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

D4.1: First Year Microproject Results on Societal, Ethical and Responsible AI

Author(s): Dino Pedreschi, Sahan Bulathwela, John Shawe-Taylor, George Kampis, Letizia Milli
Contributing partners: UNIPI, UCL, DFKI
Date: October 14, 2021
Approved by: Paul Lukowicz
Type: Report (R)
Status: Final
Contact: dino.pedreschi@unipi.it

Dissemination Level		
PU	Public	x

DISCLAIMER

DOCUMENT INFO

0.1 Authors

Authors	Institution	e-mail
Dino Pedreschi	UNIPI	dino.pedreschi@unipi.it
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
Letizia Milli (LM)	UNIPI	letizia.milli@di.unipi.it

0.2 Document History

Revision		
Date	**Lead Author(s)**	**Comments**
30.09.2020	GK	Empty template
10.09.2021	SB, JST, DP, LM	Initial and Second draft
13.09.2021	GK	Prettifying
22.09.2021	SB	Added Minor Modifications
14.10.2021	JST, DP, LM	Separated from D1.1

HUMANE AI NET

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 each of the main work packages (WPs) focused on societal, ethical and responsible AI. These sections 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 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.

[1] This and the next subsection are shared with Deliverable 1.1.

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 work programme 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 4: Societal AI

1.1 Overall Summary for Work Package 4:

The work package "Social AI" strives to shape the research on the societal dimension of AI, as increasingly complex socio-technical systems emerge, made by interacting people and intelligent agents. Examples range from urban mobility, with travellers helped by smart assistants to fulfill their agendas, to the public discourse and the markets, where diffusion of opinions as well as economic and financial decisions are shaped by personalized recommendation systems. In principle, AI could empower communities to face complex societal challenges. Or it can create further vulnerabilities and exacerbate problems, such as bias, inequalities, polarization, and depletion of social goods. Unfortunately, a crowd of (interacting) intelligent individuals is not necessarily an intelligent crowd. On the contrary, it can be stupid in many cases, due to network effects: the sum of many individually “optimal” choices is often not collectively beneficial, because individual choices interact and influence each other, on top of common shared resources. Navigation systems suggest directions that make sense from an individual perspective but may exacerbate congestion if too many drivers are directed on the same route. Personalized recommendations on social media often make sense to the user, but may artificially amplify polarization, echo-cambers, filter bubbles, and radicalization. Profiling and targeted advertising may further increase inequality and monopolies, with harms of perpetuating and amplifying biases, discriminations and “tragedies of the commons”.

How to understand and mitigate the harmful outcomes of social AI systems? How to design transparent mechanisms for decentralized collaboration that help social AI systems evolve towards agreed collective outcomes, such as sustainable mobility in cities, diversity and pluralism in the public debate, fair distribution of resources?

The set of micro-projects that have been proposed initially cover the wide spectrum of tasks envisaged in the work package.

The first of these is about "graybox models of society-scale, networked hybrid human-AI systems", aimed at developing modeling methodologies for social AI systems somehow halfway between data-driven “blackbox” and mathematical “whitebox” methods.

The second task, "individual vs. collective goals of social AI systems" is also an explicit theme in more than one MP, such as those addressing polarisation and misinformation in social media platforms or traffic in personal navigation assistants. Again, there is evidence that the "network" perspective can not only foster a deeper comprehension and prediction of the dynamics of social AI systems, but also spur the development of novel interaction mechanisms helping in reaching a better balance between collective and individual objectives, with reference to common goods.

The third task is about the "societal impact of AI systems" along its multiple dimensions. A first micro project in this line has tackled the social perception of AI through an extensive survey. It is, however, expected that more studies in this theme be further developed in the next stage of the project.

The fourth technical task of "self-organized, socially distributed information processing in AI-based techno-social systems" is aimed at understanding how to design distributed information processing in techno-social systems and what are the corresponding rules of delegating information processing to specific members (AI or human). Agent-based models and simulations play an important role in the micro-projects associated with this line.

A transversal line of activity is showcasing the power of combining AI/ML models with the network theory and complex system perspective, common to various micro-projects and well exemplified in the application to cultural heritage.
The final task is the consolidation and coordination of the research agenda which has been initiated by 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, as well as across different work packages.

