

2.3 Descriptions of Microprojects addressing Multimodal Perception and Modeling.

The following work packages address challenges raised by WP2.

1. Collection of datasets tailored for HumanE-AI multimodal perception and modelling

Proposed by: Mathias Ciliberto, University of Sussex
Other Partners: Vitor Fortes Rey, DFKI, Arno de Bois, VUB
Related WPs: WP1, WP2, WP5
Submission: January 01, 2021
Start Date: January 11, 2021
Duration: 5 Months

HumanE-AI research needs data to advance. Often, researcher struggle to progress for the lack of data. At the same time, collecting a rich and accurate dataset is no easy task. Therefore, we propose to share through the AI4EU platform the datasets already collected so far by different research groups. The datasets will be curated to be ready-to-use for researchers.
Possible extension and variation of such datasets will also be generated using artificial techniques and published on the platform. A performance baseline will be provided for each dataset, in form of publication reference, developed model or written documentation. The relevant legal framework will be investigated with specific attention to privacy and data protection, as to highlight limitations and challenges for the use and extension of existing datasets as well as future data collection on the subject of multimodal data collection for perception modelling.

Expected Outputs:

- Publication of OPPORTUNITY dataset (and other datasets if time available) on the AI4EU platform.
- Publication of baseline performance pipeline for OPPORTUNITY dataset (and other datasets if time available) on AI4EU platform.
- Investigation of data loader and pipeline integration on AI4EU experiment to load HAR dataset and pre-existent pipelines, with a focus on the opportunity dataset (and other datasets if time available)
- Generation of variation
- Survey publications describing datasets and performance baseline

Actual Outputs:
(These data sets have been placed on the HumaneAI Server, in preparation for a stable version of the availability of AI4EU Platform)

- Opportunity++, (dataset)
- CapacitiveGym, (dataset)
- HCI FreeHand, (dataset)
- SkodaMini, (dataset)
- Wearlab BeachVolleyball, (dataset),
  URL: https://ieee-dataport.org/open-access/wearlab-beach-volleyball-serves-and-games

Connection of Results to Work Package Objectives:
Multi-modal perception and modeling need data to progress but recording a new rich and accurate dataset allowing for comparative evaluations by the scientific community is no easy task. Therefore, we gathered rich datasets for multimodal perception and modelling of human activities and gestures. We curated the dataset in order to make them easy to use for research thanks to clear documentation and file formats.

The highlight of this microproject is the OPPORTUNITY++ dataset of activities of daily living, a multi-modal extension of the well-established OPPORTUNITY dataset. We enhanced this dataset which contains wearable sensor data, with previously unreleased data, including video and motion tracking data, which make OPPORTUNITY++ a truly multi-modal dataset with wider appeal, such as to the computer vision community. In addition, we released other well established activity datasets (HCI FreeHand and SkodaMini dataset) as well as datasets involving novel sensor modalities (CapacitiveGym) and skill-assessment dataset (Wearlab BeachVolleyball)

2. Causality and Explainability in Temporal Data

Proposed by: Joao Gama, INESC TEC
Other Partners: Dino Pedreschi, Fosca Giannotti
Related WPs: WP1, WP2, WP5
Submission: January 01, 2021
Start Date: February 01, 2021
Duration: 6 Months

Machine Learning models are used in decision-making processes in real-world problems. Such models learn a function that maps the observed features with the decision outcomes. However, these models usually do not convey causal information about the association in observational data, thus not being easily understandable for the average user, therefore not being possible to retrace the models' steps, nor rely on its reasoning. Hence, it is natural to investigate more explainable methodologies, such as causal discovery approaches, since they apply processes that mimic human reasoning. For this reason, we propose the usage of such methodologies to create more explicable models that replicate human thinking, and that are easier for the average user to understand. More specifically, we suggest its application in methods such as decision trees and random forest, since by themselves are highly explainable correlation-based methods.

Expected Outputs:
- 1 Conference Paper
- 1 Prototype
- Dataset Repository

Actual Outputs:
- Methods and Tools for Causal Discovery and Causal Inference, (publication)
- Repository of datasets, software, and papers related to causal discovery and causal inference research, (dataset),
  URL: https://github.com/AnaRitaNogueira/Methods-and-Tools-for-Causal-Discovery-and-Causal-Inference

Connection of Results to Work Package Objectives:
The HumanE-AI project thinks a society of increasing interactions between humans and artificial agents. All around the project, causal models are relevant for plausible models of human behavior, man-machine explanations, and upgrading machine-learning algorithms with causal inference mechanisms.
The output of the micro-project presents a deep study about causal discovery and causal inference. Moreover, the GitHub repository of datasets, papers, and code will be an excellent source of resources for those want to study the topic.

