HUMANE AI NET

Proposal: HumanE AI Net; **Duration of the project:** 36 months; **Topic:** ICT-48-2020

No.	Participant organization name	Country
1	DEUTSCHES FORSCHUNGSZENTRUM FÜR KÜNSTLICHE INTELLIGENZ GMBH (Coordinator)	DE
2	AALTO KORKEAKOULUSAATIO SR	FI
3	AIRBUS DEFENCE AND SPACE SAS	FR
4	Algebraic AI S.L.	ES
5	ATHENA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS	EL
6	VYSOKE UCENI TECHNICKE V BRNE	CZ
7	BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION	ES
8	Közép-Európai Egyetem (CEU)	HU
9	CONSIGLIO NAZIONALE DELLE RICERCHE	IT
10	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS	FR
11	AGENCIA ESTATAL CONSEJO SUPERIOR DEINVESTIGACIONES CIENTIFICAS	ES
12	UNIVERZITA KARLOVA	CZ
13	CONSORZIO INTERNUVIERSITARIO NAZIONALE INFORMATICA	IT
14	EOTVOS LORAND TUDOMANYEGYETEM	HU
15	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	CH
16	FONDAZIONE BRUNO KESSLER	IT
17	FORTISS GMBH	DE
18	FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.	DE
19	Generali Italia S.p.A.	IT
20	GERMAN ENTREPRENEURSHIP GMBH	DE
21	INESC TEC - INSTITUTO DE ENGENHARIADE SISTEMAS E COMPUTADORES, TECNOLOGIA E CIENCIA	PT
22	ING GROEP NV	NL
23	INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE	FR
24	INSTITUTO SUPERIOR TECNICO	PT
25	INSTITUT JOZEF STEFAN	SI
26	KNOWLEDGE 4 ALL FOUNDATION LBG	UK
27	LUDWIG-MAXIMILIANS-UNIVERSITAET MUENCHEN	DE
28	OREBRO UNIVERSITY	SE
29	PHILIPS ELECTRONICS NEDERLAND B.V.	NL
30	SAP SE	DE
31	SORBONNE UNIVERSITE	FR
32	STICHTING VU	NL
33	THALES SIX GTS FRANCE SAS	FR
34	TELEFONICA INVESTIGACION Y DESARROLLO SA	ES
35	TILDE SIA	LV
36	TECHNISCHE UNIVERSITAT BERLIN	DE
37	TURKIYE BILIMSEL VE TEKNOLOJIK ARASTIRMA KURUMU	TR
38	TECHNISCHE UNIVERSITEIT DELFT	NL
39	TECHNISCHE UNIVERSITAET KAISERSLAUTERN	DE
40	TECHNISCHE UNIVERSITAET WIEN	AT
41	UNIVERSITY COLLEGE CORK - NATIONAL UNIVERSITY OF IRELAND, CORK	IE
42	KOBENHAVNS UNIVERSITET	DK
43	UNIVERSITE GRENOBLE ALPES	FR
44	UNIVERSITEIT LEIDEN	NL
45	UMEÅ UNIVERSITET	SE
46	ALMA MATER STUDIORUM - UNIVERSITA DI BOLOGNA	IT
47	UNIVERSITA DI PISA	IT
48	UNIVERSITY COLLEGE LONDON	UK
49	UNIWERSYTET WARSZAWSKI	PL
50	THE UNIVERSITY OF SUSSEX	UK
51	UNIVERSIDAD POMPEU FABRA	ES
52	VRIJE UNIVERSTEIT BRUSSEL	BE
53	VOLKSWAGEN AG	DE

1. Excellence

Over the course of the last decade, artificial intelligence (AI) researchers have made groundbreaking progress in hard and longstanding problems related to machine learning, computer vision, speech recognition, and autonomous systems. In combination with continuing advances in related technologies such as the Internet of Things (IoT), mobile computing and mechatronics, AI is quickly becoming an integral part of nearly all areas of our daily lives, from smartphones and smart watches to personal digital assistants such as Amazon Echo and Google Home to autonomous vehicles, smart cities, Industry 4.0, and beyond. By packaging AI functionality in cloud services and libraries, the hurdle for using AI technologies has been lowered, pushing forward a wealth of applications in many domains.

There is a strong consensus that AI will beget changes **far more profound than any other technological revolution in human history**. Depending on the course that this revolution takes, AI will either empower our ability to make more informed choices or reduce human autonomy; expand the human experience or replace it; create new forms of human activity or make existing jobs redundant; help distribute well-being for many or increase the concentration of power and wealth in the hands of a few; expand or endanger democracy in our societies. Europe carries the responsibility of shaping the AI revolution. The choices we face today are related to fundamental ethical issues about the impact of AI on society—in particular, how it affects labor, social interactions, healthcare, privacy and fairness.

The HumanE AI Network (hereafter referred to as HumanE AI Net or HumanE AI) will leverage the synergies between the involved centers of excellence to develop the **scientific foundations and technological breakthroughs** needed to shape the AI revolution in a direction that is **beneficial to humans both individually and societally,** and that adheres to **European ethical values and social, cultural, legal, and political norms**. The core challenge is the development of robust, **trustworthy** AI systems capable of what could be described as "**understanding" humans**, adapting to complex real-world environments, and appropriately **interacting in complex social settings**. The aim is to facilitate AI systems that **enhance human capabilities and empower individuals and society as a whole** while respecting human **autonomy and self-determination**. The HumanE AI Net project will engender the mobilization of a research landscape far beyond direct project funding, involve and engage European industry, reach out to relevant social stakeholders, and create a unique innovation ecosystem that provides a manyfold return on investment for the European economy and society.

Facilitating this vision requires new solutions to fundamental scientific questions—not just within narrow classical AI silos, but **at the interstice** of various AI areas such as learning, reasoning, and perception on one side, and particularly for other disciplines, human-computer interaction (HCI), cognitive science, and the social sciences. The following are specific research areas where substantial gaps exist today that will be addressed in the course of the project:

1. *Human-in-the-loop machine learning, reasoning, and planning.* Allowing humans to not just understand and follow the learning, reasoning, and planning process of AI systems (being explainable and accountable), but also to seamlessly interact with it, guide it, and enrich it with uniquely human capabilities, knowledge about the world, and the specific user's personal perspective.
2. *Multimodal perception and modeling.* Enabling AI systems to perceive and interpret complex real-world environments, human actions, and interactions situated in such environments and the related emotions, motivations, and social structures. This requires enabling AI systems to build up and maintain comprehensive models that, in their scope and level of sophistication, should strive for more human-like world understanding and include common sense knowledge that captures causality and is grounded in physical reality.
3. *Human-AI collaboration and interaction.* Developing paradigms that allow humans and complex AI systems (including robotic systems and AI-enhanced environments) to interact and collaborate in a way that facilitates synergistic co-working, co-creation and enhancing each other's capabilities. This includes the ability of AI systems to be capable of computational self-awareness (**reflexivity**) as to functionality and performance, in relation to relevant expectations and needs of their human partners, including transparent, robust adaptation to dynamic open-ended environments and situations. Overall, AI systems must above all become **trustworthy partners** for human users.
4. *Societal awareness.* Being able to model and understand the consequences of complex network effects in large-scale mixed communities of humans and AI systems interacting over various temporal and spatial scales. This includes the ability to balance requirements related to individual users and the common good and societal concerns.

5. *Legal and ethical bases for responsible AI.* Ensuring that the design and use of AI is aligned with ethical principles and human values, taking into account cultural and societal context, while enabling human users to act ethically and respecting their autonomy and self-determination. This also implies that AI systems must be "under the Rule of Law": their research design, operations and output should be contestable by those affected by their decisions, and a liability for those who put them on the market.

1.1. Objectives

1. Leverage synergies between the involved centers of excellence to **develop the scientific foundations and technological breakthroughs** with respect to the gaps that need to be closed to facilitate our vision of Human Centric AI as outlined above (see section 1.3.1 for specific scientific concepts and goals).

   Verification criteria and KPIs: Publishing >= 100 articles in refereed venues (see examples in section 2), building up 10 benchmark sets, running 5 challenges with the community..

2. **Develop strong links between the HumanE AI Net centers of excellence within the consortium** that will contribute to a sustainable European Human Centric AI community. To this end we will be conducting the research through series of **collaborative** "micro-projects" involving researchers from different groups from within and in some cases from outside the consortium as described in 1.3.2.2 working together at one site for extended periods of time.

   Verification criteria and KPIs: Number of publications with authors from several centers of excellence (>50), number of PMs worked by project members at a location that is not their own (sum over consortium >300)

3. Closely **synchronize** the HumanE AI Net work on scientific and technological breakthroughs with **industry and society needs to foster synergies** and strengthen Europe's position in the global marketplace.

   Verification criteria and KPIs: annually updated **applied research agenda** for each of the domains in WP 6 produced by the respective industrial champion together with representatives of relevant research partners within stakeholder workshops, 6 industrial use cases per year successfully implemented within micro-projects where researchers from research partners will work under the leadership of the respective industrial champion.

What will HumanE AI do?

It will be able to have a rich and reflective discussion with a human. To understand the practical implications, consider a judge, doctor, policymaker, or manager facing a complex decision based on a large, noisy dataset comprising multiple aspects, not all of which are the decision maker's core expertise. Because such decisions often have grave personal and/or social consequences (and they typically include complex ethical and emotional facets as well), a complete replacement of human decision makers by AI tends to be undesirable, even if it were feasible. Existing decision support systems are mostly about guiding a person through a predefined decision tree, which means that while the human may formally make the decision, it is often largely predetermined by the system. Data mining and analytics systems leave much more freedom to the user, but at the price of potential information overload. Instead of these, we envision a system that can **genuinely, constructively discuss problems** with human users. Beyond being able to merely rebut individual arguments (like the IBM Debater program), the system should develop a differentiated understanding of human lines of reasoning, relate to human motivations, emotions, moral assumptions, and implications in this reasoning; help human partners challenge their own assumptions as well as provide simulations with consequences; and explain alternate "AI angles" on seeing the problem.

4. **Facilitate cross fertilization and knowledge transfer** between the HumanE AI Net centers of excellence and Industry through "human resources" by making active use of AI-on-demand platform infrastructures, supporting internships (both ways students/academic personnel in industry and industrial R&D personnel at academic labs) and running an integrated pan-european Human Centric AI Ph.D. and postdoc program.

   Verification criteria and KPIs: Number of industrial Ph.Ds associated by the program >20, number of internship months >20, existence of the brokerage mechanism.

5. **Contribute to broad take up of our Human Centric AI technology** by **European SMEs**

   Verification criteria and KPIs: Number of contacts with SMEs such as visits, talks given, personnel from SMEs participating in micro-projects **(>50)**

6. **Contribute to innovation and creation of new startups in the domain of Human Centric AI** through the establishment of an incubator infrastructure and startup support measures.

Verification criteria and KPIs: Presence of an innovation ecosystem with measures such as match making, education of researchers with respect to innovation strategies etc. in place. We expect actual start-ups to emerge from the project.

7. **Contribute to the AI4EU AI-on-demand platform** (and any additional future platform efforts within the ICT 49 call, such as the European Language Grid platform for speech and NLP resources, tools and services). Specifically we aim to ensure and demonstrate broad takeup of the platform within the community, contribute new methods, models, tools and data sets to the platform and help make it more usable and accessible.

Verification criteria and KPIs: presence of the infrastructure for scientific collaboration and challenges within the platform, presence of project results in the platform, usage statstics of the platform from the consortium

8. **Interface and collaborate with relevant related national and European initiatives** including in particular Digital Innovation Hubs, SoBigData, META-NET, CLARIN ERIC and its technical centres, EIT, WASP and WASP-HS programs…,

Verification criteria and KPIs: Number of talks given at relevant events, presence in the respective news letters, joint events.

9. **Contribute to public and political debate on AI** and its consequences through events directed at the general public, the political decision makers and.

Verification criteria and KPIs: Number of public events (1 per year), publishing project brochure, social media statistic

10. Set up a **Virtual Laboratory** as an AI one-stop shop for researchers and practitioners both inside and outside of the consortium to disseminate the latest knowledge (using online courses, overview papers and pointers to relevant web resources such as from the AI4EU and other platforms), and to lower the threshold for students and additional researcher to profit from and advance the state of the art.

Verification criteria and KPIs: Presence of the virtual laboratory, accessibility of all project results through the Laboratory, laboratory usage statistics

11. **Actively cultivate a outreach and knowledge dissemination** throughout Europe's entire AI community

Verification criteria and KPIs: Annual summer school on human centric AI, at least one other summer school each year on a focused topic, a total of 5 tutorials and 4 workshops during the course of the project, access statistics to MOOCs (>1000)

What will HumanE AI do?

Empower an employee to start a new career. This idea builds on the current notion of intelligent assistants that detect user actions, and situations in the environment to offer warnings, suggestions, and supporting information to the user. Such systems help people do their jobs in a more efficient, less error-prone way. However, the HumanE AI notion of augmenting human capabilities goes far beyond merely assisting people to do their jobs better. Instead, the vision is to enable a person to perform activities that he or she would **otherwise not be capable of doing at all**. As an analogy, consider the comparison between a simple excavator and an exoskeleton. People can easily dig holes without an excavator, but they can dig faster and deeper if they happen to have one. An exoskeleton, on the other hand, can enable a paralyzed person to walk, something that he or she would otherwise obviously not be able to do at all. Furthermore, while an excavator is a tool that a person must operate consciously and explicitly, an exoskeleton amplifies human actions synergistically, implicitly supporting the user. Currently, various mobile and wearable systems are being deployed as part of Industry 4.0, including support for assembly line work. They enable workers to do their jobs faster and better, but they do not fundamentally change the nature of what these workers actually do. In HumanE AI, we will pursue the vision of a *cognitive amplifier* to enhance a person's cognitive abilities to a level where someone who lost his or her previous job could quickly take up a **more qualified** new career. Thus, a former assembly line worker could become a technical support person relying on AI to overcome the limitation that would otherwise prevent her/him from doing this job.

1.2. Relation to the work program

1.2.1. Technology Focus

As indicated on page 1 of this proposal, the core scientific challenge is "the development of robust, trustworthy AI systems capable of what could be described as "understanding" humans, adapting to complex real-world environments and appropriately interacting in complex social settings. Of the five main research areas (and the corresponding WPs) one is dedicated to multimodal perception and modeling (WP 2) and one to human-AI interaction and collaboration (WP 3). Two other research areas (and respective work packages)—societal AI (WP 4, devoted to the impact of AI on society and ensuring social benefits of AI) and ethics and responsible AI (WP 5)—are situated within the core of the European Vision of Human-Centric AI (as outlined, for example, in the High-Level Expert Group Guidelines for Trustworthy AI). Thus the proposal is clearly situated within the scope of the "Advanced perception or interaction with humans (for human-centered AI) and environments" focus of the call. The fifth research area on learning, reasoning, and planning with humans[1] in the loop has a strong connection to the "Advances in Foundations of AI" focus, but from a strong perception and interaction angle. Within perception and interaction, we also will consider some robotics-related angles (in particular, through the participation of Prof. Ana Paiva from Lisbon, Profs. Raja Chatila and Mohamed Chetouani from Sorbonne, and Prof. Frank Kirchner from DFKI). Finally Tasks 6.2 and 6.3 are devoted to hardware platforms, in particular low-power implementations creating a link to "AI at the edge and hardware for AI" focus.

What will HumanE AI do?

Support and manage common resources in a sustainable way/region. Good management of common resources is a key aspect of well-being for human groups, societies, and humanity. Yet humans are notoriously bad at this task. In the well-described tragedy of commons [Harding 1968], individuals acting in their individual self-interest usually deplete or damage the common resource whether the task involves sharing limited amounts of water, or joint use of common pastures and fisheries. On a global level, climate crisis, depleting natural resources, or environmental damage are perhaps the best-known examples. Non-material goods, such as group reputation or intellectual property, can also be conceptualized as a common resource. There are well-established principles that allow groups to constructively share common resources [Ostrom1990; Baland1996]. In techno-social groups, AI agents also use and contribute to common resources. AI agents should be aware not only of their goals, but also of how their actions affect common resources. They should follow good practices of common resource management by knowing principles that enable groups to manage common resources and by acting according to them. They also should encourage other users to follow these principles.

1.2.2. Network Composition

"Each network should be driven by leading figures in AI from major excellent research centers, bringing the best scientists distributed all over Europe. They will bring on board the necessary level of expertise and variety of disciplines and profiles to achieve their objectives"

The HumanE AI net project amasses key European AI research centers (e.g., DFKI[2], Fraunhofer in Germany; INRIA, CNRS from France; CNR CINI, FBK from Italy; ATHENA from Greece; FCAI Aalto in Finland; INESC in Portugal; and AI Research Institute (IIIA-CSIC) in Spain) and top European Universities (ETH Zurich, Sorbonne, LMU Munich, TU Berlin, TU Vienna, UCL London, and TU Delft). It combines competences from machine learning (Prof.[3] John Shaw Taylor from UCL, Prof. Klaus Müller TU Berlin, and Prof. Samuel Kaski AAlto), reasoning and symbolic AI (Prof. Frank van Harmelen, Amsterdam, Prof. Paolo Traverso FBK, Thomas Eiter TU Wien, and Prof. Tomasz Michalak UW), multimodal perception and modeling (Prof. James Crowley, Inria, Prof. Paul Lukowicz DFKI), NLP (Prof. François Yvon, LIMSI/CNRS, Prof. Jan Hajič, Charles University, Dr. Bernardo Magnini, FBK, and Prof. Jan Černocký, Brno Univ. of Technology), HCI (Prof. Albrecht Schmidt LMU, Prof. Yvonne Rogers UCL, Wendy McKay INRIA, and Prof. Antti Oulasvirta), computational social science (Prof. Andrzej Nowak, UW, Prof. Frank Dignum, UMU, and Prof. Ana Paiva, IST), AI explainability, ethics, and design for values (Prof. Fosca Giannotti CNR, Prof. Dino Pedreschi UNIPI, Prof. Virginia Dignum, UMU, Prof. Jeroen van den Hoven, TUD, Dr. Nardine Osman, and Prof. Carles Sierra IIIA-CSIC) and others.

Industrial participation is ensured through industrial research teams and also in bringing expertise to identify important technological limitations hampering deployment in an industrial context.

The consortium includes European "Industrial Champions" from key sectors of the industry. such as Volkswagen (automotive/mobility), Airbus Group (Aerospace), Generali (Insurance), ING (FinTec), Philips (Health), Thales (Industry 4.0), Telefonica (Telco provider), Tilde (civil service with a focus on NLP), and SAP (with key contributions to security issues), and Volkswagen (automotive and mobility) who will drive the industrial use cases and provide industrial grounding for the research agenda.

Each network must demonstrate access to the required resources and infrastructure to support research and design (R&D), such as data, high-performance computing/HPC (central, GPUs, edge computing), storage, robotics equipment, and Internet of Things (IoT) infrastructure, support staff, and engineers to develop experiments.

The consortium includes the Barcelona Supercomputing Center, who not only has one of the most advanced computer infrastructures in Europe but is also a member of the PRACE (Partnership for Advanced Computing Europe) and, where needed, will help HUmanE AI Net partners access PRACE resources. Because HumanE AI partners are leading research centers and universities in Europe, most also have access to state-of-the-art local supercomputing infrastructure. Furthermore, the coordinator (DFKI) is a partner in the NVIDIA NVAIL program and hosts one of the largest GPU systems in Germany.

Network Objectives

"... networks will develop mechanisms to spread the latest and most advanced knowledge to all the AI-labs in Europe and prepare the next generation of talent in AI."

The strategy for spreading knowledge is described in section 1.3.2.7 and in the Dissemination section 2. In summary, it involves (in addition to high quality scientific publications) a series of summer schools, tutorials and seminars, a planned publication of a Handbook of Human Centric AI, and a series of online courses (MOOCs). Furthermore, the project will also organize challenges for the community and provide benchmark datasets.

What will HumanE AI do?

Enable a robot to be elected "teammate of the month" by her fellow humans. While today's robots are routinely deployed in hazardous environments, they are typically used for simple tasks with little autonomy. Thus, a reconnaissance robot may be sent to provide images from a danger zone, typically under heavy remote human supervision (if not outright remote control). In contrast, we envision systems that can become full-fledged members of the intervention team. On the one hand, this implies the ability to act with a high degree of autonomy. As a team member, a robot may be instructed to "go and take care of the area on the right," which indicates that it must autonomously explore an unknown, dynamic, unstructured environment, take any action needed to control the fire, and then help victims (this includes dealing with ethical issues involved in potentially prioritizing what/who to attend to first). On the other hand, it also implies the ability to fit and act within the team's social structure and group dynamics. This includes "reading" the subtle signals that indicate emotions, strain, and tension on both individual and collective levels and then reacting in a way best-suited to support positive group dynamics and empower the group to reach optimal performance with minimal friction. Such systems are by no means limited to emergency response forces such as firefighters; they can be applied in an analogous way to, for example, construction or medical teams, ship crews, or even teams of scientists or technicians.

"... develop synergies and cross-fertilization between industry and these networks of excellence centers, in particular through internships of academic staff (at all levels) in industry, or PhD programmes with industry."

Proposals will include common academic/industrial PhD programmes and post-PhD programmes with a focus on industrial challenges.

Task 8.4 is devoted to establishing an industrial PhD, postdoc, and internship program. The concept is described in section 1.3.2.8. Note that the notion of a collaborative microproject in which industry from both within and outside the consortium can participate is also an important internship and personnel-exchange instrument. When it comes to deciding about assigning funds to external participants (see section 1.3.2.3) reaching participants from industry (including SMI) will be an important constraint. Here, connections will be made through WP 7 and in particular T7.2 (Platform for Matching People, Ideas, Research, and Resources) and 7.7 (AI Innovation Networking Events) will be crucial.

... networks[4] *will form a common resource and will become shared facility, as a virtual laboratory offering access to knowledge and expertise and attracting talent. It should become a reference, creating an easy entry point to AI excellence in Europe and should also be instrumental for its visibility.*

Tasks 8.1 and 8.2 and are devoted to the implementation and operation of a virtual laboratory. The Virtual Laboratory will be closely integrated with the AI4EU platform, both in terms of being a resource on the platform and in terms of using platform resources. The Virtual Laboratory together with the AI4EU platform will provide easy access to all project generated content and allow easy contact with consortium members. It will also include the innovations resources such as the brokerage platform (T7.2).

1.2.3. Network Activities

"… the proposals will focus on important scientific or technological challenges with industrial relevance and where Europe will make a difference, either in building on strengths, or strengthening knowledge to fill gaps critical for Europe."

The proposal addresses the question of how AI systems can develop complex "understandings" of humans and the situations in which they are acting and interacting (including a complex social context), and from that premise, **enhance human capabilities and empower humans** on individual and social levels. The work will be conducted under strict observance of European ethical and legal standards and respect for human autonomy and self-determination. This vision of AI is what is needed for AI's positive impact on European society and economy, while building up a specific European brand of AI. As explained in the proposal, this vision poses fundamental research questions in many areas of AI, but especially has gaps at the interfaces of various parts of AI, HCI, social science, psychology, and complexity science. The project focuses on collaborative work that connects top European researchers at such interfaces to create a new generation of uniquely European AI.