1.2 Completed Micro Projects

The following microprojects address challenges raised by WP4.

1. Using Social Norms to counteract misinformation in online communities

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

The goal of the project is to investigate the role of social norms on misinformation in online communities. This knowledge can help identify new interventions in online communities that help prevent the spread of misinformation. To accomplish the task, the role of norms will be explored by analyzing Twitter data gathered through the Covid19 Infodemics Observatory, an online platform developed to study the relationship between the evolution of the COVID-19 epidemic and the information dynamics on social media. This study can inform a further set of microprojects addressing norms in AI systems through theoretical modelling and social simulations.

Expected Outputs:
Diagnosis and visualization map of existing social norms underlying fake news related to COVID19

Actual Outputs:
Functional pre-requisites of pluralistic ignorance in online settings., (publication)

Connection of Results to Work Package Objectives:
Top-down “debunking” interventions have been applied to limit the spread of fake news, but so far with limited power. Recognizing the role of social norms in the context of misinformation fight may offer a novel approach to solve such a challenge, shifting to bottom-up solutions that help people to correct misperceptions about how widely certain opinions are truly held. The results of this microproject can inform new strategies to improve the quality of debates in online communities and counteract polarization in online communities (WP4). These results can be also relevant for WP2 (T 2.4), e.g., by giving insights about how human interactions can influence and are influenced by AI technology, WP3 (T 3.3) by offering tools to study the reactions of humans within hybrid human-AI systems and WP5 (T 5.4) by evaluating the role of social norms dynamics for a responsible development of AI technology.

1.3 Ongoing Micro Projects (About 50% Complete)

The following are partially completed microprojects that address challenges raised by WP4.

1. Social AI gossiping

Proposal Submission Date: November 16, 2020
Actual Start Date: May 17, 2021
Expected Duration: 6 Months
Actual Duration: 6 Months

We envision a human-AI ecosystem in which AI-enabled devices act as proxies of humans and try to learn collectively a model in a decentralized way. Each device will learn a local model that needs to be combined with the models learned by the other nodes, in order to improve both the local and global knowledge. The challenge of doing so in a fully decentralized AI system entails understanding how to compose models coming from heterogeneous sources and, in case of potentially untrustworthy nodes, decide who can be trusted and why. In this micro-project, we focus on the specific scenario of model "gossiping" for accomplishing a decentralized learning task and we study what models emerge from the combination of local models, where combination takes into account the social relationships between the humans associated with the AI. We will use synthetic graphs to represent social relationships, and large-scale simulation for performance evaluation.

Expected Outputs:

- Paper (most likely at conference/workshop, possibly journal)
- Simulator (fallback plan if a paper cannot be produced at the end of the micro-project)

Actual Outputs:

- SAIsim, (program/code)

Connection of Results to Work Package Objectives:
The simulation engine is a modular one, that can be exploited (also by the other project partners) to test decentralised ML solutions. The weighted network used to connect nodes can represent social relationships between users, and thus one of the main objectives of the obtained results it to understand the social network effects on decentralised ML tasks.

Deviations from the Initial Plan:
The micro-project has started 6 months later than expected (17 May 2021 instead of 16 November 2020) and will last until the end of November 2021. The main reasons were some delays in hiring relevant people in the CEU and CNR groups to be allocated on the project activities, largely due to the COVID pandemic.

2. What idea of AI? Social and public perception of AI

Proposal Submission Date: December 01, 2020
Actual Start Date: February 01, 2021
Expected Duration: 4 Months
Actual Duration: 7 Months

This proposal wants to conduct empirical research that explores the social and public attitudes of individuals towards AI and robots.
AI and robots will enter many more aspects of our daily life than the average citizen is aware of while they are already organizing specific domains such as work, health, security, politics and manufacturing. Along with technological research it is fundamental to grasp and gauge the social implications of these processes and their acceptance into a wider audience.
Some of the research questions are:
Do citizens have a positive or negative attitude about the impact of Ai?
Will they really trust a driverless car, or will they passively accept a loan or insurance's denial based on an algorithmic decision? Do states alone have the right and expertise to regulate the emerging technology and digital infrastructures? What about technology governance?
What are the dominant AI's narratives in the general public?