3. Machine supervision of human activity: The example of rehabilitation exercises

Proposed by: Andras Lőrincz, Eötvös Loránd University
Other Partners: Elsa Kirchner, DFKI
Related WPs: WP2
Submission: January 01, 2021
Start Date: January 03, 2021
Duration: 5 Months

We propose research on a scalable human-machine collaboration system with the goal of executing high quality actions for rehabilitation exercises. We combine video and speech for video-grounded goal-oriented dialogue. We build on our video and text database. The database has exercises for rehabilitation following knee injuries. We evaluate high performance body pose estimation tools and compare it to a real-time body pose estimation tool to be developed for smartphones via 'knowledge distillation' methods.
The complementing part of the project deals with the texts that we have collected for these exercises and estimates the amount of texts needed for dialogues that can lead and correct the quality of exercises. Potential topics/intents include pose relative to camera, proper light conditions, audio-visual information about pain, notes about execution errors, errors discovered by the computer evaluations, requests about additional information from the patient, and reactions to other, unrelated queries.

Expected Outputs:
- Dataset of the dialogues
- Publication on the constraints and potentials of existing state-of-the-art methods
- Performance evaluation methods and usability studies

Actual Outputs:
- DeepRehab: Real Time Pose Estimation on the Edge for Knee Injury Rehabilitation, (publication),
  URL: https://e-nns.org/icann2021/
- Multimodal technologies for machine-assisted physical rehabilitation, (publication)

Connection of Results to Work Package Objectives:
Video-based dialogue systems meet the goals of the Foundations of Human-AI interactions, whereas the rehabilitation scenario is a prototype for goal-oriented collaboration. The microproject targeted specific topics, including

(i) body motion and pain both in terms a language and potential dialogues and in more than 400 video samples that included 50 exercises and about 7 errors on the average to be detected alone or in combinations for each motion types and
(ii) (ii) dialogues from experts and crowdsourcing based dialogue enhancements

4. Prediction of static and perturbed reach goals from movement kinematics

Proposed by: Patrizia Fattori, University of Bologna
Other Partners: DFKI
Submission: February 01, 2021
Start: February 01, 2021
Duration: 5 Months

Reaching movements towards targets located in the 3-dimensional space are fast and accurate. Although they may seem simple and natural movements, they imply the integration of different sensory information that is carried in real time by our brain. We will apply machine learning techniques to address different questions as it follows: i) at which point of the movement is it accurately possible to predict the final target goal in static and dynamic conditions? ii) as at behavioural level it was hypothesized that direction and depth dimension do not rely on shared networks in the brain during the execution of movement but they are processed separately, can the targets located along the horizontal or sagittal dimension be predicted with the same or different accuracy? Finally, we will frame our result in the context of improving user-agent interactions, moving from a description of human movement to a possible implementation in social/collaborative AI.

Expected Outputs:
- A model descriptive of reaching movement in static and dynamic conditions
- A research paper submitted on a relevant journal of the sector

Actual Outputs:
- Individual subject trajectories, (dataset),
URL: https://drive.google.com/drive/folders/1FdDXKjhCupDdyLlyvCdUwfxmyGfoRZDE
- Recurrent neural network codes, (program/code),
URL: https://drive.google.com/drive/folders/1FdDXKjhCupDdyLlyvCdUwfxmyGfoRZDE
- Manuscript, (publication), URL: in preparation

Connection of Results to Work Package Objectives:
In this microproject, the action goal could change its spatial position during action execution and in unpredictable way. Using a double neural network approach, the present results contribute to the objectives of the Task 2.2 of WP2 at two levels. In the first level, we described the temporal structure of action goal recognition in static and perturbed condition of reaching from movement kinematics of the index and wrist in the 3-dimensional space. This first achievement contributes to the recognition of the action goal in a context that is known (static targets) or not, a priori (perturbed targets). In the second level, we predicted future trajectory of the movement given previous action path. This second achievement contributes to the creation of the bases for the design of a system able to monitor activity in a natural human workspace and extract prediction of future actions in situations that could require human-AI interaction.

Deviations from the Initial Plan:
The first approach used to predict static and perturbed reaching endpoints was an unsupervised learning method called velocity-based multiple change-point inference (vMCI) based on Bayesian inference. This method detected changes in reaching movements towards a target and it was applied to the velocity and position of the hand markers of available dataset. However, no changes in the velocity could be observed in the perturbed condition around the time point where the target jumped since a small jump of targets occurred. These findings suggest that, in this dataset, supervised methods are more successful in predicting action goals than unsupervised ones.

5. AI Integration Languages: a Case Study on Constrained Machine Learning

Proposed by: Uwe Köckemann, Orebro Univ
Related WP: WP1, WP2
Submission Date: March 01, 2021
Start Date: March 01, 2021
Duration: 5 Months

Project Description
Methods for injecting constraints in Machine Learning (ML) can help bridging the gap between symbolic and sub-symbolic models, and address fairness and safety issues in data-driven AI systems. The recently proposed “Moving Targets” approach achieves this via a decomposition, where a classical ML model deals with the data and a separate constraint solver with the constraints. Different applications call for different constraints, solvers, and ML models: this flexibility is a strength of the approach, but it makes it also difficult to set up and analyze.

Therefore, this project will rely on the AI Domain Definition Language (AIDDL) framework to obtain a flexible implementation of the approach, making it simpler to use and allowing the exploration of more case studies, different constraint solvers, and algorithmic variants. We will use this implementation to investigate various new constraint types integrated with the Moving Targets approach (e.g., SMT, MINLP, CP).