What will HumanE AI do?

It will expose people to diversity to favor informed opinions on controversial issues. Humans tend to search for information consistent with their opinions and beliefs, a mechanism known as *confirmation bias*. The tendency is exploited by online platforms for information search and social networking and media, which employ recommendation algorithms to catalyze users' attention. As a side effect, the platform amplifies and reinforces individual bias, resulting in extreme polarization of opinions and filter bubbles at the social level, with dramatically negative consequences on the pluralistic public debate needed to nurture democracy. In addition, often access to information is maliciously biased by either commercially or politically motivated influence agents. Human-centric AI has a clear social dimension that may help us design novel platforms and mechanisms for access to news and information, focused on counterbalancing our built-in confirmation bias and transparently striving to expose people to assorted opinions, intelligently. We imagine mechanisms for helping individuals and communities become informed on controversial issues by offering multiple perspectives, connecting opposing views and conflicting arguments, and fostering critical thought. For example, a robot in a group conversation can highlight information that was unavailable to the group, or suggest omitted sources of important information. Advances in person-machine interaction models based on explainable AI have the potential to reach novel cognitive trade-offs between our confirmation bias and our curiosity of novelty and diversity, making it possible for more sustainable and humanized information ecosystems to emerge.

Strong links will be developed among the members of the networks, notably through collaborative projects, exchange programmes, or other mechanisms to be defined by the consortium.

As described in section 1.3.2.2 the main mechanism for implementing the research agenda are collaborative microprojects. These involve researchers from **several** partners **jointy** working for a period of up to a **few months** at a **single location** (which will be the location of one of the involved partners, the host). This is a much stronger collaboration mechanism than the typical project that involves collaboration with each partner's researchers remaining at their own labs and only coming together for occasional meetings.

"… the networks will develop and implement common research agendas. The main vision and roadmap with targets within the projects, as well as methodology to implement and monitor progress will have to be specified in the proposal and can be further developed during the project."

The research agenda is described in detail in section 1.3, including both the overall vision and specific research questions within the five broad research areas. Within each of the areas, section 1.3.1 defines further concrete subareas with more specific research directions. Each of these areas is a task in a corresponding WP.

"progress will be demonstrated in the context of use-cases, also helping to foster industry-academia collaboration"

WP 6 is devoted to "applied research with industrial and societal use cases." It focuses on six domains, each led by an industrial champion, each with a designated task. Each of those tasks will run stakeholder workshops to align with the research agenda, with the needs of the specific industry and will conduct microprojects that are devoted mostly to applying WPs 1–5's results to industrial use cases.

"The proposals should define mechanisms to foster excellence, to increase efficiency of collaboration, and to develop a vibrant AI network in Europe.

What will HumanE AI do?

It will add resilience to financial markets. Automated trading systems that combine AI with a variety of abstract mathematical models are pervasive in financial markets. Although different market players have their own justification for using AIs, these AIs have also been identified as a key source of market fragility [Cespa2017]. This fragility includes, in particular, so-called *flash events*, such as the May 6, 2010 flash crash, in which key US stock indices collapsed and rebounded again within minutes (the Dow Jones lost 9% and then recovered most of that loss within an hour), temporarily erasing up to a trillion dollars in value. Clearly, within the hour in question, no real-world events would have justified the fluctuation of value in the companies that make up the involved indices. There is general agreement that complex, unpredictable interactions between "blackbox" systems making decisions that human participants could neither follow nor understand were a key factor in the flash crash [Kirilenko 2107]. Had experienced humans been trading in the place of the automated systems, and had they been given enough time between trades to consider the market situation (including looking for relevant political/economic news) to speak to each other and reflect on the overall situation, the event would very likely not have happened. The question is how can AI systems trading at ultra-high frequencies be endowed with the ability to reflect on their actions in the background of a complex political and macro-economic situation, assess human market participants' probable emotions and reactions, "discuss" with each other, and then interact with relevant humans while using "common sense" to prevent a crash?

Each network will disseminate the latest and most advanced knowledge to all the academic and industrial AI laboratories in Europe, and involving them in collaborative projects/exchange programmes" [5]

In addition to the measures already described (summer schools, tutorials, workshops, challenges, and MOOCs) that actively will be distributed and supported within the community, 1 million euro will be devoted to engaging researchers from outside the consortium in microprojects and inviting them for visits within the HumanE AI Net excellence centers (see section 1.3.2.3). Through this mechanism, we envision 50–100 researchers from different groups becoming involved with HumanE AI Net on concrete projects.

Each network will develop collaboration with the relevant Digital innovation Hubs, to disseminate knowledge and tools, and understand their needs

The strategy toward interfacing with the DIH is described in section 1.3.2.5 and Task 9.2 is devoted to it. The responsible partner (Fortis) is also a partner in the SmartAnythingEverywhere coordination action, and thus well-positioned to coordinate an efficient collaboration

… networks should also foster innovation and include mechanisms to exploit new ideas coming out of the network's work (for instance via incubators)

WP 7 is devoted to "Innovation Ecosystem and Socio-Economic Impact," including a brokerage platform (T7.2), various networking events (T7.7), and training and support to help researchers create innovation (T7.4).

each proposal will define mechanisms to become a virtual center of excellence, offering access to knowledge and serve as a reference in their chosen specific field, including activities to ensure visibility

WP 8 is devoted to creating a Virtual Center of Excellence (including a Virtual Laboratory in Task T8.1 And T8.2) and capacity building. As described in section 2.2.1.3 the Virtual Laboratory will connect closely to the AI4Eu platform, provide easy access to all materials and tools produced by the project, and provide easy mechanisms for anyone to get in contact and interact with the project. WP 9 is devoted to interacting with various parts of the European research community related to AI.

1.2.4. Synergies with the AI on Demand Platform

As described in section 2.2.1.3 AI4EU is crucially at the center of the HUmanE AI net strategy for making knowledge and content available to the community, collaborating internally, and communicating. Task 9.1 is devoted to coordinating the collaboration with AI4EU and run by Thales, AI4EU's coordinator. Task 8.3 is devoted to implementing the infrastructure needed to use the AI4EU platform for content exchange and collaboration within the scientific community and run scientific challenges. It is led by Allessandro Saffiotti from Orebro, who is the research manager in the AI4EU project. The Virtual Laboratory implementation (Task 8.1) has an explicit focus on integration with AI4EU. The research WPs (1–6) all have the deposition of papers, tools, and datasets on the AI4Eu platform as deliverables (D1.1, D2.1, D3.1, and D4.1).

1.3. Concept and Methodology

The project concept starts with the vision of what a European brand of human-centric AI should be: beneficial to individuals and the society as a whole, trustworthy, ethical and value-oriented, and focused on enhancing user's capabilities and empowering them to achieve their goals. We then consider the gaps in AI knowledge and technology that must be closed to make this vision a reality and allow products and services to be developed around it.

The HumanE AI Net consortium has been built on the basis of the HumanE AI FET preparatory action to ensure coverage of all the required competences and engagement of all the key European players.

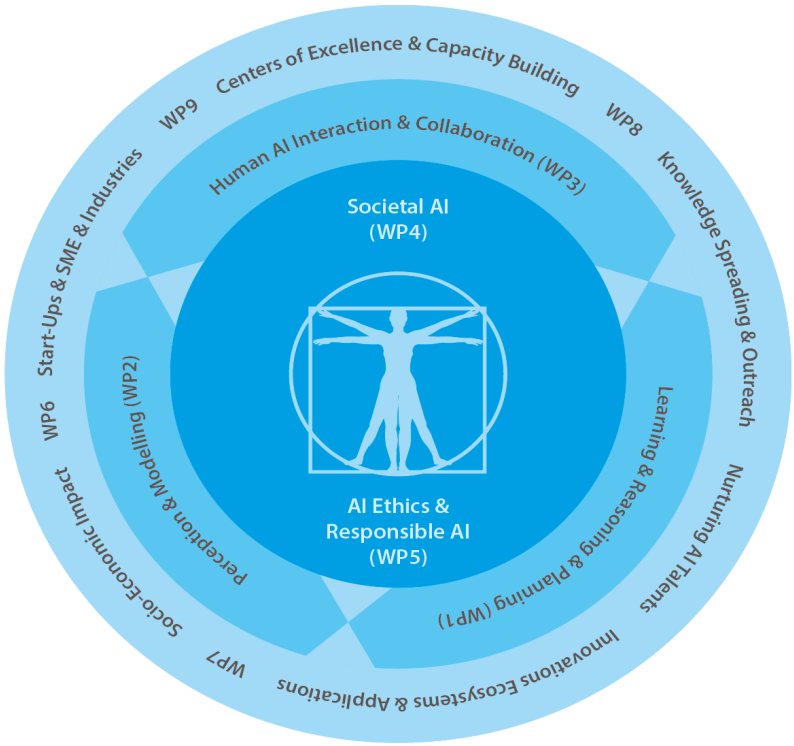


Figure 1. The HumanE AI Net Vision.

1.3.1. Concept

1.3.1.1. Overall R&D Vision

At the core of our human-centric AI concept is the need to let people interact and collaborate with AI systems and AI-enhanced environments in a way that facilitates synergistic co-work, co-creation, and enhancing each other's capabilities. The interaction must be closely connected to the computational models and systems' perceptual capabilities, as well as factoring users' social and cultural diversity. Collaboration with humans requires that humans and AI systems work together as partners to achieve a common goal, sharing a mutual understanding of each other's abilities and respective roles. Human-level performance in collaboration will require integration of learning, reasoning, perception, and interaction.

One of the pertinent issues concerning human-centric interaction and collaboration is to go beyond HCI challenges, and to ensure the human maintains control over these interactions and collaborations. It is our human values that should shape these interactions, it is our goals that must be fulfilled, and it is our benefit that must be achieved through these interactions. Giving the human control over the interactions is key, and this includes understanding how interactions are being driven (transparency) and having a say in changing how we interact, and even why we interact (the goals) whenever needed.

To achieve this research agenda, our vision (Figure 1) is built around ethics, values, and trust. These are intimately interwoven with the impact of AI on society, including problems associated with complex dynamic interactions between networked AI systems, the environment, and humans. In the project, such ethical and social aspects are not just boundary conditions but important research topics at the interstice of AI, philosophy, social science, and complex systems. They are addressed in WPs 4 and 5 and the concepts are explained in more detail below. With respect to core AI topics, fundamental gaps in knowledge and technology must be addressed in three closely related areas.

The first area is learning, reasoning, and planning methods, which allow for a large degree of interactivity. Thus, to facilitate a collaboration between humans and AI systems based on trust and enhancing each other's capabilities, AI must not only be able to provide explanations at the end of the learning or reasoning task. Rather, it must continuously give feedback on its progress and be able to incorporate complex high-level human input incrementally. We refer to such novel methods that go beyond merely explainable AI as "human-in-the-loop learning reasoning and planning" and consider them in WP 1 (see also section 1.3.1.2).

Second, building on the aforementioned learning and reasoning methods, is multimodal perception of dynamic real-world environments and social settings, including the ability to build and maintain subtle yet comprehensive models of such environments and the humans interacting within such environments (including the associated social structures and processes). In essence, to work together in a seamless, synergistic way, humans and AI must share an "understanding" of not only the immediate problem at hand, but also the problem's larger context.

Third, understanding the human, environment, and social setting is a necessary but insufficient condition for seamless, effective, and human-friendly collaboration and co-creation in mixed human-AI settings. In addition,

appropriate novel interaction and collaboration mechanisms must be developed on both the individual and collective level.

1.3.1.2. Human-in-the-Loop Machine Learning, Reasoning, and Planning

Learning, reasoning, and planning are interactive processes involving close synergistic collaboration between AI system(s) and user(s) within a dynamic, possibly open-ended real-world environment. Key gaps in knowledge and technology that must be addressed toward this vision include the following:

1. Hybrid representations that combine symbolic, compositional approaches with statistical and latent representations [Garcez 2019][vanHarmelen 2019]. Such hybrid representations are needed to allow the benefits of data-driven learning to be combined with knowledge representations that are more compatible with the way humans view and reason about the world around them. A wide variety of representations will be investigated by the consortium, including hybrids of logic and neural networks, such as logic tensor networks [Donadello2017], and latent representations of knowledge graphs through embeddings [Wang2017] and narratives [Meghini2019].
2. Methods for leveraging the above representations to not just present humans with explanations based on simple links between the input and output spaces (e.g. [Koh2017]), but to be able to reason about shared internal representations just like humans can intuitively explain to others how they arrive at certain conclusions. This is closely connected to work on human-AI collaboration (see 1.3.1.4) that will study how to present such reasoning to humans in various situations.
3. Methods for interactively including high-level human understanding in the learning and reasoning process, which is difficult with current data driven approaches. The injected knowledge can take the form of conceptual categories, knowledge of causality, and common-sense knowledge. Infusing such human knowledge into the machine-learning process increases data efficiency, and improves the learned results' generalisability and robustness [Marcus2019].

Under this vision, the knowledge of an AI system evolves and is influenced by its behavior in the world and its human interactions. We want to facilitate systems that can learn, reason, plan, act, and observe the world, by continuously and cyclically interleaving all these activities.

1.3.1.2.1. Linking symbolic and sub-symbolic learning

The construction of hybrid systems that combine symbolic and statistical methods of reasoning is widely seen as one of the grand challenges facing AI today. For example, Pearl and colleagues noted, "Our general conclusion is that human-level AI cannot emerge solely from model-blind learning machines; it requires the symbiotic collaboration of data and models" [Pearl2018]. Marcus and colleagues stated, "By pushing beyond perceptual classification and into a broader integration of inference and knowledge, artificial intelligence will advance greatly." [Marcus2018], [Marcus, Davies, 2019]. Going further, Darwich[6] noted, "the question is not whether it is functions or models but how to profoundly integrate and fuse function-optimisation with model-based reasoning" [Darwiche2019[7]]. However, as shown in two of our survey papers [Garcez 2019][vanHarmelen 2019], there is no consensus on how to achieve this, with proposed techniques in the literature ranging from graph theory to linear algebra, and from propositional logic and fuzzy logic to continuous differentiable functions.

An interesting approach is the consideration of **narratives**—which are particularly natural representations for humans that might well offer a fruitful common ground with machine representation, an insight that goes back to early work in AI on scripts [Schank1975]. However, to avoid the limitations of earlier work, such scripts will need to be automatically *generated* (see [Jorge2019] for an overview of the state of the art), and we must develop techniques for *using* such scripts for shared human-machine understanding [Bosser2018] and explaining [Jentner2018][Calegari2019].

1.3.1.2.2. Learning with and about narratives

People share knowledge by narrating stories. A story places a series of events in a larger context. A narrative is, essentially, an interpretation of a story that recounts a series of events and their consequences. A narration interprets the story in a manner that can be generalized and used to predict and explain events.

People use narratives to understand phenomena. Narratives make it possible to provide rich descriptions for events that are not directly observable, including prior events, and hypothetical or abstract events. Narratives enable predictions for possible future events, and to reason about how to create or avoid events.

We will investigate the use of narratives to provide human-understandable descriptions for complex situations, and subsymbolic representations [Urbaniak2018][Gilpin2018]. We also will research how narratives can be adapted as a bridge between human reasoning and understanding, on the one hand, and internal AI representation on the other [Vlek2016]. Specifically, we will address the following questions:

1. How can AI systems process and learn from human knowledge expressed in the form of narratives and stories [Jorge2019]?
2. How can AI systems explain their reasoning, learning, and acquired knowledge in the form of narratives and stories that humans can easily understand and relate to [Pasquali2019][Gervás2019]?
3. How can AI systems and humans jointly create, adapt, and interpret narratives or stories as a means of interactively reasoning and learning together [Bosser2018] ?

1.3.1.2.3. Continuous and incremental learning in joint human-AI systems

One specific challenge we will tackle toward AI systems with humans in the loop is the use of hybrid representations in **joint human-machine learning and planning**. An early example of this in reinforcement learning is [Garnelo2016]. Rather than the typical opaque representations usually learned in deep reinforcement learning systems, the goal is to learn an intelligible abstraction of the state-space (the world) and the possible transitions, and then learn a reward function over this abstract model, rather than the latent representation. More recent examples by consortium members are in [Toro_Icarte2018] and [Lever2016].

A second challenge is the use of hybrid representations in **generating explanations based on shared models between humans and machines.** This implies upgrading the knowledge-discovery process with the capability of generating high-quality machine-learning models equipped with their own human-comprehensible description, which in turn requires a novel blend of mathematical and statistical models with logic and causal inference and reasoning [Peters et al., 2017]. Work by consortium members such as [Tiddi2015] shows how background knowledge in the form of very large knowledge graphs can be used to generate intelligible explanations that are not constructible from data alone. [Guidotti2020] exploits auditing methods of machine-learning models to generate explanation rules reconstructing both factual and counterfactual knowledge. Other work exploits symbolic representations such as Inductive Logic Programs to explain the neural network-generated labels of objects in images [Yang2019].

The above will provide the basis for learning with humans in the loop, e.g., by exploiting rich human feedback ("this is wrong *because*..."), exploiting implicit feedback (by obtaining feedback from behavior, voice, and face), through imitation, and via active learning (the machine asking the human "Should we explore **this**?").

1.3.1.2.4. Compositionality and automated machine learning (Auto-ML)

Major breakthroughs in recent AI developments have come when well-understood learning components have been composed to create more complex behaviors and systems as, for example, in AlphaGo [AlphaGo], where a deep learner analyzing the board value is combined with a reinforcement learner implemented by using a deep learner to estimate the value function and a probabilistic method of prioritizing the exploration of the search space. These compositions are typically ad hoc and heuristic, requiring trial and error to deliver stable solutions. The HumanE AI project will develop a theoretical foundation for the composition of learning components, be they symbolic or subsymbolic, enabling the reliable engineering of systems that can deliver specified complex cognitive behaviors. The HumanE AI project will thus enable the combination of symbolic and statistical AI methods and further extend them with theoretical models that allow continuous adaptation.

The compositional approach to delivering AI systems has a number of advantages apart from the obvious inspiration of general software engineering. In addition to reducing more complex problems to well-understood components, it renders systems more transparent in that the decisions of one component can be traced to outputs of others via well-understood functionality. A key approach to unlocking the potential of the compositional approach is its link to optimization. Overall objectives for the cognitive system can be translated into optimization criteria that respect various constraints: there is then a natural correspondence between distributed strategies for solving the optimization problem and decompositions of the cognitive system. This also underpins our proposed approach for rendering AI systems interpretable by learning to decompose them into simpler components, an approach that can identify structure in the solution, hence rendering it more robust and explainable.

The recent success of general-purpose algorithm configuration and selection methods—and notably, the rise of automated machine learning (or AutoML)—already leverages this insight [KotEtAl17[8]]. In work on HumanE AI Net, we build on this foundation, mostly devising methods for automating the development, deployment, and maintenance of AI systems that are performant, robust, and predictable, without requiring deep and highly specialised AI expertise. The key to achieving this vision of automated AI (or AutoAI) is our proposed approach for rendering AI systems interpretable by learning to decompose them into simpler components, which can automatically identify key structure in the solution, hence rendering it more robust and explainable.

1.3.1.2.5. Quantifying model uncertainty

While it is well-documented that our ability to reason about probabilities has its limitations, it is equally clear that likelihood estimation forms an important part of both our understanding of the world and the way in which we communicate that understanding to others [JL]. For AI to interact meaningfully with humans it must use the vocabulary and semantics of probabilistic arguments in a way that is accessible and understandable to humans. However, uncertainty quantification is not just important as a vocabulary of communication, it is also a vital component if an agent is to weigh different alternative interpretations of a situation, to assimilate information from different sources, and to make decisions about what new information would be most useful in disambiguating a concept or question.

Indeed, it can be argued that probability provides the most natural measure for cross-modal calibration and integration of information. It therefore is natural that HumanE AI will investigate methods for both assessing and quantifying uncertainty of individual models, but also of the ways in which this can be inferred when models and/or information are combined, hence propagating measures of uncertainty through composite systems. Uncertainty will be important at all of the aforementioned levels, from assessing the confidence of individual estimations to the likelihood of logical relations or narratives in a particular context.

There are a variety of methods for estimating blackbox uncertainty, such as Bayesian posterior distributions estimated for example in dropout models for deep learning networks, or more precise measures such as conformity that can guarantee accurate percentile bars that hold with high confidence [candes, vovk]. Extending such approaches to composite and dynamical systems will be an important focus to inform decisions made by an agent, either to increase its information or alternatively trade information gain with expected success, as in bandit-style algorithms [bandit ref]. Such uncertainty estimates will also link with hard or soft constraints that must be placed on a system in order for its behavior to be "safe" or "desirable." The best way to monitor and model the uncertainty will be investigated with approximate reasoning techniques as well as separate modeling "watching" networks. The approaches will be important in driving lifelong learning algorithms in which uncertainties will determine which models need refinement and/or verification from new data, which might also be sought through interaction with humans by asking for clarifications.

1.3.1.3. Multimodal Perception and Modeling

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.

Actions can have quite different meanings, depending on contexts. Human interaction and human collaboration depend on the ability to understand the situation and reliably assign meanings to events and actions. People infer such meanings either directly from subtle cues in behavior, emotions, and nonverbal communications or indirectly from the context and background knowledge. This requires not only the ability to sense subtle behavior, and emotional and social cues, but an ability to automatically acquire and apply background knowledge to provide context. Acquisition must be automatic because such background knowledge is far too complex to be hand-coded.

To illustrate the challenges of perception for human-centric AI, consider the following simple example. A person holds an object. Another person approaches saying "give me that" and attempts to take the object from the first person's hands. Today's systems are able to recognize the object, the action of holding, to identify the two persons and to recognize the spoken command. Understanding the larger story of the two people's interaction is a far more difficult task that is currently beyond the state of the art.

In such a situation, a human completes the observations with a rich supply of contextual information based on experience. The human is able to provide a narrative that can explain why one person is attempting to take the object from the other, and predict the likely consequences. For example, the second person may be stealing the object, or he or she may be recovering a stolen object, and the second person may be helping the first person for whom the object is too heavy, or the first person may be helping the second one out by fetching the object in question. In general, a human observer is able to choose between a large set of possible narratives based on subtle clues such as clothing, each person's age and gender, and the nature of the place where the scene occurs. Research on artificial systems with such abilities will require a strong foundation for perception of humans, human actions, and human environments. In HumanE AI Net, we will provide this foundation by building on recent advances in multimodal perception and modeling sensory, spatiotemporal, and conceptual phenomena.

1.3.1.3.1. Multimodal interactive learning of models

Perception is the association of external stimuli to an internal model. Perception and modeling are inseparable. Human ability to correctly perceive and interpret complex situations, even when given limited and/or noisy input, is inherently linked to a deep, differentiated, understanding based on human experience. Current limitations of computer perception are rooted in an inability to acquire and use such background knowledge.

We will develop technologies for models that integrate perception from visual, auditory and environmental sensors to provide structural and qualitative descriptions of objects, environments, materials, and processes. Such models are required to organize and provide context for perception of objects, events, and actions. Models should make it possible to associate and organize spatio-temporal auditory and visual perception, with the geometric structure of an environment, and the functional and operational properties of objects and structures.