Expected Outputs:

- Two scientific papers in journals like (depending on peer reviewing guidelines): -Ai & Society (https://www.springer.com/journal/146); -Journal of Artificial intelligence research https://www.jair.org/index.php/jair -Big Data and Society https://journals.sagepub.com/home/bds - also other potential journals will be considered (Social science computer review, Public understanding of science)
- Public presentation at scientific (e.g., HICSS 2022, AAAI) and more general interest conferences in the second half of 2021 and first half 2022.

Actual Outputs:

- A sociotechnical perspective for the future of AI: narratives, inequalities, and human control, in Ethics and Information technology., (publication)
- Minding the gap(s): public perceptions of AI and socio-technical imaginaries, (publication)

Connection of Results to Work Package Objectives:
The microproject addresses issues related to social trust, cohesion, and public perception. It has clarified how and to what degree Ai is accepted by the general public and highlighted the different level of public acceptance of AI across social groups.
Goals of T4.3 are met since the Sartori and Bocca article highlighted how perception and narratives differ by the main socio-demographics variables. Notable is the (expected) gender effect: women do have less knowledge and do trust less AI systems. Especially when it comes to depicting future and distopic scenarios, women tend to fear technologies more than men.
Goals of T5.3 are met by the work presented in Sartori and Theodorou. Focusessing on the main challenges associated with AI as autonomous systems spread within society, the article points to biases and unfariness as being among the major challenges to be addressed in a sociotechnical perspective.

Deviations from the Initial Plan:
Deviation resulted in the low number of questionnaires collected in Umea and CNR. We are confident to counterbalance the situation with a second reminder in September.

3. Agent based modeling of the Human-AI ecosystem

Proposal Submission Date: July 01, 2021
Actual Start Date: September 15, 2021
Expected Duration: 4 Months
Actual Duration: 9 Months

The project aims at investigating systems composed by a large number of agents belonging to either human or artificial type. The plan is to study, both from the static and the dynamical point of view, how such a two-populated system reacts to changes in the parameters especially in view of possible abrupt transitions. We are planning to pay special attention to higher order interactions like three body effects (H-H-H, H-H-AI, H-AI-AI and AI-AI-AI). We hypothesize that such interactions are crucial for the understanding of complex Human-AI systems. We will analyze the static properties both from the direct and inverse problem perspective. This study will pave the way for further investigation of the system in its dynamic evolution by means of correlations and temporal motifs.

Expected Outputs:

- 1 paper in a complex systems (or physics or math) journal

Actual Outputs:

- Simulation of delegation of information processing in techno-social groups., (program/code)

Connection of Results to Work Package Objectives:
The results of this project are central to
Task 4 - Self-organized, socially distributed information processing in AI-based techno-social systems.
This research will contribute to understanding how to optimize distributed information processing in techno-social systems and what are the corresponding rules of delegating information processing to specific members (AI or human).

4. Algorithmic bias and media effects

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

Recent polarisation of opinions in society has triggered a lot of research into the mechanisms involved. Personalised recommender systems embedded into social networks and online media have been hypothesized to contribute to polarisation, through a mechanism known as algorithmic bias. In a recent work [1] we have introduced a model of opinion dynamics with algorithmic bias, where interaction is more frequent between similar individuals, simulating the online social network environment. In this project we plan to enhance this model by adding the biased interaction with media, in an effort to understand whether this facilitates polarisation. Media interaction will be modelled as external fields that

affect the population of individuals. Furthermore, we will study whether moderate media can be effective in counteracting polarisation.
[1] Sîrbu, A., Pedreschi, D., Giannotti, F. and Kertész, J., 2019. Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. PloS one, 14(3), p.e0213246.

Expected Outputs:
- A paper on opinion dynamics in a complex systems or interdisciplinary journal.

Actual Outputs:
- TBD, (publication)

Connection of Results to Work Package Objectives:
The recent polarization of opinions in society has triggered a lot of research into the mechanisms involved. Personalized recommender systems embedded into social networks and online media have been hypothesized to contribute to polarisation, through a mechanism known as algorithmic bias.
In recent work we have introduced a model of opinion dynamics with algorithmic bias, where interaction is more frequent between similar individuals, simulating the online social network environment.
In this project, we plan to enhance this model by adding the biased interaction with media, in an effort to understand whether this facilitates polarisation. Media interaction will be modelled as external fields that affect the population of individuals. Furthermore, we will study whether moderate media can be effective in counteracting polarisation.