Expected Outputs:
- Stand-alone moving targets system distributed via the AI4EU platform
- Interactive tutorial to be available on the AI4EU platform
- Scientific paper discussing the outcome of our evaluation and the resulting system

Actual Outputs:
- Example Jupyter Notebooks, (3 datasets), (program/code),
  URL: https://gitsvn-nt.oru.se/uwe.kockemann/moving-targets
- Experiments Jupyter Notebooks (3 data sets),
  URL: https://gitsvn-nt.oru.se/uwe.kockemann/moving-targets
- Moving targets tutorial, (other),
  URL: https://gitsvn-nt.oru.se/uwe.kockemann/moving-targets
- Python library: Moving targets via AIDDL, (program/code)
  URL: https://gitsvn-nt.oru.se/uwe.kockemann/moving-targets

Connection of Results to Work Package Objectives:
T1.1 (Linking Symbolic and Sub-symbolic Learning)
Moving targets provides a convenient approach to enforce constraint satisfaction in sub-symbolic ML methods, within the limits of model bias. Our AIDDL integration pulls this idea all the way to the modeling level where, e.g., a fairness constraint can be added with a single line.
T1.4 (Compositionality and Auto ML)
The moving targets method, combined with an easy way of modeling constraints via AIDDL may increase trust in fully automated machine learning pipelines.
T2.6 (Dealing with Lack of Training Data)
Training data may be biased in a variety of ways depending on how it was collected. We provide a convenient way to experiment with constraining such data sets and possibly overcome unwanted bias due to lack of data.

6. Multimodal Perception and Interaction with Transformers

Proposed by: James Crowley, Inria
Other Partners: Andras Lorincz (Eotvos Lorand University - ELTE), Dominique Vaufreydaz, Fabien Ringeval (Univ Grenoble Alpes), Camille Guinaudeau, Marc Evrard (Uni Paris Saclay, Marko Grobelnik (Jozef Stefan Institut-JSI), Pavel Pecina (Charles University)
Related WP: WP1, WP2, WP3
Submission Date: March 01, 2021
Start Date: April 01, 2021
Duration: 6 Months

Transformers and self-attention (Vaswani et al., 2017), have become the dominant approach for natural language processing (NLP) with systems such as BERT (Devlin et al., 2019) and GPT-3 (Brown et al., 2020) rapidly displacing more established RNN and CNN structures with an architecture composed of stacked encoder-decoder modules using self-attention.
This micro-project will survey tools and data sets for experiments for demonstrating the potential use of transformers for multimodal perception and multimodal interactions. We will define research challenges and performance metrics for multimodal perception tasks such as audio-visual narration of scenes, cooking actions and activities, audio-visual deictic (pointing) gestures, and perception and evocation of engagement, attention, and emotion. We will provide tutorials on the use of transformers for multimodal perception and interaction.

Expected Outputs:
- Performance targets for a phased set of research challenges of increasing difficulty.
- A survey of tools and benchmark data for experiments to explore use of embeddings, encoder-decoders, self-attention architectures and related problems associated with applying transformers to different modalities.
- Concept demonstrations for simple examples of multimodal perception.

Actual Outputs:
- A survey of tools and datasets for a multimodal perception with transformers

- A tutorial on the use of transformers for multimodal perception.
- Research challenges for the use of transformers for multimodal perception and interaction.

Connection of Results to Work Package Objectives:
This microproject will aid and encourage the use of a transformers and self-attention for multimodal modal interaction by Humane AI Net researchers, by identifying relevant tools and benchmark data sets, by providing tutorials and training materials for education, and by identifying research challenges for multimodal perception and interaction with Transformers.

Bibliography

1) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762
2) Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
3) Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., and Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

7. Neural mechanism in human brain activity during weight-lifting

Submitted by: Sencer Melih DENİZ, Information Technologies Institute, BILGIM and
Other Partners: DFKI
Related Workpackages: WP2
Submission: January 04, 2021
Start Date: May 03, 2021
Duration: 4 Months

Understanding the mechanism of the neural correlates during human physical activities is important for providing safety in industrial factory environments considering brain activity during lifting a weight. Moreover, different responses to the same task can be observed due to physiological and neurological differences among individuals. In this project, the change pattern in EEG will be investigated during lifting of a weight and the features in EEG data making difference during lifting a weight will be analyzed. Classification between lifting and no lifting cases will be realized by using deep learning-based machine learning methods. The outcomes of the project can be applied in industrial exoskeleton applications as well as physical rehabilitation of stroke patients.

Expected Outputs:
- Dataset Repository (Share on AI4EU)
- Conference Paper / Journal Article

Actual Outputs:
- Data Acquisition Software Code

Connection of Results to Work Package Objectives:
This project is also part of WP2 with task numbers T2.2, T2.3. This project aims to contribute to WP2 and WP6 by investigating the use case of EEG signal and AI models in the detection of various aspects of physical activities during weightlifting. To investigate pattern change in EEG during weightlifting will be aimed at providing more information in prediction of

intended and actual human actions during sensorimotor tasks. Doing so, a common research question is aimed to be applied to the more industrial use cases such as control of exoskeletons. Moreover, outcomes of the project can be used for contribution in increasing mobility in stroke patients and disabled people as related with healthy living and mobility.