1.3.1.3.2. Multimodal perception and narrative description of actions, activities and tasks

People perceive and understand the world not just as objects and events, but as narratives that situate objects and events within a context and establish causal relationships. Context and causality enable rich descriptions for events that are not directly observable, including hypothetical or abstract events, and events that occurred in the past.

Current approaches to action recognition simply detect actions from spatiotemporal signatures and state changes in the environment, without placing the activities in the larger context of an activity or task. Such abilities will be required to predict the intended and actual consequences of the action, and explanations for the purpose of the action.

Monitoring of manipulation activity requires recognition of manipulation actions in the context of an activity. The activity context provides constraints that can be used to focus attention on the objects and materials to be manipulated, and to disambiguate recognition results that are uncertain or ambiguous. This disambiguation applies to recognition of actions as wells objects and materials. The use of activity context can reduce both the error rate and the computational cost for action recognition.

A manipulation activity can be formalized as a process, modeled as a series of state transitions. With this approach, the activity is monitored as a series of states, where the process state is the composition of the states of the individual objects. This process state is referred to as a situation [Johnson-Laird 85]. The situation model provides context for the action, making it possible to describe the action as part of story [Genette 72] with context information about why and how the action was performed. This story may then be interpreted as part of a narrative, associating contextual information that make it possible to explain the action and predict its consequences. The results may be used to drive a natural language generation (NLG) tool to communicate and interact with a human collaborator.

1.3.1.3.3. Multimodal perception of awareness, emotions, and attitudes

Human awareness is constrained by limits to working memory and perceptual abilities. Modeling awareness is required to permit a system to predict human abilities and construct explanations. Awareness can be perceived from fixation, head orientation, posture, and vocal interjections, as well spoken language interaction. Emotions play a fundamental role in human reason, and can be perceived from physiological signs such as micro-expression, heart rate, posture, self-touch, prosody, and paralinguistic expressions. Attitude condition (how humans react to phenomena) can be determined from patterns of reactions, as well as direct spoken language interaction.

Much of human activity is reactive and unconscious. At the most basic level, sensory signals directly drive human muscles and emotional responses at the signal level in a tightly coupled interaction. Multiple sensor modalities, including tactile, visual, and auditory may be combined in such signal-level interaction. Systems that interact and collaborate with humans must appropriately respond to such signals.

Going beyond emotions to understanding human intentions, attitudes, and related values is an important topic for psychology, sociology, and philosophy with so far little work within AI. Some previous work has focused on agent-based modeling [Georgeff98], with few results in real-time recognition in interactive, real-world scenarios. Our approach is based on the assumption that comprehensive world models, combined with the ability to seamlessly involve humans in the learning and reasoning process, will be instrumental in addressing this topic. We leverage synergies with the respective activities within HumanE AI Net to develop AI systems that can, at least to a degree, recognize and reason about user motivations, attitudes, and values; meanwhile, the systems' interactions with humans will greatly contribute toward making the vision of a European brand of human-centric AI a reality.

1.3.1.3.4. Perception of social signals and social interaction

Most research on perception of human interaction tends to focus on recognizing and communicating linguistic signals. However, much human-human interaction is nonverbal and highly dependent on the social context. A

technology for situated interaction will require abilities to perceive and assimilate nonverbal social signals, to understand and predict social situations, and to acquire and develop social interaction skills. Brezeal and colleagues [Breazeal2016] have recently surveyed research trends in social robotics and its application to human-robot interaction (HRI). They argue that sociable robots must be able to communicate naturally with people using both verbal and nonverbal signals, and engage users on both cognitive and emotional levels to provide effective social and task-related services.

Our goal is to develop methods to endow an artificial agent with the ability to acquire social common sense using the implicit feedback obtained from interaction with people. We believe that such methods can provide a foundation for socially polite HCI, and ultimately for other forms of cognitive abilities. We propose to capture social common sense by training the appropriateness of behaviors in social situations. A key challenge is to employ an adequate representation for social situations.

Knowledge for sociable interaction can be encoded as a network of situations that capture both linguistic and nonverbal interaction cues and proper behavioral responses. Stereotypical social interactions can be represented as trajectories through the situation graph. We will explore methods that start from simple stereotypical situation models and extend a situation graph by adding new situations and splitting existing situations.

1.3.1.3.5. Distributed collaborative perception and modeling

People have a shared ability to explain observed phenomena and predict future phenomena based not only on direct experience, but on experience learned from others. Sharing of information provides a cultural background that is accepted as true within a culture and provides a powerful foundation for reasoning and communication through common sense. Human narratives convey information concerning what sequences of behaviors are required in specific social situations. Narratives are the source of prediction for how other actors will behave in a specific situation, as well as determining the agent's appropriate reaction to these behaviors, along with the consequence of these actions. Narratives also are used to explain the behavior of others. We need an ability for intelligent systems to learn common sense from experience shared by others. To participate as members of technosocial groups, and engage in collaborative perception and modeling, intelligent systems must be able to represent narratives, understand narratives communicated by other group members, communicate their own knowledge in the form of narratives, and integrate their own narratives with the narratives of other group members.

1.3.1.3.6. Methods for overcoming the difficulty of collecting labeled training data

Getting sufficiently labeled training data is a core concern for many ML domains. For example, much of the recent progress in computer vision and language processing has been related to the availability of huge public datasets (e.g., the 1-million-picture ImageNet dataset [Deng2009]), which enabled public ML challenges (e.g., the Large-Scale Visual Recognition Challenge with ImageNet [Russakovsky2015]). However, for multiple reasons, it is particularly grave when it comes to the perception of complex real-world situations, such as those involving humans, which besides performing actions also engage in social interactions, perceive emotions, and so on. Some reasons for these challenges include the following:

- Annotation, often executed by visualizing video-recording of experiments, can be extremely slow. Some researchers report that annotating 10 minutes of video may take as much as 10 hours of time, when very fine-grained and accurate annotations are required [Roggen2010]. Other researchers surveyed publicly available datasets and reported the costs of creating them, with many public datasets costing in the hundreds of thousands of dollars [Welbourne2014].
- Annotations often carried from video recordings cannot be used when natural behavior must be captured, as such recordings are likely to influence how naturally people behave, and in a number of situations it may not be possible to collect such datasets for ethical reasons, which limits collection to in-lab "naturalistic" emulation of everyday scenarios.
- Annotating datasets require someone previously enumerating a set of elements to identify (e.g., interesting situations or actions) within a recording. In long-duration datasets—which are especially used for unsupervised and open-ended perception and modeling—an a priori exhaustive enumeration of such interesting situations is infeasible, and post hoc re-annotation from videos may not be possible, depending on the nature of the experiment, informed consent, and ethical considerations.
- Finally, interactive systems, where an activity-aware system interacts with humans (e.g., in a human-robot interaction task) will also influence people's behavior dynamically. Therefore, such experiments cannot be captured easily in a static dataset.

These challenges open a wide number of fundamental research areas to enable the collection of larger scale and more realistic datasets, which will be explored in this project. Some of the approaches that will be pursued include:

- Leveraging the large availability of online datasets, and devising methods to transform such datasets to make them suitable for the sensor modalities available. For instance, recent work has shown that sensor data can be seamlessly transformed across modalities (e.g., between RGB images and wearable sensors [Fortes2019] or depth sensors and wearable sensors [Banos2012]). These can be combined with the availability of online datasets that are potentially of different modalities (e.g., YouTube data that comprises textual annotations to be converted to data suitable for wearable sensors).
- Leveraging crowdsourcing to annotate datasets, with AI approaches to support the efficient annotation (e.g., identifying relevant time segments), improve robustness (inter-rater reliability), and design the technical infrastructure to integrate these approaches (e.g., [Satybaldiev2019]).
- Exploiting human-in-the-loop learning, such as letting users provide annotations at their own pace through a combination of active learning (e.g., prompting users about their activities) and semisupervised learning. A primary challenge is to identify when it is most valuable to prompt a user, as a combination of information gain and minimal distraction.

1.3.1.4. Human AI Interaction and Collaboration

Beyond considering the human in the loop, the goal of human-AI is to study and develop methods for combined human-machine intelligence, where AI and humans work in cooperation and collaboration. To achieve this, we will investigate principled approaches to support the synergy of human and artificial intelligence, enabling humans to continue doing what they are good at but also be in control when making decisions. Our mutual motivation of this goal is that it is critical for Europe, which has set trustworthy and controllable AI as its goal. Within the US, attempts to operationalize similar agendas are already far. For example, Stanford University proposed that AI research and development should follow three objectives: (i) to technically reflect the depth characterized by human intelligence; (ii) improve human capabilities rather than replace them; and (iii) focus on AI's impact on humans [Li2018]. There has also been a call for the HCI community to play an increasing role in realizing this vision, by providing their expertise in the following: human-machine integration/teaming, UI modeling and HCI design, transference of psychological theories, enhancement of existing methods, and development of HCI design standards [Xu2019].

1.3.1.4.1. Foundations of human-AI interaction and collaboration

Here, we break down human-AI into three main types: (i) collaboration, (ii) interaction, and (iii) symbiosis. When studying interaction, we study AI methods that understand people and can anticipate the consequences of their actions on people, and communicate their purposes so as to ground collaboration. This level also involves seeking more natural ways to communicate with AI, including multimodality, conversations, and augmented reality (AR) interfaces. At the level of collaboration, we consider concepts like cooperation, emotional intelligence, collective intelligence, and group work. At the level of symbiosis, we study emergent properties of AI systems where people and AI combine their processes, skills, and experiences to achieve something greater together than just by themselves.

The knowledge that HCI can bring to bear on these three forms of human-AI include user modeling, inference, and machine-learning methods suitable for interactive settings with humans, deep empirical research and the design of interaction techniques, and user interfaces for interaction with artificially intelligent partners. Empirical methods in HCI can provide a way of discovering the mental representations people develop when using AI systems, their expectations when interacting with an AI system (e.g., a robot, chatbot, recommender system, or diagnostic tool), and their acceptance of the decisions it suggests or actions it makes itself. Understanding these aspects better is crucial for humans to be able to collaborate with AI, but also for AI methods such as inverse RL[9] or machine theory of mind. There are other questions that must be answered: How is user acceptance and the adoption of AI systems affected by the cultural and social background of the user? What is the overall effect (short and long term) of interacting with intelligent systems on humans and the environment?

A core capability in human-AI interaction is *understanding* of human partners. While present-day ML research mostly approaches this as a classification or prediction task in supervised or unsupervised learning, we seek a new foundation from theories of human behavior. In particular, we believe that models and theories from computational psychology [Sun2008], computational cognitive sciences [Kriegeskorte2018], and computational social sciences [Lazer2009] can underpin artificial understanding of human behavior. This research calls for plausible models of human behavior that we will use for artificial agents that can—thanks to causal models that link behavior with cognitive, emotional, and other latent factors—better infer, plan, and act without extensive data on an individual [Lake2017]; this, of course, also presupposes high-quality language capabilities in both speech and text domains (T3.6), to analyze the speech and/or text to a formalized representation suitable to provide "input" to the aforementioned theories and models that will then be able to arrive at the true understanding of human behavior.

In a number of microprojects, we assemble the complementary expertise of our consortium members to open new avenues to explore the three types of human-AI and address associated research with these questions. One such example is a study of social practices on greeting rituals.

1.3.1.4.2. Human-AI interaction and collaboration paradigms

Given a basic understanding of the way humans approach AI systems, concrete interaction paradigms must be developed. Furthermore, for humans and AI to be able to collaborate toward common goals, they must be able to interact and *understand* each other, establish common ground, and see the other's perspective (thus having a type of Theory of Mind). As such, there are several research questions:

- When should the AI system's processes be externalized (and which ones), so that system functionality metes the right level of transparency to the user?
- How should relevant information about internal processes of the AI system be represented, to make it intuitively understandable to the user?
- How can humans intuitively express their complex thoughts and suggestions as a form of dialogue with respect to AI reasoning?
- What types of interaction are suited to different situations and with other humans they need to work with?

Our approach to answering these questions combines theoretical analysis with empirical user-centered design. First, in terms of *theoretical analysis*, we will analyze interactive problems as games and decision problems. For example, we will use Markov Decision Processes, which can be solved in simulation or in some cases analytically. These will be simplified such that we can infer conditions under which information disclosure between the two partners does or (or does not) work. Second, in terms of *empirical user-centered design*, these ideas will be developed in a user-centered manner with prestudies of the particular applications, and evaluated empirically with representative user groups.

1.3.1.4.3. Reflexivity and adaptation in human-AI collaboration

Key questions for systems where humans and AI work with each other synergistically to support each other as partners in co-creation are:

- How do humans and machines continuously adapt to each other and the context?
- How can machines understand their impact on humans before taking action?
- How do we design self-aware systems that can monitor and self-diagnose their interactions with the environment and other humans to self-improve their interactions?

Our approach is to build on meta-reasoning methods, wherein the behavior of the artificial agent is supervised at a higher level, which the consortium has explored earlier, for example, in ubiquitous computing.

In particular, our work will entail methods for meta-reasoning between the human and AI system, where they can ask together or to each other "Are we doing the right thing?" or "Is it ethical what we are suggesting?" On the interaction side, our goal is to enhance reflection by having a small dialogue at particular times. Often AI systems are developed to advise or suggest without the opportunity for negotiation or understanding. A recent suggestion is that AI systems should explain their decisions. Our work will develop solutions that determine what to ask and when and how, which at the machine-learning side will combine aspects of active learning, sequential planning, and reasoning.

1.3.1.4.4. User models and interaction history

User models can be divided according to how humans mind is represented in interaction (e.g., neural, mathematical, simulation, RL-based, and Bayesian) [Kriegeskorte2019], and which factors are included (e.g., cognitive, physiological, emotions, and motivational). We here pursue two important capabilities that user models should have: (1) *forward modeling,* or providing a richer and more generalizable account of human behavior suitable for real-world interactive AI, which has been an issue in cognitive and user models for decades, and (2) *inverse modeling,* or fitting models to individual users. Both are needed for deployment in interactive AI, that must on the one hand update its model representations with interactions and, on the other, select actions while anticipating their consequences on users (counterfactuality) [Lake2017]. Recent user models have also used reinforcement learning, wherein the state-space quickly explodes with longer user history, or embeddings (Rabinowitz2018) that collapse multidimensional behavior to a lower-dimensional, but uninterpretable, representation. We here seek model-based

(e.g., probabilistic graph models) approaches that allow combining inferential and learning capabilities with explicitly specified structures.

In addition, the research will develop **interaction history trails** that can: (1) keep a record of previous encounters so that they can be referred to in subsequent interactions between the users and the AI system and (2) decide on what should be forgotten in a human-AI encounter or interactions (ethically, legally, and morally, to stay feasible).

1.3.1.4.5. Visualization interactions and guidance

Visualization remains an important aspect of interaction between humans and complex systems. Visual analytics (VA) supports the information-discovery process by combining analytical methods (from data mining to knowledge discovery) with interactive visual means to enable humans to engage in an active "analytical discourse" with their datasets ([Keim, et al. 2010], [Thomas et al., 2005]). However, for humans/users, who are usually experts in their application domains but not in VA, it is difficult to determine which VA methods to use for particular data and tasks. Guidance is needed to assist humans/users in selecting appropriate visual means and interaction techniques, using analytical methods, and configuring instantiation of these algorithms with suitable parameter settings and combinations thereof. After a VA method and parameters are selected, guidance is also needed to explore the data, identify interesting data nuggets and findings, and collect and group insights to explore high-level hypotheses, and gain new knowledge.

Guidance has its roots in HCI and can be seen as a mixed-initiative process [Horvitz, 1999]. As Ceneda and colleagues note, "Guidance is a computer-assisted process that aims to actively resolve a knowledge gap encountered by users during an interactive visual analytics session" ([Ceneda et al., 2017], p. 112). A mixed-initiative process is an approach whereby both humans and systems can "take the initiative" and contribute to the process. The central elements are the time, degree, and type of involvements. Guidance is a dialogue between humans and systems in which humans provide—implicitly or explicitly—their own needs and issues as input and the system provides possible answers to alleviate problematic situations [Ceneda et al., 2018].

1.3.1.4.6. Natural language processing and conversational AI

NLP is an enabling technology for several, if not most, areas of the HumanE AI Net proposal, especially the perception, interaction, and HCI areas, with fundamental methods being tackled in machine learning, too. Natural language is a natural way of communication with humans, be it in speech or text form (without any prejudice toward nonverbal components of communication, such as emotion and gesture).

Today, as a rule, NLP uses deep-learning techniques. The main focus in this proposal, however, is to move away from this "blackbox" (yet often highly successful) approach, to connect the high performance of the neural network paradigm with symbolic methods—especially in the area of semantics and understanding, where existing, human-understandable databases and ontologies are used to approximate world knowledge.

Also, NLP (speech and text analysis, as well as NLG) is a necessary component of communication and interaction, such as in dialogue systems of all sorts. Similarly, when a computer must generate a narrative (to explain reasoning or arguments), NLG must be used; and conversely, (almost) any human input must be tackled first by an NLP component (or at least be integrated with it).

In keeping with WP 3, T3.6, the main areas of research would then be (1) analyzing natural language speech and text beyond the current state of the art; (2) NLG from planned, formally represented communication; (3) explaining "why" in deep general understanding systems; (4) multilingual issues in all of the above, and machine translation where needed for cross-lingual understanding and communication; and (5) creating and unifying an ontology where needed (e.g., on event types).

A key research theme building on NLP is to enhance human reflection on the actions they are carrying out (e.g., decision making, problem solving) by having a small dialogue with the AI system at particular times. This is often referred to as conversational AI. Often AI systems are developed to advise or suggest without the opportunity for negotiation or understanding. An increasingly accepted notion is that AI systems should explain their decisions. Here, we propose something different, which is to support human-AI dialogues, where the human can.

1.3.1.4.7. Trustworthy social and sociable interaction

Reeves and Nass argue that a social interface may be the truly universal interface [Reeves 98]. Current systems lack ability for social interaction because they are unable to perceive and understand humans, human awareness, and intentions, and to learn from interaction with humans. Building on the research on the perception of human emotions the modeling of social context and complex, evolving world models we will address key challenges in enabling AI systems to act appropriately within complex social contexts.

Breazeal has proposed a hierarchy of four classes of social robots, from socially evocative to sociable. As one moves progressively up the hierarchy, robots' abilities to engage in social interaction increase. Within this hierarchy, socially evocative robots are designed to encourage people to anthropomorphize technology to interact with it. Socially communicative robots use human-like social cues and communication modalities to facilitate interactions with people. Socially responsive robots are able to learn from their interaction and social partners. Sociable robots are socially participative, and maintain their own internal goals and motivations.

Kendon [Kendon 75] argues for understanding social interaction as a form of dialogue. With this view, prosody and gestures are seen as annotations to the linguistic contents of interaction, serving to guide attention as well as to communicate nonlinguistic signals. Pentland [Pentlan2005] proposes an approach based on social signaling of attitude and attention, using such vocal cues as amplitude, frequency, and timing of prosodic and gestural signals. Such unconscious signals provide important cues about social situations and social relations that are not available in measures of affect. The importance of such signals is one of the reasons that we propose extending our investigation beyond visual perception into acoustic and tactile perception modes [Ta2015].

A second important issue is that the AI systems, when interacting with one or more persons (and possibly other autonomous AI systems), should consider the broader social context in which they interact. For instance, an e-health system should not recommend taking a walk at dinnertime as the whole family gets to the table. It should be aware of practices, narratives, norms, and conventions to fit the interaction within those structures. Social sciences have for decades studied emergent properties of social groups (Durkheim 1932); however, technosocial systems' emergent properties are much less understood.

1.3.1.5. Societal AI

As increasingly complex sociotechnical systems emerge, consisting of many (explicitly or implicitly) interacting people and intelligent and autonomous systems, AI acquires an important societal dimension. A key observation is that a crowd of (interacting) intelligent individuals is not necessarily an intelligent crowd. On the contrary, it can be idiotic in many cases, because of undesired, unintended network effects and emergent aggregated behavior. Examples abound in contemporary society. Anyone who has used a car navigation system to bypass a traffic jam knows. Each navigation system generates recommendations that make sense from an individual point of view, and the driver can easily understand the rationale behind the recommendations. However, the sum of decisions made by many navigation systems can have grave consequences on the traffic system as a whole: from the traffic jams on local alternative routes to ripples propagating through the system on a larger spatial scale, to long-term behavior changes that may lead to drivers permanently avoid certain areas (which can have a negative economic impact on disadvantaged neighborhoods), or artificially increase the risk of accidents on highly recommended roads.

The interaction among individual choices may unfold dramatically into global challenges linked to economic inequality, environmental sustainability, and democracy. In the field of opinion formation and diffusion, a crowd of citizens using social media as a source of information is subject to the algorithmic bias of the platform's recommendation mechanisms suggesting personalized content. This bias will create echo chambers and filter bubbles, sometimes induced in an artificial way, in the sense that without the personalization bias the crowd would reach a common shared opinion. Again, a recommender system that makes sense at an individual level may result in an undesired collective effect of information disorder and radicalization.

Aggregated network and societal effects and of AI and their (positive or negative) impacts on society are not sufficiently discussed in the public and not sufficiently addressed by AI research, despite the striking importance to understand and predict the aggregated outcomes of sociotechnical AI-based systems and related complex social processes, as well as how to avoid their harmful effects. Such effects are a source of a **whole new set of explainability, accountability, and trustworthiness issues,** even assuming that we can solve those problems for an individual machine-learning-based AI system. Therefore, we cannot concentrate solely on making individual citizens or institutions more aware and capable of making informed decisions. We also need to study the emerging network effects of crowds of intelligent interacting agents, as well as the design of mechanisms for distributed collaboration that push toward the realization of the agreed set of values and objectives at collective level: sustainable mobility in cities, diversity and pluralism in the public debate, and fair distribution of economic resources.

We therefore advocate the emergence of *societal AI* as a new field of investigation of potentially huge impact, requiring the next step ahead in transdisciplinary integration of AI, data science, social sciences, psychology, network science, and complex systems.

1.3.1.5.1. Graybox models of society scale, networked hybrid human-AI[10] systems

The general challenge is to characterize how the individual interactions of individuals, both humans and AI systems, with their own local models, as well as the social relationships between individuals, impact the outcome of AI models globally and collectively. Using a combination of machine learning, data mining, and complexity theory, we strive at understanding the networked effects of many distributed AI systems interacting together, some (or all)

possibly representing human users, therefore comprising a complex human and technical ecosystem. The different layers of this system are in mutual interaction, producing emergent phenomena which may range from synchronization to collapse.

Naturally, several questions regarding the considerable challenges emerge: How can systems be modeled adequately and predict these networked effects? What are the typical scenarios of system evolution? What are the relevant mechanisms and quantities to control to prevent a system from unpredicted/harmful behavior? How can researchers design collaborative, distributed learning and data-mining methods for AI systems that are motivated by the social mechanism for accumulating "common knowledge" and "collective wisdom" without unnecessary, unsustainable, and harmful centralized collection of raw personal data? What are the best ways to design and manage such a complex system, so that it behaves in a way that is compliant with ethical principles, while dealing with the Collingridge dilemma (i.e., designers must select solutions at the beginning from a broad variety of possibilities, but with little information about the perception of the suggested solutions, while proceeding in the process and accumulating feedback, the degree of available freedom for the design shrinks)?