5. Pluralistic Recommendation in News

Proposal Submission Date: April 01, 2021
Actual Start Date: April 01, 2021
Expected Duration: 6 Months
Actual Duration: 8 Months

The micro-project aims at designing a Recommender System able to foster pluralistic viewpoints in news pieces suggestions. The first step consists of quantifying the political bias of a news article. While such an issue has been widely investigated in the USA domain, as far as we know, no work has been performed in the European domain. In this scenario, we have already built a dataset with more than 8 million European news articles labeled by their political leaning, popularity, and distribution area. Since a publicly available dataset of such size and richness of annotations does not exist in the EU media landscape, we think that it could have an enormous potential value for subsequent academic studies. Additionally, we are currently leveraging AI-based techniques for NLP to define a topic modeling algorithm and a multilingual classifier able to identify the main topics and the political leaning of each article.

Expected Outputs:
- A European-wide dataset of News with political bias (in completion) plus a data-paper describing it.

Connection of Results to Work Package Objectives:
The dataset we built consists of a very wide multilingual corpus, where most of the text comes from Europe. The value of a dataset such as this one in the context of Europe cannot be overstated. It contains text from every country of Europe, and its focus on political bias makes it a perfect fit for socially responsible AI.
We also believe that socially responsible AI should not be a privilege for English speaking individuals only. A multilingual dataset is both a better fit to better model the current European landmark, and a fairer solution to the "English" domination in NLP.

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

The following are newly started microprojects addressing the challenges raised by WP4.

1. Venice – creation of stories and narratives from data of cultural heritage

Proposal Submission Date: January 01, 2021
Actual Start Date: July 01, 2021**Expected Duration:** 4-6 Months

Creation of stories and narrative from data of Cultural Heritage

in this activity we want by tackle the fundamental dishomogeneity of the cultural heritage data is by structuring the knowledge available from user experience and methods of machine learning. The overall objective of this microproject is to design new methodologies to extract and produce new information, as well as to propose scholars and practitioners new and even unexpected and surprising connections and knowledge and make new sense of cultural heritage by connecting and creating sense and narratives with methods based on network theory and artificial intelligence

Expected Outputs:

- Publications about maps of Social Interactions across ages
- Publication about AI algorithm for the automatic classification of documents
- Database usable by the HumaneAI community as a pilot case

2. Network effects of mobility navigation systems

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

Study of emergent collective phenomena at metropolitan level in personal navigation assistance systems with different recommendation policies, with respect to different collective optimization criteria (fluidity of traffic, safety risks, environmental sustainability, urban segregation, response to emergencies, …).
Idea: (1) start from real big mobility data (massive datasets of GPS trajectories at metropolitan level from onboard black-boxes, recorded for insurance purposes), (2) identify major road blocks events (accidents, extraordinary events, …) in data, (3) simulate the effect (modify the data) that users involved in a road block were previously supported by navigation

systems the employ policies to mitigate the impact of the block, by using different policies different from individual optimization, aiming at collective optimization (aiming at diversity, randomization, safety, resilience, etc.)
Compare the impact of the different choices in term of aggregated impact.

Expected Outputs:

- (Big-) data-driven simulations with scenario assessment
- Scientific paper

2. Work Package 5: AI Ethics and Responsible AI

2.1 Overall Summary for Work Package 5:

WP5 is dedicated to ensuring that AI systems operate within an ethical, legal and social framework, in verifiable and justified ways.
Theory and methods are needed for the Responsible Design of AI Systems as well as to evaluate and measure the 'maturity' of systems in terms of compliance to law and to ethical and societal principles. These concerns need to be combined with robustness, social and interactivity design, and must be possible to be evaluated during the lifecycle of the system.
The micro-projects below contribute to these aims in different, but interconnected ways. The first MP, focus on the validation of ethical principles, namely fairness, and studies possible differences between post-processing versus in-processing approaches.
The second MP investigates the societal and ethical consequences of decision-making using chatbots to support the user, in the specific case of COVID19 vaccine information.
In the third MP, the HumaneAI-net project is linked to Europe's Green Deal and the EU Data Strategy. It aims at developing a socially relevant visualisation of air quality in large cities.
Several MPs are currently still running, and others planned to start soon

2.2 Completed Micro Projects

The following microprojects address challenges raised by WP5.