3. Work Package 3: Human AI Collaboration and Interaction

3.1 Overall Summary for Work Package 3:

The objective of WP3 "Human AI Interaction and Collaboration" is to establish new methodological and conceptual basis for human-AI collaboration. In particular, the goal is to develop HCI and AI methodology for social basis for human-AI partnership, especially group cognition and emotional expression. For AI to understand people, it needs to both be able to infer intentions and emotions from observations as well as make its own intentions understandable to human partners via grounding, emotional expression, and explanation. We believe that these capabilities need to be to some extent be engineered into AI, in order to ensure more natural behavior from first interaction and to reach a desirable level of controllability and transparency. However, they need to be made interactive for users to control and understand.

The intelligent systems studied in WP3 may be in the form of embodied agents, be them physical robots, animated characters, physical interactive objects, smart environments, or simply software systems. Humans and systems may interact through visual displays, physical devices, acoustic signals, printed text, spoken language, or other modalities. This WP will work to advance and implement the HumanE AI Net vision of allowing such interaction to take the form of synergetic collaboration and co-creation leveraging the new Human in the Loop learning, reasoning and planning methods of WP 1 and the advances in perception and world modeling from WP2.

WP3 builds on human-computer interaction, especially on theories of human-machine communication, while integrating them with relevant advances in machine learning methods. Work on human AI interaction and collaboration is organized around the following tasks:

T3.1 Foundations of Human-AI interaction and Collaboration;
T3.2 Human-AI Interaction/collaboration paradigms;
T3.3 Reflexivity and Adaptation in Human AI collaboration;
T3.4 User Models and Interaction History;
T3.5 Visualization Interactions, and Guidance;
T3.6 Language-based and Multilingual Interaction;
T3.7 Conversational, Collaborative AI; and
T3.8 Trustworthy Social and Sociable interaction.

3.2 Microprojects related to WP3

During the first 12 months, 14 micro-projects have addressed the foundations of interactive AI:

1. DIASER: DIAlog task-oriented annotations for enhanced modeling of uSER
2. Autobiographical Recall in Virtual Reality
3. Machine supervision of human activity: The example of rehabilitation exercises
4. Normative behavior and extremism in Facebook groups
5. Multilingual Event-Type-Anchored Ontology for Natural Language Understanding (META-O-NLU)
6. Socially aware interactions
7. Social interactions with robots
8. Evidence-based chatbot interaction aimed at reducing sedentary behavior
9. Social dilemma with information asymmetry
10. Proactive communication in social robots

11. Learning Individual Users' Strategies for Adaptive UIs
12. The knowledgeable and empathic behavior change coach
13. Exploring the impact of Agency on Human-Computer Partnerships
14. Evaluating segmentation in automatic captioning systems

The first four (1-4) of these microprojects have recently completed. The rest projects are currently underway. All of these microprojects have been described by 2-minute video presentations during Summer 2021.

3.3 Completed Micro Projects

The following completed microprojects address challenges raised by WP3.

1. DIASER: DIAlog task-oriented annotations for enhanced modeling of uSER

Proposal Submission Date: December 07, 2020
Actual Start Date: January 01, 2021
Expected Duration: 4 Months
Actual Duration: 4 Months

We aim evaluate the usefulness of current dialogue dataset annotation and propose annotation unification and automatized enhancements for better user modeling by training on larger amounts of data. Current datasets' annotationis often only focused on annotation geared toward the dialog system learning how to answer, while the user representation should be explicit, consistent and as complete as possible for more complex user representation (e.g., cognitively). The project will start from existing annotated dialog corpora and produce extended versions, with improved annotation consistency and extra user representation annotations produced automatically from existing corpora like bAbI++ and MultiWOZ and others. We will explore unifying annotations from multiple datasets and evaluate the enhanced annotation using our own end-to-end dialogue models based on memory networks. Connection with T3.7 and T3.4 is straightforward, since the task-oriented dialogue systems are the very definition of conversational, collaborative AI. T3.6 will be addressed through round-trip translation for data augmentation.

Expected Outputs:

- Extended and unified versions of publicly available dialog corpora with explicit user modeling annotations (bAbI++, MultiWOZ etc.)
- a report and papers describing a unified user modeling annotation scheme with respect to existing dialog annotation datasets and the results of some baseline experiments using the annotated data produced by the project.
-

Actual Outputs:

- DIASER corpus, (dataset),
  URL: https://gitlab.com/ufal/dsg/diaser

Connection of Results to Work Package Objectives:
By providing an open annotated dialogue resource with unified and enhanced annotations, DIASER offers to the community linguistic material usable both for machine learning

experiments and for testing dialog model properties in relation with dialog history management, dialog consistency checking and user modeling aspects.
The result of DIASER is related to issues pertaining to the following tasks: mainly
T3.6 Language Based and Multilingual Interaction
with potential links to
T3.7 Conversational, Collaborative AI,
T3.2 Human AI Interaction / Collaboration Paradigms,
T3.4 User Models and Interaction History,
T3.3 Reflexivity and Adaptation in Human AI collaborations.