1.3.1.5.2. AI systems' individual versus collective goals

Social dilemmas occur when there is a conflict between individual and public interest. Such problems may appear also in the ecosystem of distributed AI and humans with additional difficulties due to the relative rigidity of the trained AI system on the one hand and the necessity to achieve social benefit and keeping the individuals interested on the other hand. What are the principles and solutions for **individual versus social optimization** using AI and how can an optimum balance be achieved?

As already illustrated, these complex systems should work on fulfilling collective goals (or requirements). However, requirements change over time, as they also change from one context to another. How can we design and manage such complex sociotechnical systems that **adapt to our evolving requirements**? How can we maintain humans' control in such systems to ensure that it is the humans' requirements and values that are being considered?

A related question is how to design mechanisms that support distributed sociotechnical systems made of self-interested agents. Such systems should be both efficient and ethical. In other words, the challenge is to develop mechanisms that will result in the system converging to an equilibrium that complies with the European values and social objectives (e.g. income distribution) but without unnecessary losses in efficiency. Interestingly, AI can play a vital role of enhancing desirable behaviors in the system, e.g., by supporting coordination and cooperation that is, more often than not, crucial to achieve any meaningful improvements. Thus far, teaching and learning in repeated strategic situations were already studied theoretically [Camerer2002] and the experiments were conducted involving human players [Hyndman2012]. Importantly, however, AI technologies bring many new possibilities to the table [Crandall2018, Peysakhovch2018] because, unlike with human players, we now have a unique chance to design both not only the rules of interaction but also some of the participants as such. Our ultimate goal is to build a scheme of a socio-technical system in which AI not only cooperates with humans but if necessary helps them to learn how to cooperate as well as other desirable behaviors.

1.3.1.5.3. Societal impact of AI systems

How to **evaluate societal impact** of competing AI technologies and promote the ones more compliant with the European values? As one of the possible approaches, explore how to construct **in vitro (controlled) experiments** of the interaction between AI technologies and humans, in order to select the technological setup most suitable for the ethical standards.

Understand and model the way algorithms and AI technologies **reinforce/generate certain undesirable human behaviors** and emerging societal phenomena, like producing echo-chambers, opinion polarization, and bias amplification at the collective level. Ultimately, the goal is to develop AI systems that contribute to improving the **quality of and access to information**, deal with information noise and fake news, detect and counter manipulation, and deal with information overload.

AI will soon change substantially the relation between the governing and governed. New opportunities open to obtain feedback and make predictions to the effect of (intended) measures and several new ways for participation in decision making will emerge. What are the possibilities, the risks and the impact of **AI on governance**, considering the opportunities of AI assisted participatory technologies? How to understand and model strategies with which AI can enhance public involvement, help foresee social consequences of policies, facilitate social adaptability to change? How can AI contribute to the handling of the conflict between the different time scales of individual interests, legislation periods and the solution of global problems?

1.3.1.5.4. Self-organized, socially distributed information processing in AI based techno-social systems

Understand 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). It has been already

argued for long that social influence is the most fundamental and pervasive social process [Allport, 1932; McGuire1985]. For instance, mutual influences among individuals underlay formation of public opinion [Nowak1990], group decisions, and actions [DeDreu2008]. At the group level, social influence is tantamount to *distributed, optimizing information processing*. In particular, in this process, individuals optimize their decision-making and judgment by delegating information processing to potential sources of influence. In this context, the Regulatory Theory of Social Influence [Nowak2020] specifies four factors —*trust*, *coherence*, *issue importance*, and *own expertise*—that play a critical role in the processes of determining the target's choice of sources and the level of abstraction in the information sought from these sources. Beyond maximizing the cognitive efficiency of the target and the quality of his or her outcomes, these processes also enhance the functioning of the social group in which the target is embedded, because the most expert on the topic and the most reliable group members gather and process the information. Our intention is to use the research on human groups [Petty1986; Mullen1994; Nowak2020, Nowak2017] as a starting point of designing **AI members of socio-technical groups**, which can improve the functioning of the group in reaching optimal decisions and judgments. AI agents need to understand their role in distributed information processing social systems, be aware of the competence and reliability of group members, the importance of the issue at hand and their own limitations. Another challenge is to design the rules by which on the basis of this knowledge, AI agents can decide which information to process themselves and which to delegate to humans, and who in the group is most capable of processing which information. We propose to design mechanisms that enable AI agent to easily and as naturally as humans estimate trustworthiness of both human and other AI members of the group and to use trust estimates as a guidance for optimal information processing in social groups.

The ultimate goal is to develop enhance distributed information processing in socio-technical systems so that they provide a platform for common action. To this end, we will study the mechanism of self-organization in socio-technical groups at different scales from **common action**, e.g., in emergency response to societal movements. In this context, it is also important to understand how to achieve **robustness** of the human-AI ecosystems with respect to various types of malicious behavior, such as abuse of power and exploitation of AI technical weaknesses. Ultimately, we will develop principles for designing schemes of AI systems that are robust or resilient to manipulation and are at the same time incentive compatible.

1.3.1.6. AI Ethics, Law and Responsible AI

Responsible AI is about the processes by which AI is developed (ethics in AI design and development), accountability for the results of AI system deliberation (ethics by design) and making sure that those developing AI systems are aware of their role and impact on the values and capabilities of those systems (ethics for designers) [Dignum2019]. Design methods, verification techniques, and codes of conduct are all aspects that need to be developed alongside the computational design of algorithms [van den Hoven 2015].

Every AI system should operate within an ethical and social framework in understandable, verifiable and justifiable ways. Such systems must in any case operate within the bounds of the rule of law, incorporating fundamental rights protection into the AI infrastructure. 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 with legal, ethical and societal principles. This is not merely a matter of articulating legal and ethical requirements, but involves robustness, and social and interactivity design. Concerning ethical and legal design of AI systems, we will clarify the difference between legal and ethical concerns, as well as their interaction (Hildebrandt 2020), and ethical and legal scholars will work side by side to develop both legal protection by design and value-sensitive design approaches. The focus here is the prioritization of ethical, legal, and policy considerations in the development and management of AI systems to ensure responsible design, production and use of trustworthy AI. This requires integration of engineering, policy, law and ethics approaches. The following are the fundamental issues to be addressed by a research roadmap on ethics for human centered AI Systems.

- Adequate account (and criteria of adequacy) of **what a moral value is,** e.g. a Topos (used in moral narratives), especially in the context of the human AI interaction (with reference to some prominent values in the Ethics & AI debate, e.g. accountability, privacy, fairness.)
- Adequate account for **Designing for HumanValues and Ethical principles** (as non-functional requirements) in the field of AI. Design for Values, value hierarchies, functional decomposition of non-functional (moral) values.
- Methods for **measuring values and norms** in the human-AI ecosystem as required by an agile approach to designing for values.
- Understanding **how values can change** (or their balance is modified) as a side effect of complex interaction between humans and AI systems in a complex socio-technical ecosystem, also with respect to the above mentioned value hierarchy.
- Emergence and resolutions of **value conflicts** by design (epistemic power of machine learning versus data protection, explainability, responsibility vs adaptability (and emergent properties).

- Theory and methods to deal with ethical dilemmas and **value prioritization**, ensuring that such decisions are open, transparent and amenable to argumentation and participation of a wide range of stakeholders.
- Moral importance of epistemic conditions for **responsibility** for design and use of AI systems (e.g. contextuality of notions such as 'understanding' 'explaining' and making 'transparent' the working of deep learning).
- Understand the **relation of Humane-ness, human centeredness, human dignity** in the application of AI.

The overall goal is to boost research aimed at developing methods and methodological guidelines for the entire lifecycle of the AI system: design, field validation with stakeholders (simulations, sandbox), deployment and feedback through continuous oversight. This will include:

- Ensuring that **design processes** result in systems that are robust, accountable, explainable, responsible and transparent
- **Ethics for designers**: and making sure that those developing AI systems are aware of their role and impact on the values and capabilities of those systems
- Methods to elicit and align multi-stakeholder values and interest and constraints capable of **balancing societal and individual values** and rights
- Methods to integrate and validate a combination of **different possibly conflicting values** (Design for Values) describe dilemmas and priorities, and integrate them into the computational solutions
- **Compliance with laws** and regulation and with guidelines for ethical AI
- **Explainable AI** systems in support of high-stakes decision making (e.g., in health, justice, job screening)
- **Feedback methods** to inform policy-makers and regulators on missing elements in current regulations and practices.

The major research challenges are articulated in the following subsections.

1.3.1.6.1. "Legal Protection by Design" (LPbD)

Legal aspects will entail a focus on **the preconditions for ethical conduct,** for instance but not only: (1) **acountability** of those who take risks with other people's rights, freedoms and interests (by processing their data or targeting them based on data-driven inferences), (2) effective and **meaningful transparency** concerning the logic of automated decision systems that enables safe and meaningful interaction with such systems (both in the case of online search, social media and commerce and in the case of real-world navigation as in Internet of Things and robotics), (3) actionable **purpose limitation** to enable users (inhabitants) of AI environments to foresee the behavior of the myriad systems they may encounter (from connected cars to self-executing insurance contracts based on real-time data-driven input and care robots for the elderly), (4) reliable **proportionality testing** in the context of impact assessments, balancing the interests of providers against the rights, freedoms and interests of users or third parties that will suffer the consequences (whether algorithmic or data protection or safety and security impact assessment), in a way that enables them to contest such assessments, (5) built-in human-machine interaction that allows users or those targeted to **exercise their fundamental rights and freedoms** (enabling access to meaningful information, withdrawal from invasive targeting, detecting and contesting prohibited or unfair discrimination, and violations of the presumption of innocence).

This Legal Protection by Design (LPbD) entails the incorporation of fundamental rights protection into the architecture of AI systems. This plays out at **two levels**. -

1. This first concerns are the **checks and balances of the Rule of Law**, notably a concrete and effective set of interventions at the level of the research design, the subsequent development of core code, choice of programming language, foreseeable system behaviors, design of the APIs, and various types of interfaces. This concerns the **choice architecture instituted by law** that confronts (1) developers, (2) manufacturers, (3) sellers, (4) users (e.g. service providers, governments), and (5) end-users, providing them with leeway, proper constraints, transparency, accountability and foreseeability.
2. The second level concerns **requirements imposed by positive law that elaborates fundamental rights protection**, such as the GDPR, non-discrimination legislation, labour law and the more. Here, the point is to follow up on concrete legal norms (e.g. the right to withdraw consent as easily as it has been given), translating them into technical requirements and specifications when developing applications. LPbD differs from ethics by design because it concerns legal obligations that are **democratically legitimated** and **enforceable under the Rule of Law**.

The choice of which norms must be built-in therefore does not depend on the ethical inclinations of e.g. developers or service providers, but on **constitutional preconditions** for ethical behavior (e.g. ensuring that those who act ethically will not be pushed out of the market), and on **enforceable law**.

LPbD differs from mere techno-regulation, 'legal by design', or 'compliance by design' because it does not aim to nudge or technologically enforce e.g. administrative law, trying to turn legal obligations into technical measures, but instead aims to build transparency, accountability and contestability into these systems enabling a.o. **protection against "compliance by design**."

The **concrete results** will consist of:

A. a **listing of relevant design principles** (including a detailed cross-disciplinary review by computer scientists and legal scholars) that concern the research design of machine-learning applications,
B. **a set of case-studies** based on the project's microprojects, demonstrating how LPbD principles can be integrated into the architecture of AI systems,
C. **a dedicated assessment** of how these principles interact with HMI design, suggesting new research lines on the cusp of machine learning, HMI, and law.

1.3.1.6.2. "Ethics by design" for autonomous and collaborative, assistive AI systems

Methods will be investigated that aim to understand how values can be *wired* into sociotechnical systems and what it means to do so. These may include (but are not limited to) studies **in compliance, security, data protection and privacy by design, fairness, explainability**, and how to implement these in combination with AI techniques and algorithmic governance through formal analysis and representation of regulatory principles, allocating rights, distributing liability, and ensuring legal protection by design.

A core challenge concerns the shaping of AI technologies and ecosystems, comprising autonomous and collaborative, assistive technology in ways that express shared moral values and ethical and legal principles as expressed in (but not limited to) binding EU legal treatises. This involves understanding, developing, and evaluating reasoning abilities of artificial autonomous systems (such as artificial agents and robots).

Even though AI systems are such that they allow us and even encourage us to defer to humans for decision making and performing actions that have grave moral impact, AI systems are artefacts and therefore are neither ethically or legally responsible. Individual humans or human corporations should remain the moral (and legal) agent. We can delegate control to purely synthetic intelligent systems without delegating responsibility or liability to them. To this effect, computational and theoretical methods and tools will be investigated, that support the representation, evaluation, verification, and transparency of ethical deliberation by machines with the aim of supporting and informing human responsibility on shared tasks with those machines.

Research is needed to discern suitable constraints on system behavior, and to elicit desiderata on the representation and use of moral values by AI systems. Furthermore, we need to provide design principles for meaningful human control over autonomous AI systems. This includes, but is not limited to, **ensuring that privacy is respected, diversity is fostered in our communities, discrimination and biases are avoided**, societal and environmental well-being is respected, and basic rights and liberties are guaranteed.

An important topic is to boost research on developing tools for **discrimination and segregation discovery**, as well as **discovery and protection of novel vulnerabilities**. AI-based complex sociotechnical systems may amplify human biases present in data. Further, they may also introduce new forms of biases. As a result, AI-based systems may produce decisions or have impacts that are discriminatory or unfair, both under a legal or ethical perspective. Auditing AI-based systems is essential to discover cases of discrimination and to understand the reasons behind them and possible consequences (e.g., segregation). It may be that decisions informed by AI systems could have discriminatory effects, even in the absence of discriminatory intent. Moreover, discriminatory decisions take place on an individual in isolation, and **segregation** is the result of interactions among people in complex sociotechnical systems, nowadays largely governed by AI. Bias in AI systems can result in **both discrimination and forms of segregations.**

Explanation for high-stakes decision making. Decision making is essentially a sociotechnical system, where a decision maker interacts with various sources of information and decision-support tools, a process whose quality should be assessed in terms of the final, aggregated outcome—the quality of the decision—rather than assessing only the quality of the decision-support tool in isolation (e.g., in terms of its predictive accuracy and standalone precision). It is therefore important to develop **tools that explain their predictions in meaningful terms**, a property rarely matched[11] by AI systems available in the market today.

The explanation problem for a decision-support system can be understood as **"where" to place a boundary between what algorithmic details the decision maker can safely ignore and what meaningful information the decision maker should absolutely know to make an informed decision**. Thus, an explanation is intertwined with trustworthiness (what to safely ignore), comprehensibility (meaningfulness of the explanations), and accountability (humans keeping the ultimate responsibility for the decision).

In this context, several questions emerge: what are the most critical features for explanatory AI? Is there a general structure for explanatory AI? How does an AI system reach a specific decision, and based on what rationale or reasons does it do so? Explanations should favor a human-machine interaction via meaningful narratives expressed clearly and concisely through text and visualizations, or any other human-understandable format revealing *the why*, *why-not*, and *what-if*.
Following the same line of reasoning, the AI predictive tools that do not satisfy the explanation requirement should simply not be adopted in high-stakes decision making, also coherently with the GDPR's provisions concerning the "right of explanation" (see Articles 13(2)(f), 14(2)(g), and 15(1)(h) of the GDPR, which require data controllers to provide data subjects with information about "the existence of automated decision-making, including profiling and, at least in those cases, *meaningful information about the logic involved*, as well as the significance and envisaged consequences of such processing for the data subject.") The research challenges will intertwine greatly with the one of a "human-in-the-loop" line.

1.3.1.6.3. "Ethics in design"—methods and tools for responsibly developing AI systems

The real value of an AI system for decision support (e.g., based on machine learning, but not necessarily) is not in merely proposing an estimation on the probability that a certain relevant event will occur, or that the event is classified under a certain category, but requiring that guarantees are given that the system is developed and used in proper and verifiable ways.

This requires methods and tools for the **value-based design and development of AI systems** that ensure (a) the analysis and evaluation of ethical, legal, and societal implications; (b) the participation and integrity of all stakeholders as they research, design, construct, use, manage and dismantle AI systems; (c) the governance issues required to prevent misuse of these systems; and (d) means to inspect and validate the design and results of the system, such as formal verification, auditing, and monitoring [Durán2018].

Accountability. Accounting includes governance of the design, development, and deployment of algorithmic systems, which takes into consideration all stakeholders and interactions with sociotechnical systems. Mitigating includes introducing techniques for data collection, analysis, processing that incorporate and acknowledge the systemic bias and discrimination that may be present in datasets and models; formalizing fairness objectives based on notions from the social sciences, law, and humanistic studies; building sociotechnical systems that incorporate these insights to minimize harm on historically disadvantaged communities and empower them; and introducing methods for decision validation, correction, and participation in co-designing algorithmic systems.

The aim is to boost research on theories, methods, and tools for trustworthy AI approaches, including ethics by design and ethics in design. This will ensure that AI systems are developed in a responsible, verifiable, and transparent way, while guaranteeing that their behavior is aligned with human values and societal principles such as privacy, security, fairness, or well-being. Naturally, users' requirements, legal requirements, and ethical requirements change over time, which necessitates dynamic, continuous evaluation and feedback throughout the system's entire lifecycle, thereby allowing participants to adapt their systems to their ever-evolving requirements.

1.3.2. Methodology

1.3.2.1. Amassing Key Players from Academia and Industry

To achieve the aforementioned objectives, a comprehensive, multidisciplinary consortium is needed that goes beyond
core AI to include key relevant players from the HCI community, cognitive science, social sciences (philosophy, sociology, and law), and complexity science. Building on the HumanE AI preparatory action community, we have assembled such a consortium from Europe's most important research centers (INRIA, CNRS in France; CNR and FBK
in Italy; DFKI, Fraunhofer in Germany; IIIA-CSIC in Spain; Aalto, INESC TEC, Barcelona Supercomupting Center; and ATHENA in Greece), top universities (ETH Zürich, TU Wien, LMU Munich, UCL London, TU Berlin, U Pisa, U Umeå, VU Amsterdam, U Bologna, Charles U, and Brno U of Technology), and key industrial champions (Thales, Philips, Airbus, ING, Volkswagen, SAP, Generali, Telefonica, and Tilde).

1.3.2.2. Implementing the Research Agenda and Leveraging Synergies through Microprojects

The project will go beyond just networking, capacity building, and dissemination activities to actually implement key components of the proposed research agenda. To maximize impact within the available resources, we will focus on **critical gaps in knowledge and technology at the interstice between the competences** of the involved centers of excellence. This will be accomplished by organizing the research activities around the concept of "**collaborative**

micropojects." A collaborative microproject will involve a small group of researchers (2–5) from **different centers** of excellence working together at **a single location** for a limited period of time (1–6 months) to solve a problem related to a given gap. Microprojects will always have to produce a tangible result, such as a scientific publication, dataset, toolbox, demonstrator, or integration of a toolbox into the AI4EU. Microprojects will be situated within WPs devoted to different parts of the project agenda. Each WP will have dedicated funds for microprojects, which it will distribute through a lightweight internal proposal system based on quality and contribution to the WP agenda. Microprojects will be encouraged between WPs (with each WP contributing part of the funds) and will have the possibility of including external partners through appropriate mechanisms (see sections 1.3.2.3 and 3.1).

The concept of microprojects has multiple advantages.

1. It is well-known that assembling a group of researchers at a single location with no distractions but a project they care about is highly effective. Thus, providing several partners with 6 project managers (PMs) each to be used loosely collaborating on a 3-year project often produces little tangible results. On the other hand, if those 6 PM per partner are used to ensure that people from the respective groups spend a total of 6 months being together at a single location doing nothing else but working on a well-defined, joint project, then they can really accomplish something meaningful.

2. The collaborative aspect of microprojects—bringing together people from **different** centers of excellence—ensures that we focus on breakthroughs and developments that **leverage the synergies** between the competences of the individual centers and would not be possible without the project. It is an essential component in our vision of creating a "**multiplier effect**," where a relatively small investment represented by the microproject creates a much larger effect. Thus pieces of know-how distributed over different centers of excellence may have little impact individually, but may amount to a significant innovation/breakthrough, with a value far beyond the funds invested in the microproject that gathered them.

3. As researchers go back to their institutions after the microproject, they will bring the results back with them, making them part of their future research (e.g., PhD work), sharing them with colleagues, and using them in proposals. This is another component of the multiplier effect, as the knowledge will help progress on each site, shape further research at each site, and lead to new proposals, including national and industrially supported proposals.

Because of these considerations, overall more than half of the budget amounting is devoted to microprojects of various types. This represents a total of around 800 PMs, with each microproject having a scope of around 8–12 PMs. It is important to restate again that those 12 PMs will not be PMs devoted to some sort of loose collaboration of people at different sites, but will be 8–12 PMs representing a small group of researchers actually intensively working together at the same site, amassing knowledge from their respective centers of excellence, doing nothing but trying to solve a well-defined problem, and coming up with a tangible result. This is an extremely effective mode of collaboration and an activity that creates strong, sustainable links not just on institutional but also on personal levels. Another advantage of the approach is that it combines three core aims in one: (1) implementing the research agenda, (2) strengthening links between European stakeholders, and (3) spreading knowledge, **with the same funds furthering all three aims**.

1.3.2.3. Involving the Community outside the Consortium

An important concern of HumanE AI Net is to involve researchers from outside the consortium in developing and implementing the research agenda whenever they either have competences not present in the consortium or when such an involvement is seen as beneficial to disseminate knowledge to all of Europe's AI community, to ensure HumanE Al Net visibility, to create outreach for European talent, and for capacity building in general.

The mechanism that we decided to employ is to invite and finance the participation of researchers from outside the consortium in microprojects. Thus, an external researcher would be working on equal footing with researchers from different HumanE AI Net partners at the respective location for either the entire duration or just part of the microproject duration. HumanE AI Net would finance travel, subsistence, and other costs (if needed, salary, depending on formal requirements and regulations). As an alternative, a microproject may be hosted by an institution outside the consortium. For the project's purpose, such an approach has a number of advantages over a more conventional "Open Call" method of involving external partners. Open calls are a good method of using project funds to direct external resources toward certain aspect of the research agenda, bringing strong consortia into a loose collaboration with the project. However, open calls have a number of disadvantages:

- They involve significant overhead involved in the formal process, which means that they are not agile; not only in consuming resources for the process, but also being fairly slow (not too many open calls would be possible in a 3-year project).
- The cooperation between the external open call participants and the consortium is not automatically very close. Each of the external projects is a consortium on its own, that may or may not intensively interact with the original project (beyond formal requirements). The contribution to spreading knowledge from within the consortium is limited.
- The method favors large external organizations that can address such calls with no way of involving excellent individuals (in particular, young researchers) who may happen to be part of smaller universities and organizations.