1. Validating fairness property in post-processing vs in-processing systems

Proposal Submission Date: November 23, 2020
Actual Start Date: December 01, 2020
Expected Duration: 4 Months
Actual Duration: 4 Months

After choosing a formal definition of fairness (we limit ourselves with definitions based on group fairness through equal resources or equal opportunities), one can attain fairness on the basis of this definition in two ways: directly incorporating the chosen definition into the algorithm through in-processing (as another constraint besides the usual error minimization; or using adversarial learning etc.) or introducing an additional layer to the pipeline through post-processing (considering the model as a black-box and focusing on its inputs and predictions to alter the decision boundary approximating the ideal fair outcomes, e.g. using a Glass-Box methodology).

We aim to compare both approaches, providing guidance on how best to incorporate fairness definitions into the design pipeline, focusing on the following research questions: Is there any qualitative difference between fairness acquired through in-processing and fairness attained by post-processing? What are the advantages of each method (e.g. performance, amenability to different fairness definitions)?

Expected Outputs:

- Paper: That addresses the difference between in-vs-post processing methods on ML models focusing on fairness vs performance trade-offs.

Actual Outputs:

- Bias mitigation: in-processing or post-processing? Ethical decisions hidden behind engineering choices, (publication)

Connection of Results to Work Package Objectives:
T6.7. Finance domain related industrial use case that has many benefits on ML based applications where fairness is important.
T5.4. Promotes the importance of ethics in design and leads to future methods and tools for the value-based design and development of AI systems.
T5.5. Compatible with our vision of responsible AI by design.

2. Ethical chatbots

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

Building AI machines capable of making decisions compliant with ethical principles is a challenge that needs to be faced in the direction of improving reliability and fairness in AI. This micro-project aims at combining argument mining and argumentation-based reasoning to ensure ethical behaviors in the context of chatbot systems. Argumentation is a powerful tool for modeling conversations and disputes. Argument mining is the automatic extraction of arguments from natural language inputs, which could be applied both in the analysis of user input and in the retrieval of suitable feedbacks to the user. We aim to augment classical argumentation frameworks with ethical and/or moral constraints and with natural language interaction capabilities, in order to guide the conversation between chatbots and humans in accordance with the ethical constraints

Expected Outputs:

- conference paper

Actual Outputs:

- An Argumentative Dialogue System for COVID-19 Vaccine Information, (publication), URL: https://arxiv.org/abs/2107.12079

Connection of Results to Work Package Objectives:
In the context of information-providing chatbots and assistive dialogue systems, especially in the public sector, ethics by design requires trustworthiness, transparency, explainability, correctness, and it requires architectural choices that take data access into account from the very beginning.

The main features of our chatbot architecture, with respect to the objectives of HumaneAI-net WP5, are

- an architecture for AI dialogue systems where user interaction is carried out in natural language, not only for providing information to the user, but also to answer user queries about the reasons leading to the system output (explainability).
- a transparent reasoning module, built on top of a computational argumentation framework with a rigorous, verifiable semantics (transparency, auditability).
- a modular architecture, which enables an important decoupling between the natural language interface, where user data is processed, and the reasoning module, where expert knowledge is used to generate outputs (privacy and data governance).

3. Improving air quality in large cities using mobile phone and IoT data

Proposal Submission Date: January 11, 2021
Actual Start Date: July 01, 2020
Expected Duration: 3 Months
Actual Duration: 5 Months

Globally, transportation is responsible for about 30% of air pollution, and in large cities, this is even higher. Between 20%-40% of deaths due to serious diseases are caused by air pollution (source: https://www.stateofglobalair.org/sites/default/files/documents/2020-10/soga-global-profile-factsheet.pdf). In Spain, 10.000 people die every year due to air pollution (almost tripling traffic deaths) and in Madrid alone, there are 5000 pollution deaths per year (14/day).