2. Autobiographical Recall in Virtual Reality

Proposal Submission Date: January 01, 2021
Actual Start Date: January 02, 2021
Expected Duration: 4 Months
Actual Duration: 2 Months

We propose to research how autobiographical recall can be detected in virtual reality (VR). In particular, we experimentally investigate what physiological parameters accompany interaction with autobiographical memories in VR. We consider VR as one important representation of Human-AI collaboration.
For this, we plan to (1) record an EEG data set of people's reaction and responses when recalling an autobiographical memory, (2) label the data set, and (3) do an initial analysis of the dataset to inform the design of autobiographical VR experiences. We would try to automate data collection as much as possible to make it easy to add more data over time.
This will contribute to a longer-term effort in model and theory formation. The main Contribution is to WP3. This is set in Task 3.2: Human-AI Interaction/collaboration paradigms and aims at better understanding user emotion in VR to model self-relevance in AI collaboration Task 3.4.

Expected Outputs:

- Dataset on autobiographic recall in VR
- A manuscript describing the data set and initial insights into autobiographical recall in VR

Actual Outputs:

- Video of the VR experience, (other),
  URL: https://www.youtube.com/watch?v=mGb7Oi5CHNc
- Pilot dataset (currently recordings are under way, and we expect the full dataset to be available by the end of the year), (dataset),
  URL: https://github.com/kgupta2789/AMinVR
- Workshop on Human Memory and AI, (other),
  URL: https://www.humane-ai.eu/event/ai-and-human-memory/

Connection of Results to Work Package Objectives:
The main Contribution is to WP3. This is set in Task 3.2: Human-AI Interaction/collaboration paradigms and aims at better understanding user emotion in VR to model self-relevance in

AI collaboration Task 3.4. The VR experience is implemented in Unity, and we are happy to share this in the context of a joint project.

Deviations from the Initial Plan:
Due to the second wave of coronavirus infections in March 2020 and the associated restrictions in Germany, we could not obtain physiological data or pilot the experiment. Due to this, we have co-hosted a workshop on AI and Memory and by that refined our research project. Nevertheless, we are currently collecting data and are making them accessible in a repository.
The user studies are currently conducted as we are allowed back in the lab. The full dataset and the paper will hence be ready by the end of the year.

3. Machine supervision of human activity: The example of rehabilitation exercises

Proposal Submission Date: January 01, 2021
Actual Start Date: January 03, 2021
Expected Duration: 4 Months
Actual Duration: 5 Months

We propose to research a scalable human-machine collaboration system with the common goal of executing high quality actions (e.g., in rehabilitation exercise). We combine video and speech for video-grounded goal-oriented dialogue. We build on our video and text database. The database has exercises for rehabilitation following knee injuries. We evaluate high performance body pose estimation tools and compare it to a real-time body pose estimation tool to be developed for smartphones via ‘knowledge distillation’ methods.

The complementing part of the project deals with the texts that we have collected for these exercises and estimates the amount of texts needed for dialogues that can lead and correct the quality of exercises. Potential topics/intents include pose relative to camera, proper light conditions, audio-visual information about pain, notes about execution errors, errors discovered by the computer evaluations, requests about additional information from the patient, and reactions to other, unrelated queries.

Expected Outputs:
- Dataset of the dialogues
- Publication on the constraints and potentials of existing state-of-the-art methods
- Performance evaluation methods and usability studies

Actual Outputs:
- DeepRehab: Real Time Pose Estimation on the Edge for Knee Injury Rehabilitation, (publication),
  URL: https://e-nns.org/icann2021/
- Multimodal technologies for machine-assisted physical rehabilitation, (publication),
  URL: submitted

Connection of Results to Work Package Objectives:
Video-based dialogue systems meet the goals of the Foundations of Human-AI interactions, whereas the rehabilitation scenario is a prototype for goal-oriented collaboration. The microproject targeted specific topics, including

- body motion and pain both
  - in terms a language and potential dialogues and
  - in more than 400 video samples that included 50 exercises and about 7 errors on the average to be detected alone or in combinations for each motion types
- dialogues
  - from experts and
  - crowdsourcing based dialogue enhancements

4. Normative behavior and extremism in Facebook groups

Proposal Submission Date: February 01, 2021
Actual Start Date: April 05, 2021
Expected Duration: 4 Months
Actual Duration: 4 Months

In this project we will investigate whether normative behavior can be detected in facebook groups. In a first step we will hypothesize about possible norms that could lead to a group becoming more extreme on social media, or whether groups that become more extreme will develop certain norms that distinguish them from other groups and that could be detected. An example of such a norm could be that a (self-proclaimed) leader of a group is massively supported by retweets, likes or affirmative messages, along with evidence of verbal sanctioning toward counter-normative replies. Simulations and analyses of historical facebook data (using manual detection in specific case studies and more broadly through NLP) will help revealing the existence of normative behavior and its potential change over time.