By contrast, the proposed method of involving external researchers as participants in microprojects focusing on financing travel and subsistence has the following advantages:

- The formal process is extremely lightweight, with each cooperation consuming a smaller amount of resources. This means that a considerable number of corporations is possible (we estimate 50–100, given the funds set aside for the collaboration) and we can react quickly to invite promising researchers.
- Having a researcher who spent a few months with a group from our consortium is a highly intensive form of collaboration, leading to a strong knowledge transfer and building of links not just on institutional but also on personal levels. Especially for young researchers, such network building is extremely valuable.
- The methods lets us focus on individual excellence, reaching out to researchers not only at large institutions but to reach out to talent no matter where they are in Europe, and to help the talent develop and encourage talented researchers to continue their careers in Europe.

1.3.2.4. Transforming Research and Innovation into an Economic Impact and Value for Society

To foster human-centric AI and maintain Europe as a powerhouse in the key technology shaping the global economy, it is crucial to maximize the socioeconomic impact of the consortium's research roadmap. The key approach is application-driven sustained innovation to generate societal impact clearly perceptible to European citizens. Transforming results of basic and applied research into economic strength, useful products and services, and new venture is at the core of our research strategy, including mechanisms for start-up creation and means for strong innovation in existing businesses. The following mechanisms are exemplary for activities, to ensure the tight integration between all stakeholders and to ensure societal and economic relevance.

Agenda workshops. Agenda workshops for each domain will bring together representatives of the respective industrial champions, their customers, and researchers from relevant WPs of the project. They will on one hand ensure that researchers understand the needs of the industry. Simultaneously, they will help industry advance beyond incremental improvement over established solutions and identify potentially disruptive, novel approaches. The workshops will produce domain-specific R&D agendas that will be the basis for industrial microprojects.

Industry-driven microprojects. Industry-driven microprojects will be conducted by WP 6 on the basis of the above R&D agendas and the research results of WPs 1–5. Where appropriate, other WPs and external partners will be involved. The industrial microprojects will be key means for inspiring basic research to look into real-world challenges but also for transferring the results of basic research into industry and evaluating the results in use cases that are economically as well as societally relevant.

Industrial PhD studies. An industrial PhD and postdoc program will create further close links between the academic centers of excellence and European industry. As a coordinated program across the entire HumanE-AI network, it will provide industrial PhDs the opportunity to consult Europe's leading experts in their specific area, to pick the best academic supervisors for their work, to interact with peers across the continent, and to access all necessary infrastructure. This program will collaborate with national and local initiatives in AI, launched in most of the countries of this consortium. Most of the partners enroll at least 20 students per year in AI doctoral programs.

The research agenda is addressing challenges with high societal relevance, but also with significant economic perspective. The microprojects (MPs), that are key to fostering collaboration, simultaneously will produce results ranging from fundamental research insights to prototypes, software components and frameworks, tools, realistic and proven use cases, as well as scientifically evaluated and sound business ideas. The project will offer for promising MPs suitable opportunities to pursue the ideas further along the innovation funnel, supporting the researchers, inventors, and developers in transforming results into real economic and societal value. With the innovation funnel, we simplify the steps from research to business (see Figure 2). In a step-by-step approach, researchers receive the structured support for turning their research into economic and societal value with appropriate partners.

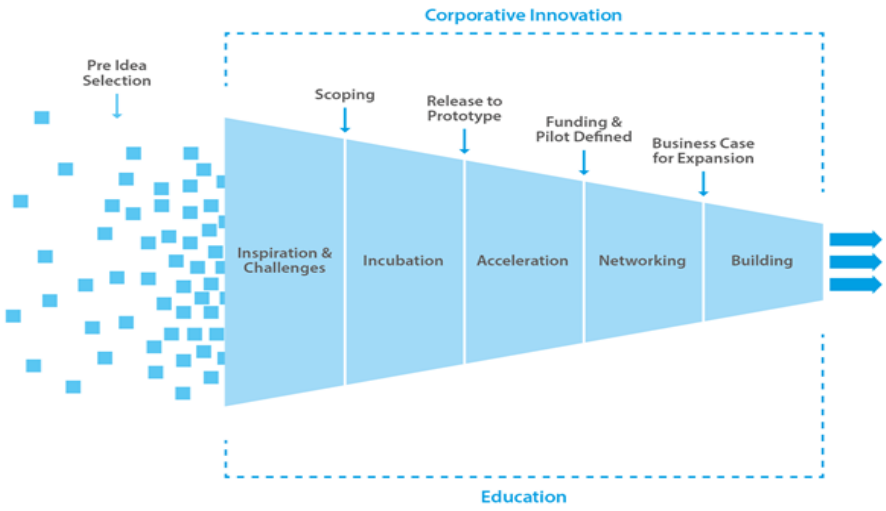


Figure 2. Innovation funnel: a structured approach to move from basic research to economic and societal value.

The innovation strategy is comprised of four main objectives:

- Unite the research and innovation community through a **platform by matching researchers and their re-search results with appropri-ate support structures and partners,** taking their research a step further into successful application and implementation.
- Catalog and **involve the best existing support structures and formats in Europe**, from incubators to innovation units in industries, and create an attractive environment for all to use the platform.
- **Design and implement new support formats and structures filling the existing gaps** in the research and innovation ecosystem.
- Use and **develop an exhaustive innovation infrastructure, taking societal impact and involvement into consideration** for the three main target groups: **(1) industry, (2) SMEs, and (3) startups**, with a clear focus on creating societal values.

1.3.2.4.1. Cooperating with AI4EU

The HumanE AI consortium has substantial connections to the AI4EU project. Key people within the HumaneE AI Net consortium also play key roles in the AI4EU project. The coordinator, Prof. Paul Lukowicz, is an active member of the AI4EU consortium. Profs. Barry O'Sullivan, Jim Crawley, Virginia Dignum, Micheala Milano, Andrejs Vasiljevs (from Tilde, is a member of the AI4EU Industrial Committee), Prof. Alessandro Saffiotti (leading Task 8.3 on integration infrastructure for scientific collaboration in the AU4EU platform) leads WP 7 in the AI4EU project, and Joachim Köher (leading tasks on creating the Virtual Laboratory) also is involved in the AI4EU implementation.

On an operational level, synergy between the HumaneE AI Net and the AI4EU platform will be achieved through the following measures:

1. Task 9.1 is devoted to cooperation with AI4EU. The task is led by Thales, specifically Patrick Gatellier who is the coordinator of the AI4EU project. This task has a dedicated budget of 18 PMs to coordinate and support contributions to the platform and facilitate its broad use within the project.
2. Task 8.3 is dedicated to the implementation of infrastructure for scientific data collaboration and running scientific challenges and benchmarks within the AI4EU plattform.
3. A key contribution of the project to the community will be benchmarks, including corresponding datasets and evaluation scripts. These benchmarks will be made available through the AI4EU platform.
4. Within the educational activities (see T8.5 and 8.6) such as summer schools, tutorials, and workshops, we will include sessions devoted to the platform. Respective materials will be developed in cooperation with the AI4Eu project within task 9.1.
5. The Virtual Lab will include an interface for the platform and coordinate with AI4EU through tasks 8.1, 8.3, and 9.1. As section 2.2.1.3 describes, the Virtual Laboratory will be usable and available as a resource from within the AI4EU platform. It will also link to additional resources.
6. The knowledge-spreading activities (T8.5 and 8.6) will be disseminated through the platform.
7. The research WPs (1–6) all will deposit the results of their microprojects on the AI4Eu platform as deliverables (D1.1, D2.1, D3.1, and D4.1). Thus, all project research results will be available through the platform within the scientific collaboration component that we will build for AI4EU.

Overall, around 500,000 euro will be devoted directly to platform-related work.

1.3.2.5. Cooperating with Digital Innovation Hubs

HumanE AI will establish a close contact and cooperation with DIHs and DIH network projects. The objective is to leverage the (regional) DIHs and (cross-border) DIH networks as mediators that enable industry, SMEs, and startups throughout Europe to develop innovation using the algorithms, data, tools, and support services provided through HumanE AI and its members.

HumanE AI will approach currently active as well as future DIH-related activities. Among the currently active networks, in particular the DIH network-cluster on Robotics and SAE (SmartAnythingEverywhere) will be targeted. This cluster is coordinated and supported by the RODIN CSA and consists of the projects DIH2, Trinity, DIH-HERO, agROBOfood, and RIMA, which cover the application areas' industrial production, healthcare, agriculture, and inspection and maintenance.

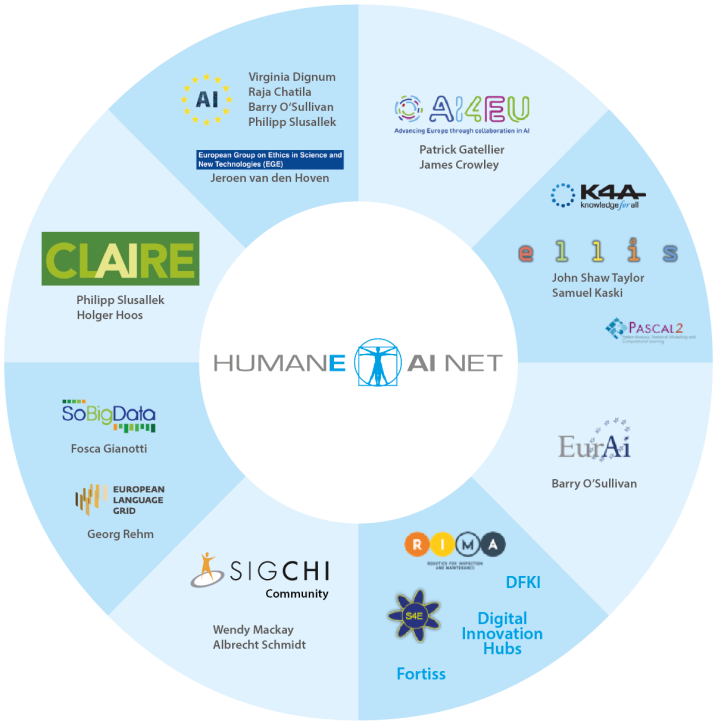


Figure 3. The HumanE AI Net embedding within the AI community.

A close connection between HumaneAI and the four DIH networks on robotics will be established through HumaneAI beneficiaries that are directly involved in the DIH projects and related organizations (such as DFKI, who is both a member of the[12] RIMA consortium and the association of the European robotics community euRobotics). HumanE AI will interface with the RODIN CSA to organize joint events, such as workshops and plenary discussions at the "European Robotics Forum ERF," an annual event organized by the European PPP on robotics SPARC (jointly run by the European Commission and euRobotics). The objective here is to bridge the still existing gap between the AI and the robotics communities and to stimulate the uptake of AI based concepts in robotics research and applications.

HumanE AI will also cooperate with other already established networks, such as I4MS, and the AI DIH network, as well as the planned DIH networks in the areas of Big Data and smart hospitals.

1.3.2.6. Cooperating with Other Relevant European and National Initiatives

The consortium is deeply connected to all relevant national and European initiatives, and will pursue an active cooperation and collaboration strategy through the corresponding tasks in WP 9. This WP is led by Barry O'Sullivan who is President of EurAi and who will coordinate the networking activities with and though EurAi. Initiatives that we will specifically address include, among others: Language-centric AI (ELG, META-NET), SoBigData, the AIBig Data and Robotics[13] PPP initiative that is currently being planned, CLAIRE, and ELLIS. For each of those, there is a dedicated task on WP 9 lead by individuals who are also important players in the respective initiative.

1.3.2.7. Disseminating Knowledge to All AI Labs within and beyond the Consortium

A detailed dissemination plan is provided in section 2.2 and is the core concern of WP 8. The core ideas toward dispersing the knowledge to all European AI Labs (and beyond pure AI toward all relevant scientific disciplines) can be summarized as follows. First and foremost, the consortium aims to produce high-quality, high- impact publications. As stated under 1.3.2.2 the microprojects must produce tangible results. Especially the ones focused on more basic research will have publications as their main KPI. Beyond individual scientific articles, we will publish edited article collections and a Handbook of Human Centric AI as a synopsis of the most important project results and guide to the community.mOther concrete deliverables for microprojects will be tools and datasets. Those will be provided to the community through the Ai4EU platform, the Virtual Laboratory (see below), and other channels typically used by the respective communities. Building on the tools and datasets, we will run challenges and competitions. These will be setup by the respective WPs and run within the Virtual Laboratory (WP 8).

Going beyond one-time competitions, we will create a benchmarking and challenge infrastructure for project-related AI methods that will be setup within the AI4EU platform. It will combine curated datasets, algorithms, and methods for results analysis. It will also allow researchers to add datasets and algorithms and compare their method to the ones stored in the repository. In doing so, we will build on existing benchmarks (e.g., in computer vision) and partners' expertise and tools for setting them up.A final pillar of our knowledge-spreading approach will be direct

educational activities (see section 2.2). These will include summer schools, tutorials, and workshops (Task 8.5). We also will produce dedicated MOOCs and online materials from the summer school, tutorial, and so forth. Presentations (T8.6) will be available through the Virtual Laboratory and AI4EU platform.

1.3.2.8. Industrial PhD, Postdoctoral, and Internship Programs

The PhD program will be run by Task 8.4, with the coordinator of WP 8 heavily leveraging the Virtual Laboratory (T8.1 and 8.2), the summer schools, tutorials (T8.5 and 8.6), and the math platfofm and events from the innovation WP (T7.2 and 7.7). The ambition of the HumanE AI Net PhD Program is to establish a training agenda to improve the education of a new generation of creative researchers and innovators, knowledgeable and skilled in AI. HumanE AI Net aims to provide excellent research and training opportunities to attract, develop, and retain talented PhD students with both a background in core AI and related domains such as social sciences or neuroscience.

The complementary research skills and training expertise within this PhD Program will transform the way each network partner is working, to foray into the necessary steps and changes in developing AI-related sciences and technologies needed for businesses and society in domains shaping the 21st century.

The objectives in terms of education and training are (i) to foster interactions between PhD students, researchers, and innovators both in academics and industry in Europe, and (ii) to define new specific, attractive, and reference curricula to form a new generation of graduate students.

The network will particularly focus on industrial challenges by offering to PhD students and postdocs the opportunity to build industry-guided showcases that integrate AI's state-of-the-art models and methods. Industrial partners' participation will support the translation of new academic results to the marketplace and a better transfer of knowledge between different sectors. These will ensure that PhD students learn how to conduct research in an industrial context and force them to think of their research project in terms of real products, leading to increased innovation outputs.

The exposure of the nonacademic sector to the PhD students has a great market potential for our industrial partners, who hope to capitalize by recruiting young talent. This is mutually beneficial to the PhD students, who will be provided with new career perspectives in AI-related industries. The HumanE AI Net PhD Program will bolster Europe's capacity in research and innovation by nurturing a new generation of highly-skilled PhD students with an entrepreneurial mindset and an understanding of AI and potential products in these emerging markets.

Implementing the HumanE AI Net PhD Program focuses on the following actions:

- *Collaborative microprojects* entail enrolling young researchers in short-duration HumanE AI Net microprojects (1-6 months), with concrete objectives such as a scientific publication, data collection, or training in specific topics. The research opportunities offered by HumanE AI Net span a girth of theoretical and application topics in key European research laboratories. PhD students and postdocs also will be afforded the opportunity to propose, lead, and manage (including budgets) collaborative microprojects.
- *Cross-sectoral secondment* involves matching PhD students and postdocs with European industry players, to gain practical experience and benefit from in-depth knowledge of AI applications in real-world situations. Cross-sectoral secondments are proposed by industrial partners.
- *Designing HumanE AI Net curriculum* for international students will need to align with the rapidly changing character of research and innovation. University-industry partnerships will help identify needs for new training, learning outcomes, ensure the relevance of software and equipment, and adequacy to regulations and laws. Such partnerships also define real-life workplace scenarios and industry-guided showcases with the aim of producing a recognized European curriculum that could be used to derive local curricula.
- *Connecting academic content to state-of-the-art results and real-world experiences* will transpire through the Virtual Laboratory. This will provide young researchers with unique access to knowledge, resources, and shared facilities by being the interface to the AI4EU platform. Young researchers will have the opportunity to access, produce, and share code snippets and benchmark datasets.S

1.4. Ambition

HumanE AI Net will strengthen active research areas that pose hard scientific challenges. These include integration of machine learning with cognitive processes, improving integration and learning for multimodal perception, action, and natural interaction with humans [Cornelio2017] [Schmidt&Hermann2017]. We will bring together the best researchers and practitioners in Europe in specific areas and create incentives for collaboration. We will also create an environment where expertise in specific areas can be jointly built up with the aim of creating teams of world-leading experts on several AI technologies.

Explanation is at the heart of a responsible, human-centered AI. Despite recent progress on interpretable machine learning (see [Guidotti et al. 2018] for a comprehensive recent survey) and on discrimination-aware data mining [Pedreschi et al. 2008, Ruggieri et al. 2010], a practical, widely applicable technology for explainable AI has not yet emerged. The challenge is hard, as explanations should be sound and complete in statistical and causal terms, yet based on shared models between humans and machines in order to be comprehensible and credible for humans with different levels of expertise. Humans must be able to understand and reason about how automated decision-making works, how a specific decision is taken and on the basis of what rationale/reasons and how the user could get a better decision in the future. Open questions include how to exploit recent advances in generative adversarial networks [Goodfellow 2014], deep reinforcement learning [Mnih 2016] in the context of systems that can explain and justify actions and decisions and the use of combined statistical and symbolic representations to facilitate shared models between humans and machines. The need to invest in such research also aligns with the EU data protection framework that requires meaningful information when people are targeted by automated decisions with a significant impact on their lives [Hildebrandt2018][14].

Joint research at the boundaries. As core AI technologies evolve, exciting opportunities will arise at the boundaries between those technologies and their application fields. This is where we can overcome boundaries in a specific layer and where we expect exciting advances to come from sharing technologies and approaches across specific fields. Research that involves stakeholders in the real world is key to a deeper understanding [Rogers2017]. One example is to combine research in language understanding with computer vision and robotics to jointly develop novel approaches and algorithms that benefit from insights realized in all the different areas involved. Similarly, AI4EU, the recent effort to create an AI-on-Demand platform will become a catalyst for joint investigations and for creating actionable results that go beyond innovation and allow for AI to make major leaps forward.

The name of the project, HumanE AI Net, describes the most complex area of investigation: **making AI valuable for humanity, advancing AI for the socialgood**. This aim is the driving force for bringing together researchers and practitioners on this project. AI must not be a threat to future human development or livelihood. With our human-centered approach, we expect major advances in the state of the art because of two points: aiding human development guides and inspires the search for novel applications services while also giving a clear metric for evaluation beyond purely technical advances. IAI should be accessible to individuals as well as society.

Ourresearch will develop new means for interaction and collaboration between humans and intelligent agents, creating AI systems that collaborate with people rather than replace them. For such systems, technology acceptance is a key factor [Hornbæk2017a], and the interaction with technology has to be fundamentally investigated [Hornbæk2017b]. Next to **technology acceptance**, legal and ethical requirements highlight that such systems must also be **acceptable** from a fundamental rights and from a moral perspective [Hildebrandt2020][15]. Advances at the boundaries between artificial and natural intelligence are expected, leading to combined teams of artificial agents and human actors that will jointly have strongly enhanced cognitive, perceptual and physical capabilities (see [Schmidt2017a] and [Schmidt2017a[16]][17]). Considering the wide range of technologies, there will not be a single right way of implanting AI in most application areas and for many use cases, e.g., policies based on a specific societal discourse could limit or prescribe what the AI is optimized for. However, there might not be a consensus in other areas, leaving it to the developer or the user to make decisions. To do this, systems must allow for parameter selection in a potentially vast space and a means to communicate what they are optimized for. This will be a major challenge, but it will move systems forward to a new level of understandability, inspectability and auditability (which will often be a legal requirement). Advances to the state of the art will be related to specific technological results, methods and approaches, as well as with regard to societal benefits.

We believe that for truly interactive AI, it is necessary to not only build interfaces for them, but to revisit the core methodology in machine intelligence. This goes beyond state-of-the-art by building a new capability for AI to *understand* human partners. This, we believe is the basis of fluent interaction and has been neglected in present day ML. Researcher in Humane AI Net will seek a new foundation from theories of human behavior, combining them with modern machine-learning methods (e.g., probabilistic methods, artificial neural networks) in such ways that allow more transparency, control, and first and foremost more natural, human-like collaboration. In particular, we believe that for AI to function in collaborative settings with human communication partners, models and theories from computational psychology [Sun2008], computational cognitive sciences [Kriegeskorte2018] and computational social sciences [Lazer2009] are needed, along with natural language processing theories, models, algorithms and systems (as described in the previous section). This calls for plausible models of human behavior that can — thanks to *causal models* that link behavior with cognitive, emotional, and other latent factors — better infer, plan, and act without extensive data on an individual [Lake2017]. This approach calls for virtual training with simulated humans (e.g., similar to AlphaGo Zero) as opposed to passive datasets. It also calls for verification of model predictions against human behavior in live interactive settings, as opposed to cross-validation within passive datasets. The corresponding joint challenge with HCI research is to construct efficient and appropriate ways for communicating

and expressing internal states between the AI and the human partner. By virtue of being partly based on human-understandable models of interaction, this goal is more plausible than with the predominant data-driven approach.

Beyond the individual the striking emergence and popularization of AI raised ample discussions about how AI may be designed to be safe and beneficial for society [Baum2017, Floridi2018]. A crucial point is understanding and modeling the way AI generates or reinforces certain human behaviors and emerging societal phenomena, considering that mutual influences among individuals underlay the formation of public opinion [Nowak1990], group decisions, and actions [DeDreu2008]. To this purpose, our ambition is to start from the research on human groups [Petty1986, Mullen1994, Nowak2020, Nowak2017] to design AI members of socio-technical groups. The challenge is how to design AI agents that can improve the functioning of the group in reaching optimal decisions and judgments. AI agents need to understand their role in distributed information processing social systems, be aware of the competence and reliability of group members, the importance of the issue at hand, and their limitations. Another challenge is to design the rules by which AI agents decide which information to process themselves and which to delegate to humans, and who in the group is most capable of processing which information. In other words, the challenge is to design the rules that allow AI to put the human in the loop in a way that is most efficient for techno-social systems. We intend to go beyond current research designing, based on the Regulatory Theory of Social Influence [Nowak2020], mechanisms that enable AI agents to estimate, as naturally as humans, the trustworthiness of both humans and other AI members of the group. In this way, AI agents can use trust estimates as a guide for optimal information processing in social groups, where the most reliable and competent in the topic group members process the most critical information.