The combination of mobility data (generated from anonymized and aggregated mobile phone data of the telecommunications sector), IoT pollution & climate sensor data from moving vehicles, and Open Data, can provide actionable insights about traffic mobility patterns and pollution such that authorities and policymakers can better measure, predict and manage cities' mobility and pollution.

This micro project is strategically aligned with Europe's Green Deal and the EU Data Strategy.

Expected Outputs:

- Demonstration with visualizations for pollution in Spanish city
- Press release, blog post on organizations' websites
- Potentially scientific publication and patent (TBC)

Actual Outputs:

- Prototype, (program/code)

- Video, (other),
  URL: https://www.youtube.com/watch?v=WBNf5F9Kp7c
- Business and government presentations, (other)

Connection of Results to Work Package Objectives:
This project uses industrial data from the telecommunications industry, combined with open data and IOT generated data to solve an important societal problem, which are the two objectives of WP6. It shows a way in which the industry can create new products and services using artificial intelligence and data, very much aligned with the European data strategy. However, using data and AI for more evidence-based policymaking and decision-making by public institutions, also has ethical issues such as bias and undesired discrimination. In the prototype, we not only measure the quality of the air but also how many people are affected by this. In the series of three micro projects, we want to study whether this kind of data driven policymaking introduces undesired bias and inequality. We want to use the results of the other work packages to mitigate those potential problems.

2.3 Ongoing Micro Projects (About 50% Complete)

The following are partially completed microprojects that address challenges raised by WP5.

1. The role of designers regarding AI design: a case study

Proposal Submission Date: February 08, 2021
Actual Start Date: February 08, 2021
Expected Duration: 6 Months
Actual Duration: 4 Months

The purpose of this micro-project is to critically reflect on the design of an AI system by investigating the role of the designer. Designers make choices during the design of the system. Analysing these choices and their effective consequences contributes to an overall understanding of the situated knowledge embedded in a system. The reflection is concerned with questions like what does it mean for the output of the system what the designer's interpretations are? In what way do they then exercise power on this system? In particular, this micro-project will examine a concrete case. It will follow the design of an agent-based social simulation that aims at modelling how inequality affects democracy.

Expected Outputs:
- An agent-based simulation on how wealth inequality affects political relations
- A conference paper critically reflecting on the design of the simulation

Actual Outputs:
- The Role of a Designer, (publication), URL: TBA
- Inequality and Democracy, (other), URL: TBA

Connection of Results to Work Package Objectives:
WP5 is concerned with the responsible development of AI systems. This MP analyzes the role of a designer, and by doing so sheds light on what is necessary for responsible AI development. In order to develop responsible AI, it is essential to understand the underlying

power dynamics behind an AI's ecosystem. For this, we need to understand how and why a designer has a special role in the development of a system.

Deviations from the Initial Plan:
There is some delay (i.e., 2-3 weeks) in finalizing the output due to holidays.

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

The following are newly started microprojects addressing the challenges raised by WP5.

1. Ensuring Fairness when Protected Attribute Information Is Unavailable

Proposal Submission Date: Not Reported
Expected Duration: 6 Months

The basic challenge regarding debiasing ML models is that in order to prevent models from generating bias on the basis of some sensitive characteristics it is necessary to have information about these characteristics. Usually, this information is not available Fortunately, there is a new approach: Adversarial Reweighted Learning which debiases the models without having sensitive attribute information. However, this approach redefines the fairness as Rawlsian max-min principle which is quite different from parity-based fairness definitions that have been hitherto used. The goal of this project is to scrutinize the implications of using Rawlsian fairness principle in order debias the models by scrutinizing three things

1. Are there sensitive attributes for which Rawlsian fairness is unsuitable?
2. What would be the parity-based fairness scores when Rawlsian fairness definition is used for debiasing the models?
3. Is it possible to use Rawlsian fairness (and ARL) as post-processing method to existing models?

Expected Outputs:

- Publication on identifying the effect of Rawlsian fairness on parity-based fairness definitions
- Publication on how ARL can be used for models that are already being used as a post-processing tool

3. Work Package 6: Applied research with industrial and societal use cases

3.1 Overall Summary for Work Package 6:

There are only two MPs in this workpackage, one about 50% complete and the other just starting. The workpackage is focused on translating core research into applications, so this is in line with the development of the overall work programme. The first MP is concerned with connected vehicles and aims to develop a simulation environment for AI-based applications. The second is concerned with a dietary coaching application and provides an exciting potential demonstration for human-centric approaches to AI. We aim to promote the development of further MPs in this workpackage during the second phase of the network with appropriate calls for proposals potentially involving additional external (application oriented) partners.