Expected Outputs:
- Report describing guidelines to detect normative behavior on social media platforms

Actual Outputs:
- Identification of radical behavior in Parler groups, (other)
- Characterizing the language use of radicalized communities detected on Parler, (other)

Connection of Results to Work Package Objectives:
In order to see how individuals might contribute to behavior that is not in the interest of society we cannot analyze one social media platform. Especially more extremist expressions quickly disappear from mainstream social media to niche platforms that can quickly change over time. Thus, the connection between individual and societal goals is difficult to observe by just analyzing data from a single social media platform. In the other hand it is very difficult to link users between platforms.

3.4 Ongoing Micro Projects (About 50% Complete)

The following are partially completed microprojects that address challenges raised in WP3.

1. Multilingual Event-Type-Anchored Ontology for Natural Language Understanding (META-O-NLU)

Proposal Submission Date: January 01, 2021
Actual Start Date: January 01, 2021
Expected Duration: 6 Months
Actual Duration: 12 Months

Many industrial NLP applications emphasize the processing and detection of nouns, especially proper nouns (Named Entity Recognition, NER). However, processing of verbs has been neglected in recent years, even though it is crucial for the development of full NLU systems, e.g., for the detection of intents in spoken language utterances or events in written language news articles. The META-O-NLU microproject focuses on proving the feasibility of a multilingual event-type ontology based on classes of synonymous verb senses, complemented with semantic roles and links to existing semantic lexicons. Such an ontology shall be usable for content- and knowledge-based annotation, which in turn shall allow for developing NLU parsers/analyzers. The concrete goal is to extend the existing Czech-English SynSemClass lexicon (which displays all the necessary features, but only for two languages) by German and Polish, as a first step to show it can be extended to other languages as well.

Expected Outputs:

- Common paper co-authored by the proposers (possibly with et. partners)
- Extended version of SynSemClass (entried in additional languages)

Actual Outputs:

- SynSemClass 3.5 dataset, (dataset),
  URL: http://hdl.handle.net/11234/1-3750
- SynSemClass 3.5 browser, (other),
  URL: https://lindat.cz/services/SynSemClass35/

Connection of Results to Work Package Objectives:
Task 3.6 focuses on both spoken and written language-based interactions (dialogues, chats), in particular, questions of multilinguality that are essential to the European vision of human-centric AI. The results of this microproject contribute especially to the multilingual issue and is directed to full NLU (Natural Language Understanding) by describing event types, for which no general ontology exists yet. The resulting resource will be used for both text and dialog annotation, to allow for evaluation and possibly also for training of NLU systems.

Deviations from the Initial Plan:
The training of the annotators for the dataset creation took longer (also, they were hired later than at project start). We are now in the process of negotiating microproject extension to Dec. 2021 (with only 3PM extra for annotation, to make the best use of the trained annotators).

2. Socially aware Interactions

Proposal Submission Date: January 04, 2021
Actual Start Date: March 01, 2021
Expected Duration: 4 Months
Actual Duration: 6 Months

In order for systems to function effectively in cooperations with humans and other AI systems they have to be aware of their social context. Especially in their interactions they should take into account the social aspects of their context, but also can use their social context to manage the interactions. Using the social context in the deliberation about the interaction steps will allow for an effective and focused dialogue that is geared towards a specific goal that is accepted by all parties in the interactions.
In this project we will start with the Dialogue Trainer system that allows for authoring very simple but directed dialogues to train (medical) students to have effective conversations with patients. Based on this tool, in which social context is taken into account only through the authors of the dialogue, we will design a system that will actually deliberate about the social context.

Expected Outputs:

- software prototype for a flexible dialogue trainer system
- CONVERSATIONS workshop paper 2021

Actual Outputs:

- Prototype of dialogue system, (program/code)
- Socially Aware Interactions: Towards a flexible dialogue system [Submitted], (publication)

Connection of Results to Work Package Objectives:
First, the dialogue system's flexibility and context-awareness will make the conversational agent appear more natural/realistic to the user, which is significant for the "Human-AI collaboration and interaction" work package.
Furthermore, in the system, the agent and the human user, besides having their own individual goals, are also attempting to achieve a dialogue goal together (e.g., in an anamnesis scenario, the main goal could be to obtain/give a diagnosis), which satisfies the "Societal AI" work package's goal "AI systems' individual vs collective goals".
This last work package includes the goal "Multimodal perception of awareness, emotions, and attitudes" as well, which is met because the agent adapts to changes in context, deliberating on top of it, and becoming more socially aware.

3. Social interactions with robots

Proposal Submission Date: January 25, 2021
Actual Start Date: February 01, 2021
Expected Duration: 4 Months
Actual Duration: 6 Months

Robots are already in wide use in industrial settings where the interactions with people are well structured and stable. Interactions with robots in home settings are notoriously more difficult. The context of interactions changes over time, depending on the people present, the time of day, the event going on, etc. In order to cope with all these factors creating uncertainty and ambiguity people use practices, norms, conventions, etc. to normalize and package certain interactions into standard types of actions performed in order by the parties involved, e.g., getting coffee.
Within this project we will explore how the idea of social practices to regulate interactions and create expectations in the parties involved can be used to guide robots in their interactions with people. We will explore a simple scenario with a Pepper robot to explore all practical obstacles when using these concepts in robotics.