AI-based complex socio-technical systems may amplify human biases present in data. Further, they may also introduce new forms of biases. As a result, AI-based systems may produce decisions or have impacts that are discriminatory or unfair, both under a legal perspective or an ethical perspective. Social discrimination is considered illegal and unethical in the modern world. Auditing AI-based systems is essential to discover cases of discrimination and to understand the reasons behind them and possible consequences (e.g., segregation). It may happen that decisions informed by AI systems could have discriminatory effects, even in the absence of discriminatory intent. The objective of equity can be achieved by embedding the fairness value in the design of such systems (Fairness-by-design) and by upholding that value (justice). On this issue there is a recent flourishing of literature with more than 20 different definitions of fairness [Becker2019, Stewart2019] aimed at quantifying different notions of bias; disparate treatment, disparate impact and disparate mistreatment, among the others. Approaches for embedding fairness in AI have been proposed mainly for machine-learning (ML) classification models acting on i) preprocessing methods focusing on the data, ii) in-processing methods focusing on the ML algorithm, and iii) post-processing methods focusing on the ML model. [McMahan2017] proposes a two-stage decision making that shows that imposing fairness at intermediate stage comes at a cost but that cost can be reduced by hiding the sensitive feature at the first stage of the decision. Our ambition is to leverage the planned advances in learning, reasoning and planning with human in the loop (see 1.3.1.2 and WP 1) combined with the ability of systems to understand humans and complex social settings and incrementally build up respective subtle world models (see 1.3.1.3and WP 2) to come up with novel more effective approaches to combating bias and unfairness in AI systems. New interaction methodologies (see 1.3.1.4, WP 3) will allow people to flexibly reason with AI systems about such problems and attempt to jointly mitigate them. This notion of **true collaboration and co-creation between humans and AI is**[18] **at the** core of our ambition with respect to truly human-centric, empowerment oriented AI. This also involves taking into account competing theories on the difference between human and machine learning, making sure that the 'machines-in-the-loop' do not get the last word on how to interpret human action [Hildebrandt[19]2017].

A related core ambition of the project is to boost research on theories, methods and tools for trustworthy AI approaches, including ethics by design, and ethics in design. This will ensure that AI systems are developed in a responsible, verifiable and transparent way, while ensuring that their behavior is aligned with human values and societal principles such as privacy, security, fairness, or wellbeing. Naturally, user, legal and ethical requirements change over time, which leads to a dynamic, continuous evaluation and feedback throughout design and operation [Dignum, 2019], thus allowing participants to adapt their systems to the ever evolving requirements. In this context it is a core concern to ensure that the design and use of AI are aligned with ethical principles and human values, taking into account the societal context while enabling their human users to act ethically and respecting their autonomy and self-determination. New-generation AI systems must be "under the Rule of Law," i.e., their design, operations and output should be contestable by those affected by their decisions, liability for those who put them on the market.

Ambition in Innovation

In order to foster human-centric AI and maintain Europe as a powerhouse in the key technologies shaping the global economy, it is the aim of the HumanE AI Net to maximize the socio-economic impact of the research roadmap. Therefore the research agenda must be most relevant to solving current challenges in European society and economy. To generate societal impact clearly perceptible to European citizens, fostering application driven innovation is key. Therefore, the aim is to intensively support the transformation from the results of the research agenda to start-up

creation and innovation in existing businesses, organisations and industry via dedicated mechanisms and close collaboration with key industry players and key innovation drivers.

2. Impact

2.1. Expected Impacts

2.1.1. Contributions to Expected Impacts of the Call

Throughout Europe there is consensus that we need to establish a European brand of AI that is human centric, with trust, ethics, European values and the benefit of European citizens and society at its focus. There is also consensus[20] that this cannot be achieved through regulation that hinders innovation, but must be a motor of innovation driving European researchers to develop new unique solutions—which therefore will also have a unique potential to spin-out innovation from the project and thus generating increased economic activity.

A key challenge is that such solutions can not be found by working within traditional AI silos but instead require breakthroughs at the interfaces of various areas of AI, HCI, cognitive science, social science, complex systems, etc. HumanE AI Net brings together a **unique community** which has the expertise both within these silos and at the interfaces between them, and can address those challenges. Building on the HumanE AI FET CSA this community already has a vision where the key gaps in knowledge are and what needs to be done to address them. This includes a key focus on AI that is compatible with the EU fundamental rights framework and dedicated to integrating legal protection at the level of research design, thus enabling the industry to develop AI applications that are robust, trustworthy and ethically sound.

Thus **a key impact will be to significantly advance the science closing the gaps needed to make the vision of a human-centric, European brand of AI a reality.**

The proposal not just involves key European industry partners but has a setup where (within WP 6) the industry can on one hand influence the research agenda and evaluate the usefulness of the results in relevant use cases and on the other hand learn and be inspired by the research visions through stakeholder workshops. In addition we have a WP (WP 7) devoted to innovation strategy that will reach out to industry beyond the consortium and facilitate the creation of spin-outs and startups. We will also engage external industry through the ability to include external researchers in microprojects.

Thus an important impact of the project will be not just to advance science but to do it in a way that is **synchronized with the needs of the European industry** and will contribute to strengthening it including **strong potential for spin-out and startup creations** leading to **increased economic activity.**

The proposal integrates seminal work in the realm of both fundamental rights and ethics, paying keen attention to AI applications that enhance rather than diminish human agency, based on a solid understanding of the interaction between research design and practical application. This will involve both **legal protection by design** and **value sensitive design** approaches into the heart of the project.

Thus a further **key impact will be holistic design practices ensuring our vision of human-centric AI can be integrated within a wide range of AI-related innovation in a reliable and trustworthy manner.**

The concepts of cooperating within and beyond the consortium through collaborative micro projects (of which we envision to have over 50) will create very strong links between the involved groups including strong personal networks. At the same time having WP 9 explore and foster connections to all relevant AI related initiatives and groups will create the corresponding institutional links and networks.

Together the impact will be to **create a vibrant, sustainable community** around vision of human-centric AI the associated science and applications that will **overcome the current fragmentation of the AI landscape** focusing it on a unique European brand of AI.

The concept microprojects - which allows to include external researchers - will also allow us to reach out to top European talent no matter where the respective people are. The industrial Ph.D. and postdoc program will also increase the attractiveness of working in and for Europe.

Thus the HumanE AI Net project will have significant impact on **capacity building in AI** in particular in **helping to keep young talented researchers in Europe.**

By putting a strong emphasis on making all results available through the Virtual Laboratory (T8.1 and 8.2) which is seamlessly integrated with the European AI on demand platform AI4EU (T8.3,T9.1), we will ensure efficient and

broad dissemination of the results. This will be further enhanced through the summer schools, workshops and MOOCs.

This will ensure **efficient spreading of advanced knowledge related to human-centric AI to all AI Labs** and researchers in Europe providing easy seamless access and fostering cooperation.

This project will enhance the AI4EU platform with the infrastructure needed for research-oriented collaboration, adding mechanisms for using it to run challenges, depositing the entire body of knowledge that the project will create on the platform, and by adding the Virtual Laboratory as a resource. This then becomes available for the entire platform community to use - as well as for users beyond the platform.

Therefore, the HumanE AI Net project will have the impact of **strengthening the AI-on-Demand-platform and enriching its capacity** in terms of tools, competencies, services, data and making it the reference and quality label for resources in AI.

Putting all the individual contributions of the HumanE AI Net project together, it is clear that it will significantly contribute to reinforce Europe's capacity and progress in AI, boosting the research capacity in Europe and the status of Europe as a research powerhouse for AI, especially in the particular brand of human-centric, value-oriented AI directed at empowering European Citizens and society.

2.1.2. Key Benefits to European Economy and Society

AI has progressed to the point where it is an essential component in nearly all sectors of today's modern economy, with a significant impact on nearly all areas of our private, social, and political lives. At the same time, the technology is still in nascence, with many fundamental grand challenges remaining to be solved and a lack of workable solutions for large-scale application to key socioeconomic problems.

Europe has a strong tradition in AI research, and could be poised to lead the next wave of the AI revolution. Unfortunately, there are many signs that Europe is starting to lose the AI race. This is not the first time that this has happened: there are many similar, past examples in which Europe started from a position of strength in an emerging area, failed to follow through, and eventually lost competitive and economic benefits. Companies like Nixdorf were global leaders in the early days of computing; Philips, Nokia, Siemens, and Alcatel (to name just a few) initially dominated the European mobile phone market; the World Wide Web, invented at CERN, took off in America and Facebook, Google, and Amazon reap most of its benefits. In these cases, key innovations were made in Europe but were productized [21]somewhere else.

A key aim of the HumanE AI Net project is to ensure that the same does not happen with AI. We believe that while there may be aspects of AI that established big players and other countries may be ahead, a specific European brand of AI that focuses on the human, interaction with humans, and the impact on society (as proposed by HumanE AI Net) can provide Europe with a unique competitive advantage and make a significant contribution to the well-being and prosperity of European society. While in the early stages of digitization, the focus of value creation was on improving connectivity, sensing, and interoperability, now value creation increasingly is dominated by AI. Already, we are seeing the rate of AI's penetration in real-world applications being limited by most systems' ability to adequately interact with humans, by related issues of user acceptance, and by the capability of dealing with complex, dynamic, unpredictable environments. These are precisely the issues that HumanE AI Net aims to address in developing the next generation of European AI technology.

Whoever leads the way in future generations of AI technology will set the standards for the values and ethics embedded in it. AI's development to date has brought us to an important crossroads: depending on the direction of future development, AI could destroy or create new jobs; empower citizens or impair their autonomy and privacy; increase the complexity of a globalized, interconnected world, thereby increasing the risk of systemic failures, or provide us with transparent, usable, understandable tools for taming that complexity. The HumanE AI Net project aims to embed such considerations in all stages of development by working closely with social scientists, ethics experts, political decision makers, and society[22]. It will go beyond classical notions of HCI, developing new ways for making AI empower and enhance human capabilities and building systems aware of and sensitive to social and ethical concerns. As a point of departure, HumanE AI will ensure that its high quality output is compatible with the European framework of fundamental rights, providing legal protection by design. As this will build the relevant checks and balances into the computational architectures that run our everyday lives, such legal protection by design enables a human-focused, reliable, and ethically sound ICT infrastructure.

Our society currently faces tremendous challenges in many areas, including but not limited to climate change, global resource constraints, erosion of confidence in the democratic process, polarization of public opinion, and the risk of cascading effects causing systemic failures in, e.g., financial systems, energy grids, or societal cohesion. Clearly, none of these problems can be solved through technology alone. However, HumanE AI Net can provide tools that

will make solutions easier to find, implement, and sustain. When resources are scarce, then flexible, intelligent, dynamically adaptive usage is crucial.

2.2. Measures to Maximize Impact

Given the nature of the call, exploitation and dissemination measures are given particular importance in the HumanE AI Net project. Of the 9 WPs (excluding management) **3 (WPs 7,8,9) are exclusively devoted to dissemination, exploitation, and impact** while one (WP 6) is meant as a bridge between exploitation/dissemination and advancing the research agenda.

2.2.1. Dissemination and exploitation of results

2.2.1.1. Standard Scientific Dissemination and Exploitation

Scientific publications: Scientific publications will be among the most important results of the microprojects and are a key KPI, as described in the objectives section. Given the highly multidisciplinary nature of the consortium, publications will take place in various communities, which is a strength of HumanE AI Net in terms of knowledge spreading:

- **AI:** AAAI, IJCAI**,** NIPS,ICML
- **HCI :** CHI, IUI, Inf. Visualisation, TVCG, CG&A, EuroVis, VAST, IEEE InfoVis
- **Ubiquitous computing:** UBICOMP, Pervasive ISWC, MOBISYS
- **Language processing and text mining:** (E)ACL, COLING, CONLL, EMNLP, CIKM,
- **Computer vision:** CVPR, ICCV, ECCV
- **Online social networks:** WWW, SocialCom, ACM SIGCOMM, IEEE INFOCOM, ´
- **Social media and social informatics:** ICWSM, WSDM, CIKM, SIGIR,
- **Data mining and pattern discovery:** CIKM, IEEE-ICDM, ACM-SIGKDD, PKDD, SDM

Datasets and benchmarks: Where perception and interaction are concerned, datasets and benchmarks are central to the community. As described throughout the proposal, they will be another important aspect of microprojects' concrete results. Task T2.7 is directly devoted to collecting datasets and benchmarks for perception, Task 8.3 will build the infrastructure to store and distribute them through the AI4EU platform. Summer schools and tutorials (see below and T8.5) will help establish them within the community.

Keynotes and invited talks: Members of the consortium are distinguished scientists often invited for keynotes at conferences and talks at academic and industrial research labs. They will use such opportunities to increase the visibility of HumanE AI Net and spread key project findings.

2.2.1.2. Specific HumanE AI Net Dissemination

Given the project's strong focus on dispersing knowledge to all European AI labs (and related communities), multiple measures that go beyond standard dissemination "best practices" will be implemented.

Summer/winter schools (T8.5) are larger events with up to one week duration and up to 50 participants devoted to on overview of a broader topic. Such events are not only a source of knowledge, but are also an excellent networking opportunity, in particular an opportunity for young researchers to expand their professional networks and interact with senior figures from the field (who are typically the lecturers at such events). We will establish an annual summer school on "human-centric AI," which will focus on lowering the barrier for entering the field for potential PhD students, postdocs, and industrial researchers. The lecturers will be senior figures from the consortium and leading international figures in relevant areas from outside the consortium when needed. We aim to establish the summer school as an institution within the field that persists beyond the project.

Tutorials (T8.5) will be more focused events dedicated to specific techniques, tools, or datasets. We will conduct two types of tutorials. The first will be directed predominantly at industry, especially SMIs, introducing basic techniques to lower the entry barrier toward human-centric AI. The second will be tutorials about the tools and datasets developed by the project. These will mostly target the scientific community, with the aim of facilitating a broad takeup of the project results within the European AI community. We aim to conduct of least two of each type per year. In addition, tutorials will be recorded and **made available online** through the Virtual Laboratory and AI4EU platform**.**

Workshops (T8.5) will be organized during major conferences such as AAAI, CHI, or UbiComp to reach out to the broader community. When possible, we will co-organize with other relevant projects and AI4EU. Workshops will focus on in-depth scientific presentations and discussions and will contribute to both shaping and disseminating the project research agenda. We will have at least two workshops per year: one at a core AI conference such as AAAI,

and one targeting a related community such as computer vision, robotics, ubiquitous computing, HCI, and data science. As a start, already in 2020, a workshop on human-centered AI will be held at Dagstuhl, organized by consortium partners, aiming to discuss the needed scientific and technological foundations for designing and deploying AI systems that work in partnership with human beings, to enhance human capabilities rather than replace human intelligence. Another relevant Dagstuhl workshop, on "Universals of Linguistic Idiosyncrasy in Multilingual Computational Linguistics," will take place in June 2020, attended by multiple project partners.

MOOCs (T8.6). In addition to the recordings of the tutorials, we will produce dedicated online courses covering all core aspects of human-centric AI. These will in particular target the industrial audience.

Handbook of Human Centric AI. As a further effort to support education in human-centric AI, we will publish a handbook that systematically introduces key concepts and techniques in a way suitable for a variety of courses and self-learning. It will align closely with the MOOCs.

2.2.1.3. Virtual Laboratory and Use of the AI4EU AI On-Demand Platform

A key goal of the HumanE AI Net project is to enable various target groups within academia and industry to quickly profit from AI technologies and knowledge created by the project (and various stakeholders around the project) and bring them to fruition. To this end, the Virtual Laboratory will provide **a single access point to all relevant HumanE AI resources** within an easily usable digital networking and collaboration environment.

The goal of the HumanE AI Virtual Laboratory (**T8.1 and 8.2**) is to quickly bridge the gap between basic research, industrial validation, and real-world solutions. In terms of artefacts, it is set between research publications on the one side and repositories for software and solutions on the other. To support the process from publication to products, its three goals are as follows:

- **To inform** different target groups of researchers and AI power users of new developments in AI research and put this research into perspective with current market developments. As a main, flexible channel to spread such information, an AI research blog will be set up, which will be fed by research from HumanE AI initially, but will focus on attracting a European community of collaborators.
- **To enable** AI developers to quickly get up to speed with new research topics and approaches in the form of short and focused tutorials and introductions. As the most promising medium to quickly share this information with a large group of potential users, stakeholders will produce short videos on YouTube.
- **To create** new and reusable building blocks for AI solutions in the form of code snippets (i.e., short, well-documented pieces of code that demonstrate the use of new AI technologies on an exemplary problem), so that this code can be quickly evaluated and adopted by AI developers for their own purposes. Snippets in Python in the form of Jupyter notebooks will be the predominant way of distributing this information.

The HumanE AI Net Virtual Laboratory will be closely interwoven with the AI4EU European AI on-demand platform, benefiting both the platform and the HumanE AI Net project:

- The HumanE AI Net Virtual Laboratory will be reachable as a resource from the AI4EU platform.
- The HumanE AI Net Virtual Laboratory will build on the AI4EU collaboration component to facilitate the communication and collaboration both within the consortium and with the broader European AI community.
- As described above the HumanE AI Net project will share publications, datasets, tools and code snippets both within the consortium and with the European AI community. For this sharing the Virtual Laboratory will link to and use the AI4EU and other Ai on demand platforms relevant for AI. As described in section 1.2.4 (see also Task 8.3) one of the contributions of this project to the AI on demand platform(s) will be to implement within the platform the mechanism needed for sharing scientific results, code snippets and and benchmark datasets and having the research community effectively collaborate using them.

An important target group of the Virtual Laboratory will be "power users" of AI, i.e., people who incorporate the latest AI research results into their own research domain, product, or service, but are not fluent in AI research themselves. These power users are early adopters of the latest research and bring innovation to the market by creatively mashing up new and proven AI technologies. Providing power users with a rich environment of solutions and ideas will help bring AI solutions more quickly and directly to the market. A special effort will be made within the Virtual Laboratory to bring the project research results to that group.

2.2.1.4. Dissemination and exploitation through Established AI Networks (CLAIRE, ELLIS, EuRAI)

The HumanE AI Net consortium is well embedded within all relevant AI organisations such as CLAIRE, ELLIS and EuRAI and projects and initiatives such as AI4EU, European Language Grid, SoBigData etc. **WP 9** is devoted to the dissemination through such initiatives It is lead by

2.2.1.5. Impact through Venture and Value Creation

The The HumanE AI network aims at creating high impact for European Citizens through sustained positive societal impact and through the creation of economic value in industries, SMEs, and start-ups. Interdisciplinary cooperation and exchange between science, culture and the public on the one hand as well as industry and research on the other hand is important in the future and can assure an early on connection between the technology push and the market pull. **WP 7** is devoted to this activity

2.2.1.5.1. Platform for AI Innovation

To ensure that real-life problems are addressed, understood, and connected to research early, a platform will be provided that links people, research, ideas, businesses, and investments to lead to successful innovation. This platform allows a centralized, integrated approach that facilitates co-creation between the relevant players, supports vital exchange across disciplines and industries, and ultimately joins forces across Europe toward a strong AI innovation strategy. The platform will be embedded in or connected to existing networks (e.g. AI4EU) and designed to attract and incentivize participation from relevant target groups within research, industry (corporate, SMEs, and startups), talents, venture capital firms, accelerators and incubators, as well as local, national, and international communities and organizations. This platform will provide a forum to research groups showcasing their microprojects, and attract institutions and groups to opportunities (e.g., university entrepreneurship centers, company builders, corporate intrapreneurship programs, mentors, experts, co-founder, investors, or venture capitalists). The platform allows a multidirectional approach for innovation-creation and enables all stakeholders to actively engage with the community to reach the common goal. The platform's transparency will help anyone track the success and progress of projects and show further gaps and needs in the European Innovation Ecosystem.

2.2.1.5.2. Collaborative European Structures and Events

To create a vivid and dynamic analog ecosystem aside from the digital platform, events and structures are created that foster exchange and collaboration. To solve the biggest common problems and allow efficient innovation in AI, three main "offline structures" are built: (1) Future vision conferences for specific domains, (2) a European Data Hub, and (3) a European regulatory Co-Development Hub.

To bring all stakeholders together, inspire, ideate, and create a future disruptive vision for different domains, interdisciplinary future conferences will be held. For these conferences, each for one specific domain (e.g., mobility), all relevant stakeholders (industry, startups, research, nonprofit, and governmental) will be invited to come and create a future vision for the domain. In cooperation with arts and culture, a final fair will be organized that is accessible to the public. In the fair cooperation across disciplines, industries and nationalities are celebrated and collaboration and networking strengthened, and public feedback and interaction is facilitated.

2.2.1.6. Management of Research Data & Knowledge Management

The consortium will enforce the capacity for the project to generate intellectual property, a topic that will be discussed in detail at the kickoff meeting. Procedures and rules for managing knowledge and intellectual property issues both for the project's lifetime and for later exploitation of project results will be discussed as well. We will address confidentiality and publication procedures, mechanisms for intellectual property rights (IPR) reporting and dispute resolutions, rights to issue patents and grant licenses, joint ownership issues, and access rights for both during and after the project's term. If the consortium decides to apply for patents, the Project Coordinator will provide support and assistance and ensure that a nondisclosure agreement is signed and IP arrangements are in line with institutional requirements. The rules in the agreements will respect the following principles:

- The main principle is that knowledge shall be the property of the contractor generating it.
- Where several contractors have jointly carried out work generating the knowledge, and where their respective share of the work cannot be ascertained, they shall have joint ownership of that knowledge.
- Confidentiality will be maintained for all information gained from partners through deliverables or by other means while carrying out the project, unless this information is already in the public domain.
- If knowledge can be translated into industrial and commercial applications, its owner must adequately protect it in compliance with all legal provisions. If the partner does not intend to respect this duty in a specific country, the European Commission shall be informed immediately and will evaluate whether the protection of such knowledge is necessary in that country (it could decide to adopt protective measures).
- Researchers will have the right to publish research results, subject to the terms of the agreements. Notifications of publication in all forms will be submitted to the project-management board, which will notify consortium members as a whole.

- Partners will include in the agreements a list of existing background IPR that they bring to the project as a basis for their research work, and that will remain their property.
- Should ideas be patentable (or subject to registration under copyright or trademark law), the partner who developed them will choose, in discussion with the Coordinator, how best to file the patents or intellectual property, in line with the terms of the grant agreement.
- Foreground knowledge created and protected by one partner will be made equally available to all other partners at preferential licensing terms (as compared to market terms). Foreground IP created jointly by two project partners and not ascertainable will be jointly owned by the contributing project partners. Any details will be addressed in the agreements.
- Licensing terms for foreground knowledge will be negotiated on a case-by-case basis..

2.2.1.7. Use of Open Data and Open Research Data to Increase Impact

With respect to any data that the HumanE AI preparatory action will produce (for example, as a result of literature research or community analysis) in preparation of the research agenda, we will participate in the EC's Pilot on Open Research Data and deposit the relevant data (as foreseen by related Horizon 2020 documents) in a research data repository and, to the maximum extent possible, implement provisions for third parties to access, mine, exploit, reproduce and disseminate this data.

2.2.2. Communication Activities

Communication activities will be carried out to promote the project and its findings during the project period. Whereas the dissemination activities will focus especially on relevant stakeholders, communication activities will focus more on the wider EU public. Task 8.8 is devoted to this purpose.

2.2.2.1. Core communication activities

We will develop and implement a HumanE AI Net dissemination and communication kit with the following:

Project logo. This is needed to establish the project's visual identity, to appear on all project presentations, online videos, webpages, and training materials.

Project presentation. This is a public, high-level presentation of the project's aims and objectives; research infrastructure scope, design, and positioning; transnational and virtual access offered; networking activities and how to get involved in the HumanE AI Net starting community.

Presentation template. This includes the project logo and visual identity, as well as funding acknowledgments. It will be used by all partners when creating dissemination and training materials arising from the project.