3.2 Ongoing Micro Projects (About 50% Complete)

The following are partially completed microprojects that address challenges raised by WP6.

1. Connected vehicle simulation for AI-based applications

Proposal Submission Date: October 01, 2021
Actual Start Date: October 04, 2021
Expected Duration: 4 Months
Actual Duration: 5 Months

Build a simulation environment to test connected car data-based applications.
Application: For instance, parking space occupancy prediction. Simulate the number of observations necessary to produce reliable parking occupancy predictions and therefore estimate the necessary number of connected cars for such a service.

Expected Outputs:

- **Paper**
- Simulation Environment

Actual Outputs:

- Publication of results of simulation and AI application, (publication)

Connection of Results to Work Package Objectives:
AI based car data applications save people's time by guiding drivers and vehicles intelligently.
This leads in Addition to a reduction of the environmental footprint of the transportation sector by reducing local and global emissions.
The development and usage of a simulation environment enables data privacy compliancy for the development of AI based applications.

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

The following are newly started microprojects addressing the challenges raised by WP6.

1. ADoN - Automated Dietary Monitoring for Nutrition Coaching

Proposal Submission Date: January 18, 2021
Expected Duration: 3 Months

This micro-project aims to explore the integration of virtual coaching and the use of Automated Dietary Monitoring wearable devices by conducting a preliminary study in which these technologies are integrated and experimented in a real-life context. Such a capability will enable the investigation about the role of real-time sensing data within a healthcare monitoring system supporting users by suggesting the most appropriate healthy behavior. The designed strategy will be validated within a living lab involving a group of 20-30 users that will wear the textile and will use the application for a period of three weeks.
Goals are to validate the capability of the smart textile in detecting chewing activities in a real-world environment, the effectiveness of the coaching system in detecting unhealthy dietary behaviors, the quality of the feedback provided, and the overall acceptability of the users with respect to a coaching system introducing a minimum invasive strategy.

Expected Outputs:

- 1 conference paper containing the description of the proposed approach together with the results and the insights gathered from the living lag we run.
- 1 dataset containing all the generated data that will be made available to the HumanE-AI network

2. Can we use ML tasks as a proof of work in a consensus algorithm?

Proposal Submission Date: Not Reported
Expected Duration: 6 Months

Blockchains such as Bitcoin and Ethereum currently are computational wasteful. On an annual basis both blockchains consume over a 70 terawatt-hours (TWh) of energy on computational resources to secure the network, which is similar to the annual energy consumption of Switzerland. These computational resources are used to reverse a cryptographic hash function (which is called a consensus algorithm) that is a solution to a puzzle but serves no other purpose. Such an amount of computational resources should be used more efficiently. Our aim is to use the large amount of computational resources more efficient by replacing the cryptographic hash function with a machine learning task. We focus on the Ethereum network as its computational power is not limited to solving hash functions only, as is mostly the case in Bitcoin. However, our intended solution can be generalized to any blockchain that is currently a computational wasteful.

Expected Outputs:

- Publication: Review paper

4. Work Package 7: Innovation Ecosystem and Socio-Economic Impact

4.1 Overall Summary for Work Package 7:

Again, the workpackage has only seen two MPs, one completed and one just starting. The themes of the two MPs are very similar and related to matching expertise and people for innovation. The completed MP has done the development and research necessary for launching the follow-on MP that is just starting. It aims to develop a platform for matching users with appropriate expertise for different projects. As such this will be an exciting demonstration of the potential for AI to deliver positive contributions to both commercial and social good projects, highlighting how AI can enhance an innovation ecosystem with beneficial socio-economic impact.

4.2 Completed Micro Projects

The following microprojects address challenges raised in WP7.