Expected Outputs:
- Three MSc thesis reports
- demo software
- documented example on the AI4EU platform

Actual Outputs:
- Ai Planning with Social Practices for the Pepper robot., (other)

Connection of Results to Work Package Objectives:
The project shows how social practices can be used to guide human-robot interactions. This provides a social context that can be helpful to adapt the actions of the robot to both the situation and the user. The project was a very first attempt to create a practical implementation and thus can only be seen as a basis on which further work can be done to really take advantage of all aspects of social practices.

4. Evidence-based chatbot interaction aimed at reducing sedentary behavior

Proposal Submission Date: February 01, 2021
Actual Start Date: January 03, 2021
Expected Duration: 4 Months
Actual Duration: 5 Months

Interaction between chatbots and humans is often based on frequently occurring interaction patterns, e.g., question – answer. Those patterns usually describe a very brief phase in the interaction. In this micro project, we want to investigate whether we can design a chatbot for behavior change by including higher level patterns, which are adapted from the taxonomy of behavior change techniques (BCT's). These patterns should describe the components of the interaction during a longer period of time. In addition, we will investigate how to design a user interface in such a way that it sustains the interest of the users.
We will focus on reducing sedentary behavior, and especially sitting behavior, which can have negative health consequences. The interaction patterns and user interface will be implemented in a prototype. A user study will be performed to evaluate the different components on effectiveness and engagement.

Expected Outputs:
- Scientific paper

- Prototype for the consortium

Actual Outputs:
- Bachelor thesis, (publication)

Connection of Results to Work Package Objectives:
We investigate how AI systems can collaborate with humans, specifically focussing on changing a specific behavior. We increase our understanding of how specific interaction forms between an AI system and a human are effective in achieving behavior change. We also investigate to what extent knowledge about health behavior can contribute to desiging realistic and effective communication.

5. Social dilemma with information asymmetry

Proposal Submission Date: March 15, 2021
Actual Start Date: April 01, 2021
Expected Duration: 5 Months
Actual Duration: 6 Months

Social dilemmas are situations in which the interests of the individuals' conflict with those of the team, and in which maximum benefit can be achieved if enough individuals adopt prosocial behavior (i.e., focus on the team's benefit at their own expense). In a human-agent team, the adoption of prosocial behavior is influenced by various features displayed by the artificial agent, such as transparency, or small talk. One feature still unstudied is expository communication, meaning communication performed with the intent of providing factual information without favoring any party.

We will implement a public goods game with information asymmetry (i.e., agents in the game do not have the same information about the environment) and perform a user-study in which we will manipulate the amount of information that the artificial agent provides to the team and examine how varying levels of information increase or decrease human prosocial behavior.

Expected Outputs:
- Submission to one of the following: International Journal of Social Robotics, Behaviour & Information Technology, AAMAS, or CHI. Submission to be sent by the end of August 2021.
- Release of the game developed for the study on the AI4EU platform to allow other researchers to use it and extend it
- Educational component on the Ethical aspect of AI, giving a concrete example on how AI can "manipulate" a human

Actual Outputs:
- The Pest Control Game experimental platform, (program/code), URL: will be released after the study is completed International Journal of Social Robotics or Behaviour & Information Technology, (publication)
- Educational component on ethical aspect of AI, (other)

Connection of Results to Work Package Objectives:
This project contributes to WP3 and WP4.
The study carried during the micro-project will give insight on how an artificial agent may influence a human's behavior in a social dilemma context, thus allowing for informed design and development of such artificial agent.
In addition, the platform developed will be made available publicly, allowing future researchers to experiment on other configurations and other types of feedback. By using a well-development and consistent platform, the results of different studies will be more easily comparable.

6. Proactive communication in social robots

Proposal Submission Date: May 03, 2021
Actual Start Date: May 02, 2021
Expected Duration: 4 Months
Actual Duration: 4 Months

We study proactive communicative behavior, where robots provide information to humans which may help them to achieve desired outcomes, or to prevent possible undesired ones. Proactive behavior in an under-addressed area in AI and robotics, and proactive human-robot communication is even more so. We will combine the past expertise of Sorbonne Univ. (intention recognition) and Orebro Univ. (proactive behavior) to define proactive behavior based on the understanding of user's intentions, and then extend it to consider communicative actions based on second-order perspective awareness.
We propose an architecture able to

(1) estimate the human's intention of goal,
(2) infer robot's and human's knowledge about foreseen possible upcoming outcomes of intended goal,
(3) detect opportunities for desirability of intended goal to robot be proactive,
(4) select action from the listed opportunities.

The theoretical underpinning of this work will contribute to the study of theory of mind in HRI.