Project posters. These contain an appealing graphical design and presents the key aim, objectives, and results of the project. The posters also will promote the project website and be presented at workshops and other dissemination events. The posters are continuously updated, with new ones created and customized for specific events and stakeholders.

Project brochure/fact sheet. This contains similar content to the project presentation, but in an appealing foldable brochure design.

Project website. This is the web face of the project, where stakeholders find information about the project, publications, public deliverables, training materials, open and past calls for transnational access, networking events, and ways to join the HumanE AI Net community. The website also will provide an attractive entry point to the HumanE AI Net virtual research infrastructure (see WP7 and WP9).

Press releases and media coverage. Press releases are published by the coordinator at key milestones and/or events. In HumanE AI Net, two major press releases are planned for the creation of the SoBigData Association and Foundation. Moreover, other press releases will be published when important collaboration agreements are signed or major international initiatives begun. Media coverage will be tracked and reported on the project website, social media channels, and dissemination reports.

Social media presence: HumanE AI Net will continue to be an active presence on social network platforms such as Twitter and Facebook, to promote events, share results, announce collaborations, or provide other news about the projects. Social media also will be used to promote other communication tools actively (e.g., website, publications, deliverables, software releases, press releases, calls for transnational access projects, blog posts, and other news). The success of the communication activities will be monitored closely and reported in the periodic activity

deliverables arising from WP 10. In more detail, download statistics for the various materials from the dissemination and communication kit will be reported, as will be statistics for website access, presentations given, peer-reviewed papers published, media coverage of press releases made, and social media activity reach (e.g., the number of Twitter followers, retweets of key messages, number of Facebook group members, number of posts, and discussions)

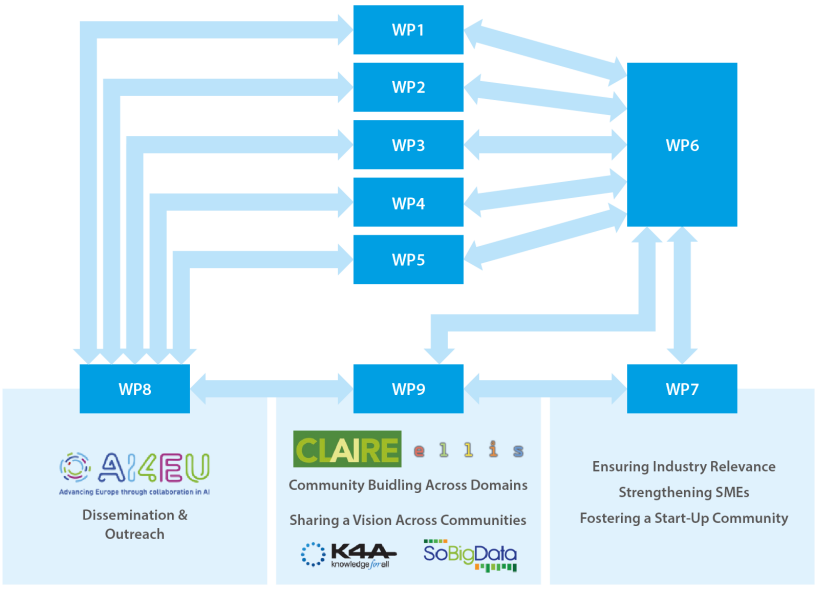


Figure 4. Relationship of the HumanE AI Net WP.

Press releases and media coverage: Press releases are published by the coordinator at key milestones and/or events. In HumanE AI Net two major press releases are planned: On the creation of the SoBigData Association andFoundation. Moreover, other press releases will be published when important collaboration agreements are signed or major international initiatives are started. Media coverage will be tracked and reported on the project website, social media channels, and dissemination reports.

Social media presence: HumanE AI Net will continue to be an active presence on social network platforms such as Twitter and Facebook in order to promote events, share results, announce collaborations or to give other news about the projects. Social media will also be used to promote actively the other communication tools, e.g. web site, publications, deliverables, software releases, press releases, calls for transnational access projects, blog posts, and other news. The success of the communication activities will be monitored closely and reported in the periodic activity deliverables arising from WP 10. In more detail, download statistics for the various materials from the dissemination and communication kit will be reported, as will be statistics for web site access, presentations given, peer-reviewed papers published, media coverage of press releases made, and social media activity reach (e.g. number of Twitter followers, re-tweets of key messages, number of Facebook group members, number of posts and discussions).

3. Implementation

3.1. Work plan — Work packages, deliverables

3.1.1. Work Package Roles and Relationship to each other

In addition to the **Management WP (WP 10)**, the project contains three kinds of WPs. First, P 1-5 are devoted to developing, advancing and implementing the research agenda through microprojects and scientific challenges (see section 1.3.1). They correspond the five core research areas described in sections 1.3.1.2 - 1.3.1.6(**WP1 Human-in-the-Loop Machine Learning, Reasoning and Planning, WP 2 Multimodal Perception and Modeling, WP 3 Human AI Collaboration and Interaction, WP 4 Societal Ai and WP 5 AI Ethics and Responsible AI).** Second, the role of **WP 6 (Applied research with industrial and societal use cases)** is to connect the research agenda of WP 1-5 to industrial needs and ensure that the results are evaluated in industrially (and socially) relevant use cases (according to the strategy described in section 1.3.2.4). Finally WPs 7-9 are devoted to translating the research successes into impact in terms of capacity building, knowledge spreading, visibility, innovation, contributions to Europe's industrial strength, social benefit and creating sustainable collaboration mechanisms. Thus **WP 7** on **Innovation Ecosystem and Socio-Economic Impact** will devise, refine and implement a strategy for going beyond technology transfer to the industrial champions toward fostering start-up creation, interfacing SMI and benefiting the broader European industry. Cearly, WP 6 and 7 will be closely cooperating with WP 6 being focused on **industrial R&D** innovation and WP 7 on translating such R&D innovation into **business innovation and impact**. **WP 8** on **Virtual Center of Excellence, Capacity building and Dissemination,** is devoted to knowledge spreading and making Human a virtual center of excellence including the implementation of a Virtual Laboratory closely linked to the AI4EU platform, the industrial Ph:d. postdoc and internship program and the organization of summer schools,

tutorials and other dissemination events to all relevant target groups. **WP 9** on **Synergies with AI on demand platform(s) and the Broader European AI Community.**

3.1.1.1. Special Common Tasks in WP 1-6

WPs 1–6 all have two common tasks:

3.1.1.1.1. Coordination and consolidation of the research agenda

This task makes sure that the activities of all tasks of a WP lead to a consistent and coherent overall research agenda, helps conduct joint microprojects, challenges and datasets between the tasks and provides an interface to the innovation ecosystem (WP 7), theVirtual Laboratory (WP 8 including contributions to the summer school and tutorial program T8.5, and the cooperation with other initiatives (WP 9).

3.1.1.1.2. Responsible Research and Innovation Assessment (RRIA)

This task ensures continuous support of the WP work with respect to ethical/legal/privacy/ Impact assessment ensuring that the research agenda and results are compatible with our vision of responsible AI by design. This includes: (1) incorporating feedback into future iterations of the agenda and for future microprojects, (2) reaching out to responsible bodies (see section 3.2) if feedback warrants a suggestion of changes to current guidelines, regulations and law, and (3) instigation discussion and reflection of aspects related to trust, robustness, explainability, fairness, rule of law etc. This will include interaction with the LPbD assessment (Tasks 5.1 and 5.2) . Each partner of the WP will be part of this task and devote part of the resources within the WP to it.

3.1.2. Deliverables and Their Roles

In our project concept the deliverables are not just means of documenting results and fulfilling obligations toward the Commission (which is of course a key aspect) but also a vehicle to ensure cooperation between WPs. Thus, WPs 1-5 (the core research agenda WPs) all contribute to **D6.1 (M09) ,6.2 (M21) and 6.3 (M36)** which are **" First/Second/Final report on research results, their application significance and the resulting evolution of the research agenda**). The deliverable is coordinated by WP 6 which, as described above, connects the reserch agenda to the industrial needs and use cases. This strengthens the link between research and industrial relevance. Simularly WPs 1,2,3 as well as 4 and 5 have joint Deliverables rerlated to collecting and depositing their micro-project results on the AI4EU D1.1-3.1 and 4.1,5.1,4.2 again fostering cooperation.

3.1.3. Role and Management of Microprojects (and Challenges)

Our strategy for implementing the research agenda is built around the concept of collaborative microprojects. In a microproject, a group of researchers (often PhD students) from different partners spend 1–6 months at the same "host" lab working toward a tangible goal such as a paper, dataset, or tool. This ensures that work focuses on those gaps in knowledge that require the **combined expertise of different partners** to be filled. With the exception of PMs for WP and task leadership, all PMs assigned to WPs 1–5 will be spent on microprojects. Partners who will not participate in microprojects will not be able to claim their PMs. Examples of microprojects are given in the respective WP forms. Overall, the process of defining microprojects will be part of the research agenda refinement work of each WP. This process will have three variations (see also section 3.2.3.1 on microproject governance):

1. The few microprojects that are sketched in the proposal will start M01 or 2 to make sure that research work is initiated without loss of time.
2. As long as they have PMs left, groups of partners will be able to define microprojects in a bottom-up way. Formally, just a short online form will have to be filled. The WP Steering Team (WST) of the WP (see section 3.2) will have the right to object (within a week) if the microproject is grossly out of scope of the research agenda of the respective WP/task, or in conflict with the RRIA, in which case the escalation process will occur. Otherwise, the microproject will proceed.
3. As described in section 3.4, funds have been set aside to be flexibly assigned to additional microprojects, and to be used to invite external researchers to participate in microprojects. To access these funds, a more elaborate proposal must be sent to the WST (but no more than 3 pages). If approved, the microproject will immediately commence; otherwise, partners can appeal through the process described in section 3.2.

Microproject including partners and topics from different WPs will also be encouraged with each WP contributing PMs from its own pool and the microproject being approved by the WST of each WP.

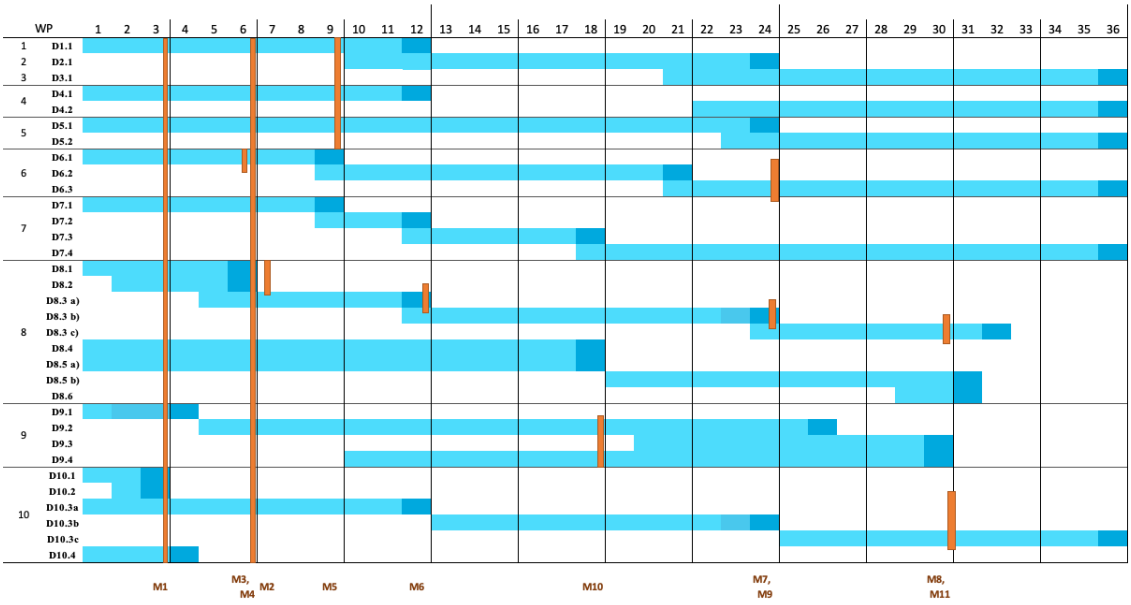
Within WP 6, the industrial procedure will be slightly different, with the microproject emerging from the stakeholder workshops and ultimately being proposed by the industrial champion of each domain (=task). Research partners will then be able to volunteer to be part of the microproject. Such microprojects could be financed by a mixture of PMs

from WP6 (for partners who are in WP 6) and from WPs 1–5, as cross-WP microprojects connecting basic research and industrial needs.

Challenges will either be run in form of microprojects or organized centrally through the coordination and consolidation of the research agenda task (run by the WP leader) in collaboration with WP 8.

3.1.4. Project Timing

As described above, all WP are essential not just to establishing the network but also to running it. This includes conducting microprojects in WPs 1–6 as well as the exploitation, dissemination, and communication activities in WPs 7–9. Thus, all WPs will be running throughout the entire length of the project. Individual tasks, such as 8.1 (setting up the Virtual Laboratory) or 8.3 (building the infrastructure for scientific collaboration within AI4EU) start at the project's beginning and must produce results by the end of year 1.



Overall, the work is driven by annual cycles given by the deliverables that summarize and collect the results. The deliverables-driven timing is shown below.

3.2. Management structure, milestones and procedures

Each WP is led by investigator/researchers with an established track record and solid experience in running projects of this scale. It is a deliberate design choice to have a mix of senior researchers and younger ones in the increasing stage of their careers, in order to blend experience and enthusiasm.

3.2.1. Management Roles

Three key management roles are defined for the project:

1. The Project Coordinator (PC), Prof. Paul Lukowicz, Scientific Director at the German Research Center for Artificial Intelligence (DFKI), will be in charge of the overall scientific coordination and project implementation, as well as administration. He is the coordinator of the HumanE AI FET preparatory action on which this proposal builds and has a long history of coordinating successful EU projects (5 projects over the last 15 years). The coordinator is the contact person for all interaction and communication with the European Commission and all legal aspects, monitoring the implementation and progress as well as the achievements of the project and reporting to the Commission. The PC acts as chair of the Project Management Board, the Project Steering Board and the Project Management Team. **Two Deputy coordinators** will assist the PC: Virginia Dignum and John Shaw-Taylor.

2. The Project Manager (PM) will report to the Project Coordinator and will be responsible for administrative management tasks (organisation of meetings and events, coordinating the preparation of reports, monitoring and maintaining the content on the website, facilitating communication within the consortium, and preparing publicity materials). The PM will be responsible for resolving financial and other administrative issues and monitoring the completion of technical and business objectives. The PM will also participate in facilitating dissemination and training activities in collaboration with the responsible WP leaders. The PM will be Dr. George Kampis (DFKI), who has had this role in numerous previous successful EU projects such as CIMPLEX, and the FET preparatory action HumanE AI (with 36 partners).

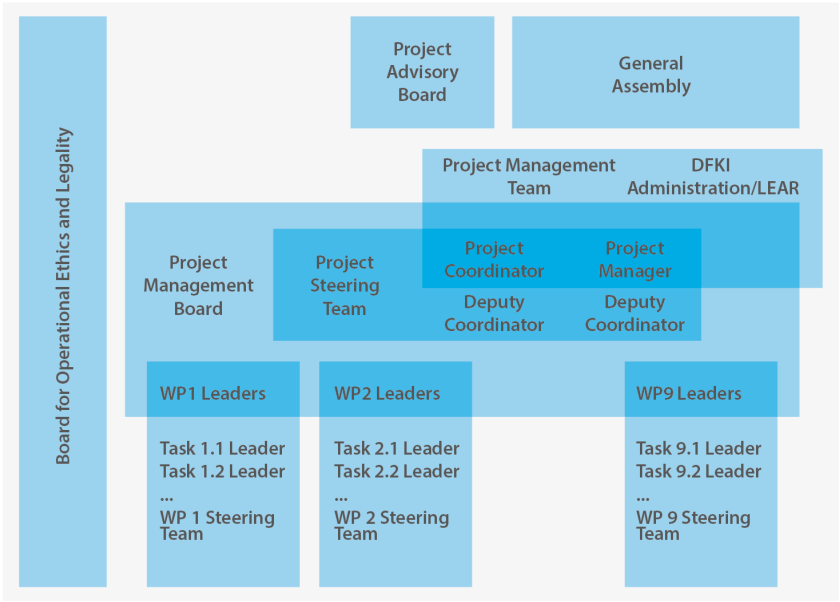


Figure 5 The HuamnE AI Net management structure.

3. The Work Package Leaders (WPL) are key investigators who will manage and monitor the scientific and technical objectives, deliverables and milestones of their respective work package and report to the Project Management Board. WP Leaders are responsible for maintaining an archive of documents, including draft documents, deliverables and meeting minutes, which are stored in the private area of the project website.

4. The Task Leaders (TL) are investigators who will manage and monitor the scientific and technical objectives, effort expenditure, deliverables and milestones of their respective tasks.

5. For each WP the WPL and all the TLs will constitute **the The Workpackage Steering Team (WST).** Its main task will be the selection and are investigators who will manage and monitor the scientific and technical objectives, effort expenditure, deliverables and milestones of their respective tasks.

6 The Project Management Team (PMT) will provide a management structure that monitors the progress and quality of the project and resources. The PMT will consist of the Project Coordinator and the Project Manager, supported by the DFKI European Office (LEAR). The Project Coordinator may co-opt into the PMT other representatives as necessary to assist her in the day-to-day running of the project. This Team is the central locus for the ultimate success of the project, so particular attention will be paid to monitoring the Gantt chart of the Description of Work for time planning, project reporting and financial control to ensure the project is delivered on time and to budget.

7 The Project Steering Team (PST) is composed of the Project Coordinator (chair), the Project Manager and the two deputy project coordinators. Its primary focus is on operationalizing the scientific vision of the project issues. The PSB is a smaller body than the PMB and is more flexible to meet whenever either the Project Coordinator or a member of the PMB considers this necessary to ensure focused and flexible action. In the first phase of the project the PSB will meet twice per month. When the project is well on track, this will be reduced to four to six times per year. The meetings will be a mix of teleconferences and face-to-face meetings as appropriate. The PSB will regularly discuss scientific progress, dissemination, intellectual property rights, exploitation, legal, ethical and regulatory issues and measures to be tabled to consortium bodies.

3.2.2. Management Boards

The main management bodies responsible for the strategic direction, management and administration of the project are defined as follows:

I The Project Management Board (PMB) is composed of the Project Coordinator, the Project Manager and the Work Package Leaders. Meetings will be chaired by the Project Coordinator. This committee is the primary executive decision making body. Each area of work will discuss the specifics of their technical input, progress and effort expenditure within their respective work packages and the PMB will balance these inputs, along practical lines, within the overall technical direction agreed in the original Description of Work. Ethics and Gender issues will

also be monitored by the PMB. There will be meetings of the PMB at least every six months and more frequently, if required. These will be virtual or face-to-face, as appropriate.

II The Project Advisory Board (PAB) will comprise recognized experts of international standing from a range of academic and non-academic organizations. Members will be senior researchers, business executives, educators and decision makers from both technical and non-technical backgrounds. The PAB will review selected reports and attend meetings to offer the project the benefit of their domain and technical expertise and help position the technical advances of the project in a wider scientific and commercial context. The PAB will meet at least three times during the course of the project. Meetings will be chaired by the Project Coordinator. Members will be required to sign appropriate confidentiality agreements covering their involvement.

III The General Assembly (GA): All the partners are members of the GA (one person (= vote) per partner). The Coordinator acts as chair. The GA decides on: the yearly technical and financial plan, the plan for use and dissemination, the strategy and procedure for publications, the Training Programme. All beneficiaries will be invited to participate in the meeting and to ask questions, give advice and propose actions. The General Assembly also play a key role in assessing and deciding whether milestones have been achieved

IV The Board for Operational Ethics and Legality (BOEL) and the High Level Advisory Board (HLAB): described in task 2.1 and 2.3 of WP2 and in detail in Section 5.

3.2.3. Management Procedures and Tasks

Note that the responsibilities of the **PMB** differ from those of the **PMT**; the PMB is responsible for managing the successful execution of the project in terms of the research and other activities described in the Description of Work, and will have members from all work packages. The PMT, on the other hand, addresses administrative, legal and financial matters, and is based entirely at the lead institution (DFKI). The PMT will be supported by the DFKI EU Administration, which will be responsible for legal issues (e.g. Grant Agreement negotiations, Consortium Agreement drafting) and financial issues (e.g. receipt, transfer, and accounting of payments, collection of audit certificates). Each beneficiary will be responsible for their local financial and administrative obligations and will report to the Project Coordinator. Secretarial support will be provided by existing departmental staff as required. Project coordination. The Project Coordinator will define and implement the management functions of HumanE AI Net, in line with the Consortium Agreement, drawn up and signed by all partners prior to the project start. The Project Coordinator will define a Management Charter that defines the practical steps for implementing the Consortium Agreement, with regard to events such as changes in the PSB membership, loss of a partner, and dealing with under-performing partners. The Project Coordinator will be the main interface between EC officials and the project. The Project Coordinator will monitor dissemination and exploitation activities in collaboration with the responsible WP leaders, provide guidance and direction as appropriate.

Project Management Board meetings. Scheduled meetings of the PMB will be announced at least four weeks in advance unless called under an ad-hoc request (see below). The agenda will be prepared by the Project Manager in consultation with the other members of the PMB. Agendas will be sent out to all participants via email at least two weeks in advance (except for PMB ad-hoc meetings). Minutes will be circulated within a week from the meeting taking place and partners will have up to fifteen days to comment on them. The PMB will monitor progress and take the necessary corrective measures, in case there are deviations from the agreed effort or work plan. Decisions will be taken by a majority vote (50% + 1 partners present at the meeting).

Ad hoc meetings. If required, any member of the PMB or PSB has the right to call for an ad-hoc meeting by notifying the other members by email or phone. These meetings will be organised as soon as practicable. Virtual meetings will be organised within one week of the request.

Project Management Team tasks. The PMT will make operational decisions necessary for the smooth running of the project. All major decisions need approval by the PMB. PMT tasks include: day-to-day management of the project, ensuring an effective communication; collecting cost and other statements from all partners for submission to the EC; preparing all progress and financial reports and documents required by the Commission; ensuring prompt delivery of all data identified as deliverables in the Description of Work or requested by the EC for reviews and audits, including the results of the financial audits prepared by independent auditors; allocation of budgets to the partners in accordance with the EC contract; taking measures in the framework of controls/audit procedures; reviewing and proposing budget reallocations to the partners (to be approved by the PMB) and making proposals to the partners for the review and/or amendment of the terms of the EC contract and the consortium agreement (to be approved by the PMB).

Website and document management. The public website will be used for knowledge dissemination generated by the project as well as relevant links to other sites and publications. Content added to the site will be monitored by the Project Manager. In addition to the public website, there will be a private wiki, protected by access management controls, that will be used to hold internal documents, draft versions of reports and other deliverables, and for

internal project discussions. Access to certain areas of the private site will be granted to EU representatives, to enable them to download deliverables ready for review.

Reporting. At the requested time points during the project, reports will be submitted to the EC. All annual reporting will be undertaken by the Project Coordinator. WP leaders will be responsible for collecting and aggregating information relating to their work package and submitting it to the Project Manager. Members of the PMT will assemble the report, which will be approved by the Project Coordinator prior to submission to the EC. The private area of the project website will act as a repository for reporting information and for the exchange of reporting information.