1. Asking the right Questions! How to Match Expertise and People for Innovation.

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

A central challenge is to bring the people together to create innovations. Traditionally this works only with a high density of experts, entrepreneurs, and customers, e.g., in the Silicon Valley. In distributed settings on a European scale this does not work, despite the existence of matching platforms.
We take a different approach that is inspired by matching people in online dating. Individuals and companies often don't know what they can offer or what they need. Hence, we suggest a more holistic approach. Based questions asking about skills, interest, values, approaches, existing collaborations and digital artifacts (code, images, algorithms) we envision an intelligent matching platform.
We will run a workshop to understand the right questions, what artifacts are telling, and what good indicators for potential collaborations are. We also want to run a workshop to identify AI techniques for making the matches and for identifying the system architecture of the platform.

Expected Outputs:

- Catalog of questions as the basis for the matching platform
- A system architecture for the platform
- Set of candidate algorithms for matching

Actual Outputs:

- Event documentation of the AI Idea Price, (other)
- Video recording of qualitative interviews, (other)
- Questionnaire of the matching platform, (other)

Connection of Results to Work Package Objectives:

The elaborated questionnaire and the developed concept of the matching platform both pursue the aim of bringing the different players in the AI ecosystem together, which are normally not in contact. The overarching goal is to strengthen the EU AI Community in order to increase the knowledge transfer between startups, academia and industry and therefore enable organizations that are compatible with European and Humane values.

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

The following are newly started microprojects addressing the challenges raised by WP7.

1. Matching the right people! Creating a functional prototype for online matching of people and expertise for innovation.

Proposal Submission Date: January 07, 2021
Expected Duration: 6 Months

In the first microproject “Asking the right questions” we could identify and verify a set of functional questions to match people for innovation. With this project we build on this understanding of relevant elements of a matching platform and the identified user needs in the context of AI based innovation.

Within this micro project, we aim to implement an AI matching platform prototype. The first version will allow users to specify and manage profile data and receive matches. Matches could be either one-to-one matches or where a set of people is identified for which live event (online or in person) for getting them together is beneficial. We aim to evaluate the platform and get feedback from technology and domain experts, scientists, entrepreneurs, and startups.

Our long-term goal is to help connect experts and entrepreneurs across physical boundaries, creating a vibrant and agile high-tech environment on a European Scale.

Expected Outputs:

- A first prototype of a working matching platform
- Matches and test users on the platform

5. Work Package 8: Virtual Center of Excellence, Capacity building and Dissemination

5.1 Overall Summary for Work Package 8:

This work package has seen only one project proposed which has only just started. It aims to showcase a system for recommending open educational resources (OER) by linking to different OER repositories and identifying semantically relevant lectures for individual learner or teacher needs. The system operates cross-lingually and to some extent cross-culturally, hence enabling a virtual center of excellence, capacity building and dissemination. The MP aims to build a collaboration with AI4EU and important thread in the overall project work programme.

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

The following is a newly started microproject addressing the challenges raised by WP8.

1. X5LEARN: Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site interface for access to openly licensed educational materials

Proposal Submission Date: December 14, 2020
Expected Duration: 3 Months

K4A proposes a micro project to extend its existing prototype of the online learning platform X5LEARN (https://x5learn.org/) developed by K4A and UCL and JSI and its new IRCAI center under the auspices of UNESCO. It is a standalone, learner-facing web application designed to give access through an innovative interface to a portfolio of openly licensed educational resources (OER) in video and textual format. Designed for lifelong learners looking for specific content wanting to expand on their knowledge, our aim is to extend it to AI-related topics. The updated application will be released via IRCAI a newly designated AI center and integrated with AI4EU with heavy HumaneAI branding. The main reason to push the product with IRCAI is that UNESCO is positioning itself as the main UN agency to promote humanist Artificial Intelligence, a major international policy on the Ethics of AI, and champion OER, which is in line with HumaneAI.

Expected Outputs:

- System prototype version 1: The initial prototype of the X5LEARN system upgraded with a novel User Interface and initial TrueLearn models
- System prototype version 2: The second prototype of the X5LEARN system that integrates into the AI4EU Platform
- Final prototype version 2: The final X5LEARN prototype that integrates the final and updated TrueLearn models in both interfaces
- X5LEARN System Expandability/Updateability database with additional OER materials
- X5LEARN Public website branded with UNESCO and HumaneAI visuals
- X5LEARN Integration Guidelines and API Reference Guide
- Paper to be published in peer-reviewed journal