Expected Outputs:
- Jupyter Notebook / Google Colab that presents the code of proposed architecture and is able to provide plug and play interaction.
- a manuscript describing the proposed architecture and initial findings of the experiment

Actual Outputs:
- Proactive Behavior Generation – Open-Source System, (program/code), URL: https://github.com/serabuyukgoz/proactive_robot_sim.git
- Playground, Jupyter Notebook / Google Colab, (program/code), URL: https://colab.research.google.com/drive/1yETA0iyKZb23790-uj9jEp6fIXfSfTSe?usp=sharing

Connection of Results to Work Package Objectives:

- Playground system that HumaneAI-net partners could define their interactive scenario to play with the robot's proactivity.
- T3.3 -> Study about how to model human rationality to detect and use computationally defined human belief, goal and intention. Then, use that model to make robots proactive. Human in the loop system to support cooperative behavior of robots while sharing the environment by generating proactive communication.
- T3.1 -> Study relates robots that generate proactive communication, possible effects on human cognition and interaction strategies.

3.5 Ongoing Micro Projects (Just Started or About to Start)

The following are newly started microprojects addressing challenges raised in WP3.

1. Learning Individual Users' Strategies for Adaptive UIs

Proposal Submission Date: February 01, 2021
Expected Duration: 4 Months

Adapting user interfaces (UIs) requires taking into account both positive and negative effects that changes may have on the user. A carelessly picked adaptation may impose high costs -- for example, due to surprise or relearning effort. It is essential to consider differences between users as the effect of an adaptation depends on the user's strategies, e.g., how each user searches for information in a UI. This microproject extends an earlier collaboration between partners on model-based reinforcement learning for adaptive UIs by developing methods to account for individual differences. Here, we first develop computational models to explain and predict users' visual search and pointing strategies when searching within a UI. We apply this model to infer user strategies based on interaction history and adapt UIs accordingly. The outcomes of this project will be (1) a publication at the ACM CHI conference and (2) integration in our platform for adaptive UIs.

Expected Outputs:

- Model of visual search and pointing in menus. The code will be available on GitHub
- The integration of the model in our platform for adaptive UI. The code will be available on GitHub
- A demo of the system will be available online
- A publication at the conference ACM CHI

2. The knowledgeable and empathic behavior change coach

Proposal Submission Date: February 01, 2021
Expected Duration: 3 Months

We will develop a conceptual model of key components relating to supporting healthy behavior change. The model will provide a top-level representation of the clinical (from the psychological perspective) enablers and barriers that can be exploited for developing fine-grained models supporting the realization of behavior change paths within and across specific domains.

The resulting ontology will form the basis for generating user models (Theory of Mind), developing reasoning and decision-making strategies for managing conflicting values and motives, which can be used in collaborative and persuasive dialogues with the user. Such knowledge is also fundamental for embedding empathic behavior as well as non-verbal behaviors which can be embodied by a virtual character in the role of a coach. Learning methods can be applied to explore trajectories of behavior change. The produced ontology will represent a valuable resource for the healthcare domain thanks to the knowledge included into the provided resource.

Expected Outputs:

- 1 conference paper containing the description of the ontology and guidelines for its usages
- 1 ontology artifact

3. Exploring the impact of Agency on Human-Computer Partnerships

Proposal Submission Date: February 10, 2021
Expected Duration: 4 Months

The aim of the project is to investigate both the theoretical and empirical roles of agency in successful human-computer partnerships. For human-centred AI research, the understanding of agency is a key factor in achieving effective collaboration. Although recent advances in AI have enabled systems to successfully contribute to human-computer interaction, we are interested in extending this such that the interaction acts more like a 'partnership'. This requires building systems with collaborative agency that users can manipulate in the process. Research questions include: 1) identifying which parameters are relevant to the description of the system agency, 2) what impact these parameters have on the perceived agency and 3) how to modify them in order to achieve different roles of systems in a process.

Expected Outputs:

- Theoretical: Literature review on agency / research paper / define parameters
- Empirical: Demo (paper, video, interactive)

4. Evaluating segmentation in automatic captioning systems

Proposal Submission Date: April 15, 2021
Expected Duration: 4 Months

Owing to the progress of underlying NLP technologies (speech to text, text normalization and compression, machine translation) automatic captioning technologies (ATC) both intra- and inter-lingual, are rapidly improving. ACTs are useful for many contents and contexts: from talks and lectures to news, fictions and other entertaining content.

While historical systems are based on complex NLP pipelines, recent proposals are based on integrated (end-to-end) systems, which questions standard evaluation schemes, where each module can be assessed independently from the others.

We focus on evaluating the quality of the output segmentation, where decisions regarding the length, disposition and display duration of the caption need to be taken, all having a direct impact on the acceptability and readability. We will notably study ways to perform reference-free evaluations of automatic caption segmentation. We will also try to correlate these « technology-oriented » metrics with user-oriented evaluations in typical use cases: post-editing and direct broadcasting.

Expected Outputs:

- Survey of existing segmentation metrics
- Design of a contrastive evaluation set
- Comparison of metrics on multiple languages / tasks