Quality assurance. The project will implement review procedures for internal and public deliverables, to ensure a consistent level of quality and scientific excellence and compliance with EC policies, such as open access submissions for project papers. More specifically, (I) Internal reports must be approved by the WP leader prior to upload to the approved area of the private website; (II) Deliverables must be approved by the related WP leader before being forwarded to the PMB for final approval. Drafts of deliverables will be required to be submitted prior to the final deadline to allow sufficient time for internal review; (III) Key selected deliverables will be forwarded to the PAB members for review. The collected feedback will be used to guide the strategic direction of the project. The Operational Plan prepared by the Project Coordinator at the start of the project will define all the relevant reporting and operational procedures.

Intellectual property management. The Project Coordinator, and representatives from the beneficiaries will confer on a regular basis with the aim of carrying out technology audits of the projects' results as they occur, with a view to determining what advances are viable of commercialisation or other forms of exploitation and will agree arrangements (within the terms agreed in the Consortium Agreement) for any protection of arising information and knowledge. IPR management plan will be prepared, to define specific procedures to be followed within the project.

Addition of beneficiaries during the lifetime of the project. The consortium will be open to new partners if one of the founding partners should withdraw from the consortium or new tasks should be identified that cannot be fulfilled by any of the existing Partners. In this case, the rules for inclusion of new partners will be followed and all existing legal documents will be adjusted according to the requirements of the Grant Agreement. In such instances, the process will be managed by the EU Research Funding Office of the Coordinator.

Voting and decision making. Each beneficiary has one vote in the consortium body of which it is a member. Nevertheless, unanimous consensus is obviously the first option to be explored at all times in decision making. In cases where no consensus can be reached the decision will be taken by majority vote. In case of equal votes, the chairperson's vote is decisive. When a decision has an impact on several WPs the decision is taken at PMB level. In case of disagreement by one of the partners affected by the decision, the final decision will be taken at PMB level.

Dispute resolution. If any conflict, whether technical, managerial or financial, should occur during the course of the project, it will be resolved by the PMB either at its next meeting or by other means of communication, in accordance with the procedure detailed in the Consortium Agreement and endorsed by the PMB. These dispositions will be in line with the current recommendations of the European Commission policy and the model Grant document.

Communication. Internal communication will be open and transparent to ensure that partners are kept fully informed of progress, results, developments and decisions. The private area of the project website and wiki will function as a day-to- day interface, as well as a pathway to distribute and exchange information. Internal project communication will be facilitated by mailing lists. Periodic newsletters documenting project achievements will be prepared and circulated by the PMT. To facilitate project internal communication, an initial workshop will be held near the project launch date, with the aim of promoting a creative collaborative environment for the project and establishing direct lines of communication between participants. External communications will be facilitated through the public area of the project website, which will be used to host all public deliverables, publicity materials, links to publications, conference reports and information on the results achieved. The project will arrange a presence at major conferences and exhibitions as well as participating in liaison and clustering activities with other related

3.2.3.1. Governance of microprojects.

As described in 3.1 , with the exception of PMs for WP and task leadership all PMs assigned to WPs 1-5 (and many in WP 6) will be designated to microprojects. Partners who will not participate in microprojects will not be able to claim their PMs.

In 3.1 we have described how microprojects will defined in a bottom-up way on WP level approved in a light way process by the WST of the respective WP. The exact process will be defined in all details at the kick-of meeting and laid out in D 10.1. The definition will include the exact obligations of microprojects which will include providing input for WP 7 to support the innovation process and for WP 5 to support ethical/legal assessment and research. We expect that this mechanism will lead to seamless operation of high quality microprojects well aligned with the research vision of HumanE AI Net. However the following quality assurance and conflict resolution mechanism will be implemented by to prevent and solve any unexpected problems.

Quality Control: We consider quality control with respect to 1) the definition and selection of the micro projects and 2) the quality of the results,

As described in 3.1 the first step in the quality control process is the review of the proposed microprojects by the WST of the respective WP. In addition the PST will go over all approved microprojects during the monthly video meeting looking for any quality issues in the selection.

With respect to the results a similar two step procedure with first the WST inspecting the results in detail and then the PST having a look at selected mciroöproject or at ones flagged by the

Conflict Resolution

We consider the following conflicts with respect to microprojects selection,

1. Proposer of a microprojects disagrees with decision of the WST.
2. A project member disagreeing with the selection of a specific microprojects.
3. Disagreement between WST and and PST over micro project selection.

Analogous is possible with respect to results quality assessment:

4. Proposer of a microprojects disagrees with decision of the WST.
5. A project member disagreeing with the selection of a specific microprojects.
6. Disagreement between WST and and PST over micro project selection.

In cases 1,2,4,5 the first escalation step is involving the PST who will set up a video call of the respective WP leader and the project member who has the complaint trying to mediate. Should it fail as well as in cases 3 and 6 the escalation procedure is to involve the PMB who will make the final decision.

The above proposed conflict and quality management procedures will be discussed refined and finalized in the project manual (D10.1)

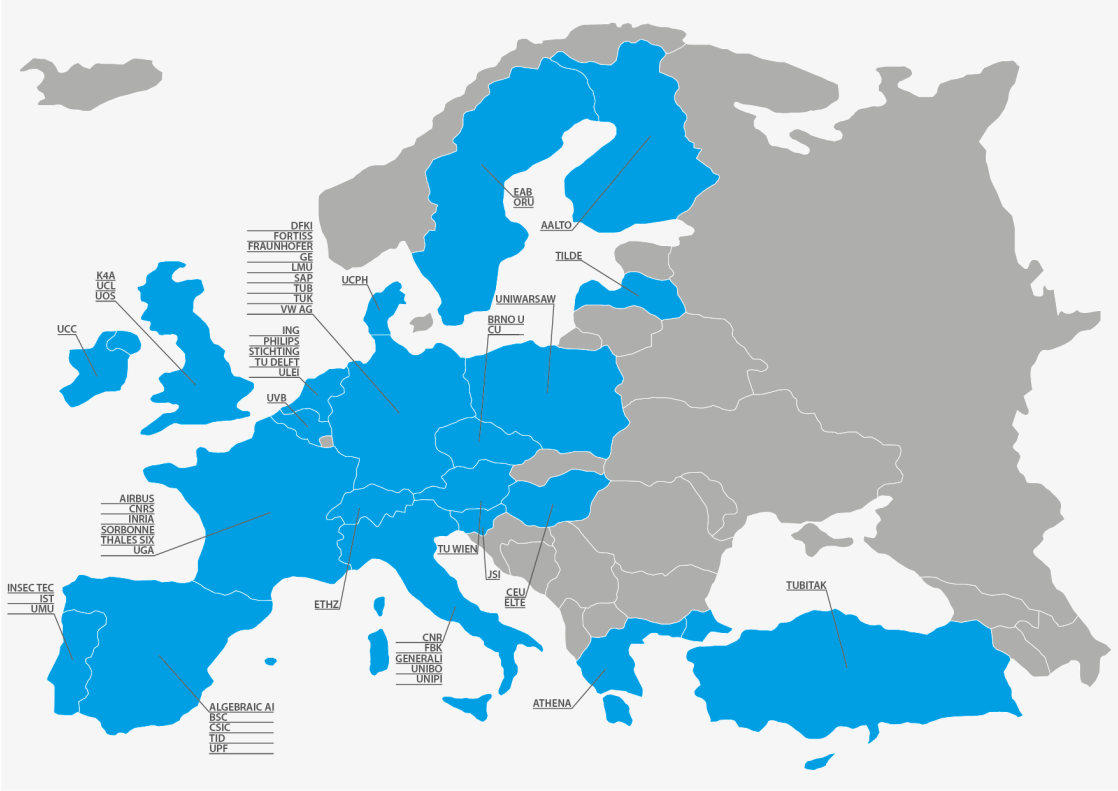
3.2.4. Critical Risks Management

The consortium has designed its work plan in a way that prevents most common risks right from the beginning. This is achieved by proactive risk management: modular work packages, clear responsibilities and minimization of critical paths. If one work package is delayed, other packages can still continue to achieve their results, project board members have proven skills and track record in international projects and research organization. One person is devoted full-time to financial and other project administration tasks. There is already a preliminary contingency plan defined for the major risks (TABLE 3.2B). This will be updated in each management reporting deliverable. Risk analysis and management will be performed by the PSB and PMB on a regular basis and taken into account in key decision making. HumanE AI Net is a highly innovative project, so changes in user requirements, business needs, time schedules and costs may occur. The PSB and PMB will make a risk evaluation at each milestone, in order to determine whether the level of risk is acceptable and appropriate for the project. Appropriate contingency planning will be carried out accordingly. A comprehensive state-of-the-art Risk Assessment and Management Plan will be implemented within the first six months of the project and will address different kinds of risk (external, internal, strategic, operational, other). This work, which is part of the management activity, will: (I) Identify risks of any nature that might occur during the project and assess their probability and potential impact on the project; (II) Plan concrete actions to prevent risk occurrence; (III) If problems do occur, then the associated contingency measure will be swiftly implement in order to minimize impact.

3.3. Consortium as a whole[23]

The consortium has been put together along three dimensions. First are the competences. As discussed throughout the proposal our vision goes beyond a narrowly defined classical core AI scope to include Human Computer Interaction, cognitive science, social science and complexity science. Participants in the project cover all those areas e.g. machine learning (e.g. Prof. John Shaw Taylor from UCL, Prof. Klaus Müller TU Berlin, Prof. Samuel Kaski AAlto), reasoning and symbolic AI (e.g Prof. Frank van Harmelen, Amsterdam, Prof. Paolo Traverso FBK, Thomas Eiter TU Wien. Prof. Tomasz Michalak UW) Multi-modal Perception and Modeling (Prof. James Crowley, Inria, Prof. Paul Lukowicz DFKI), Natural Language Processing (Prof. François Yvon, LIMSI/CNRS, Prof. Jan Hajič, Charles University, Dr. Bernardo Magnini, FBK, Prof. Jan Černocký, Brno Univ. of Technology), Human Computer Interaction (e.g Prof. Abrecht Schmidt LMU, Prof. Yvonne Rogers UCL, Wendy McKay INRIA, Prof. Antti Oulasvirta), computational social science (Prof. Andrzej Nowak, UW, Prof. Frank Dignum, UMU, Prof. Ana Paiva, IST), AI explainability, & ethics data mining and Design for Values (Prof. Fosca Giannotti CNR, Prof. Dino Pedreschi UNIPI, Prof. Virginia Dignum, UMU, Prof. Jeroen van den Hoven, TUD, Dr. Nardine Osman and Prof. Carles Sierra IIIA-CSIC) and others.

Second is the inclusion of a broad range of institution types with strong political and industrial (including key industry) connections. Thus the consortium includes most of the major AI related research centers in Europe such as the coordinator DFKI and Fraunhofer in Germany, CNR, FBK, CINI in Italy, INRIA and CNRS in Italy, INESC in Portugal, IIIA-CSIC and Barcelona Supercomputing Center in Spain, ATHENA in Greece, and FCAI in Finland. There are also many of Europe's top Universities such as Sorbonne, UCL, LMU, TU Berlin, TU Wien, ETH Zürich, TU Delft, VU Amsterdam and VUB to (name **just a few).** To ensure a synchronisation with European industry we have included key European players, each representing one specific domain with particular importance to the continent's economy. This includes Airbus (aerospace), Generalli (insurance), ING(FinTec), Philips (Health), Telefonica (Telco) and Volkswagen (mobility/automotive). In addition SAP will contribute its security related research, Thales (coordinator of the AI4EU project) will be the link to the EU on AI on Demand platform, and Tilde will be incharge of horisontal language and multilingual technologies. Furthermore we have with Algerbraic AI a startup devoted to a specific novel learning technology (Algebraic AI) and with German Entrepreneurship a consultancy devoted to technology transfer and creating and running incubators (in charge of WP 7).



Third we have aimed for a comprehensive national coverage as illustrated above. Thus we cover 20 European countries with a coverage stretching comprehensively from east to west and from south to north. Further countries will be included in the operation of the network through the mechanism for inclusion of external partners in microprojects and challenges. The choice of industrial champions have also included the national distribution considerations with the industrial 12 partners coming from 8 different countries.

3.3.1. Roles of partners

The detailed roles of partners are described in section 4 and with over 50 partners elaborating each individual role within this section is not feasible. Key considerations are as follows:

DFKI is the coordinator and also **leads WP 6** on "Applied research with industrial and societal use cases." Having the coordinator run WP 6 This underscore the importance we attach to synchronization of the research agenda with industry needs. DFKI, being a public private partnership with a focus on technology transfer and a strong record on industry related projects. **The individual tasks are lead by the respective industrial champions**.

WP 1 on "Learning, Learning, Reasoning with Human in the Loop" will be lead by UCL, specifically Jahn Shaw-Talor (**who is also deputy coordinator of the project**). Other key partners and people with respect to the topic will be **(**more ML side**)** Klaus Robert Müller (TU Berlin), Christian Igel (Copenhagen), Samuel Kaski (Aalto), Jaoao Gama (INESC), Marco Grobelnik (JSI) and **(**more reasoning**):** Frank Harmelen (VU Amsterdam), Paolo Traveso (FBK), Ramon Lopez de Mantaras (CSIC Bercelona), Holger Hoos (Leiden), Thomas Seidl (LMU).

WP 2 on Perception and Modelling is lead James Crawley from INRIA. Other key partners and people with respect to the topic will be (among others) Paul Lukowicz (DFKI), Daniel Roggen and Ron Chrisley (Susex), Ana Paiva, IST), Luc Steels (UPF), Nadia Berthouze (UCL), Raja Chatila (Sorbonne), Marco Conti (CNR),

WP 3 on Human AI Interaction and Collaboration is lead Aalto specifically Antti Oulasvirta with key people supporting him being (among others) Albrecht Schmidt (LMU), Wendy MacKey (INRIA), Kasper Hornbaek (Copenhagen), Yvonne Rogers (UCL), Raja Chatila (Sorbonne),

WP 4 on Societal AI is lead by Dino Pedreschi (PISA), with the support of (among others) Guido Calderli and Fosca Giannotti(CNR), Janos Kertes (CEU), Jan Hajič, (Charles University), Dirk Helbing (ETH, Social Sceince), Andrzej Nowak (Warsaw, Social Science) Frank Dignum, (UMU)

WP 5 on Ethical and Responsible AI is lead by Virginia Dignum (Umea, **deputy coordinator of the proposal, member of the EU High Level Experts Group on AI**), who is supported among others by Jeroen van den Hoven (Delft), and Mireille Hildebrandt (VUB who is professor of Law specializing on ICT and AI)

WP 7 on Innovation Ecosystem and Socio-economic Impact is lead by German Entrepreneurship with the support of LMU, ING, Fortiss and Volkswagen.

WP 8 on Virtual Center of Excellence, Capacity building and Dissemination is lead by Sorbonne who also run the Ph.D. program. Partners include CINI (running the summer school and dissemination to the scientific community task) K4All (foundation specializing on AI outreach) running the dissemination material creation and global outreach, Fraunhofer running the Virtual Laboratory and Alessandro Saffiotti from Orebro (leader WP 7 in AI4Eu) running the integration of research collaboration infrastructure into the AI4EU platform.

WP 9 on Synergies with AI on demand platform(s) and the Broader European AI Community is lead by Barry O'Sullivan who is president of the EuRAi association. Individual tasks are run by key people in the respective initiatives (as described in the tasks). Thus, for example, the coordinator of the AI4EU project is (THALES) is for example running T9.1 on interaction with AI4EU.

Overall it can be seen that the project brings together top experts and institutions from the respective areas within every WP including industrial expertise where needed.

3.4. Resources to be committed

The resources in the project are distributed according to the following concept:

1. Each partner gets a funds for maintain basic network activities (travel, administration, paying for open access publications etc.). The basic "unit" is 40K Euro. How much each institution gets depends on the number of researchers from this institution who are involved. Overall around 25% of the 12 Million is assigned in this category.
2. Each partner gets funds that can be used for participating in micro-projects. We have a basic "funding unit" of 60K Euro which, depending on partner costs can allow between 6- 12 months of micro-project participation. Again different institutions get different number of such "units" depending on the involved researchers from the institution. Overall 35% of the 12 Million is assigned in this category.
3. Finds are made available for task and WP leadership (which requires personnel effort). Overall 7% of the 12 Million is assigned in this category.
4. Certain partners have specific tasks that fall outside the above categories. These this include (1) leading the implementation of the innovation ecosystem within WP 2 (2) mplementing the scientific cooperation infrastructure within AI4EU (T8.3) (3) implementing and running the Virtual Laboratory (T8.1. and 8.2) (4) running summer schools etc. (T8.5), (5) interfacing to AI4EU (T9.1) (7) producing dissemination materials (e.g. MOOCs) (T8.6) (8) project management WP 10, (9) eupporting LPbD and ethics assessment (T5.1 AND 5.2)

   Overall 10% of the 12 Million is assigned in this category
5. Slightly less then 10% is reserved to dynamic distribution of micro-project funds for which partners within the consortium will compete with ideas

Slightly less then 10% is devoted to inviting researchers outside the consortium to participate in the micro-projects and challenges

Excessive travels of the 53 Partners (in 83 groups, 152 persons) are mainly serving the frequent regular project meetings planned. Besides, we have formed two sums, both allocated now to the Coordinator but used by the entire Consortium: 1mEuro (ie 800k without overhead) personnel cost to be transferred to other partners as the need arises, as well as 960k (768k net) travel costs for 3rd persons who are not supported by the project but invited to project meetings.

Table 3.1a: List of work packages

WP No	Work Package Title	Lead No	Lead	PMs	S	E
WP1	Learning, Reasoning and Planning with Human in the Loop		UCL	181,5	1	36
WP2	Multi Modal Perception and Modeling	23	INRIA	177,1	1	36
WP3	Human AI Interaction and Collaboration	2	Aalto	241,8	1	36
WP4	Societal AI	47	UNIPI	105,8	1	36
WP5	AI Ethics and Responsible AI	45	UMEA	114,7		36
WP6	Applied research with industrial and societal use cases	1	DFKI	213,1	1	36
WP7	Innovation Ecosystem and Socio-Economic Impact	27	GE	46,7	1	36
WP8	Virtual Center of Excellence, Capacity building and Dissemination	31	Sorbonne	116,5	1	36
WP9	Synergies with AI on demand platform(s) and the Broader European AI Community	41	Cork	69,3	1	36
WP10	Managemen and Governance	1	DFKI	100,2	1	36
			Total PM	1366,7		

Table 3.1b: Work package description

WP 1 Lead UCL	Learning, Reasoning and Planning with Human in the Loop
M01-36	**2 (AALTO): 12.9 | 4 (Algebraic AI): 10.8 | 5 (ATHENA): 2.9 | 9 (CNR): 8.3 | 10 (CNRS): 1.7 | 11 (CSIC): 10.1 | 14 (ELTE): 16.2 | 21 (INESC TEC): 6.6 | 25 (JSI): 5.6 | 27 (LMU): 13.3 | 32 (STICHTING): 18.1 | 36 (TUB): 13.5 | 40 (TU WIEN): 6.7 | 41 (UCC): 3.6 | 42 (UCPH): 2.7 | 44 (ULEI): 3.3 | 46 (UNIBO): 2.0 | 47 (UNIPI): 4.7 | 48 (UCL): 25.7 | 50 (UOS): 1.9 | 51 (UPF): 4.9 | 53 (VW AG): 6.0**

Objectives

This WP aims to develop the fundamental Learning, Reasoning and Planing methodologies that allow humans to be interactively involved "in the loop". As outlined in section 1.3.1.2 this goes beyond explainability (which in itself is a challenge) toward methods that allow interactive human input to influence their inner workings.

Description of work

The work in workplace is organized along topics that bring together different communities and avoids the usual split between the symbolic and the subsymbolic communities. It will closely connect to WP

T1.1 Linking symbolic and subsymbolic learning (STICHTING)

This task is devoted to research in the construction of hybrid systems that combine symbolic and statistical methods of reasoning as described in section 1.3.1.2.1. A variety of representations will be considered hybrids of logic and neural networks such as logic tensor networks [Donadello2017], latent representations of knowledge graphs through embeddings [Wang2017] and narratives [Meghini2019].

T1.2 Learning with and about narratives (INESC)

Following the approach described in section 1.3.1.2.2 this task will investigate how narratives based approaches can be used to bridge the gap between human understandable descriptions of complex situations, and subsymbolic representations [Urbaniak 2018][Gilpin2018 to bridge between human reasoning and understanding, on the one hand, and internal AI representation on the other [Vlek 2016]

T1.3 Continuous & incremental learning in joint human/AI systems (UCL)

This task is devoted to the fundamentals of learning through exploiting rich human feedback ("this is wrong *because…*"), exploiting implicit feedback (by obtaining feedback from behavior, voice & face), through

imitation, through active learning (the machine asking the human "should we explore **this**?") as described in section 1.3.1.2.3

T1.4 Compositionality and Auto ML (Leiden)

This task will pursue a compositional approach to delivering AI systems that aims to combine well-understood learning components to create more complex behaviors. As described in section 1.3.1.2.4 key aspects to be investigated will be the link to optimization and and the automation of this approach (AutoML [KotEtAl17]).

T1.5 Quantifying model uncertainty (Aalto)

As described in detail in section 1.3.1.2.5 this task will investigate methods for quantifying uncertainty of Ai systems that on one hand are applicable to composite dynamic systems (e.g. through propagation) while at the same time allowing a human understandable estimate in form of a semantically meaningful explanation of the potential risks and their causes.

T1.6: Consolidation and coordination of the research agenda (STICHTING)

This task implements the consolidation and coordination function for the research agenda of this WP according to the approach sketched in section 3.1.1.1. An important WP specific aspect is the interface to WP 3 (on interaction and collaboration) as a core aim of the developments in this WP is the ability of ML systems to seamless work and collaborate with humans,

T1.7: Responsible Research and Innovation Assessment (RRIA) (UCL)

This task provides WP specific RRIA support according to the approach sketched in described in section 3.1.1.1.2. The WP specific focus in on the degree to which the respective methods contribute to the goals of trustworthy AI, explainability and the ability to involve humans in the learning, reasoning and planning process.

Example microprojects that could begin immediately after initiating the project

Microproject (1) : Build a system that combines standard statistical opaque recommender techniques with knowledge about domain specific narratives. As and evaluation domain we will consider systems to **recommend educational paths for life-long learning (in cooperation with T**, retraining and requalification. This domain is urgent because of longer work life, a more dynamic labour market, and the disruptive effects of AI technologies. The recommender techniques that suffice to recommend the next movie to watch or the next book to buy will not suffice to convince people to invest in a particular learning trajectory, hence more insightful explanations and joint human/machine reasoning based on educational and career narratives will be required. This project can build on work such as [Liang2016], [Wang2019a], [Cao2019] and others, as well as leveraging the results of the X5Gon project (www.x5gon.org) coordinated by a project partner.

Microproject (2): Learning the compositional structure of environments. The challenge in this microproject is learning to recognize relations between objects in images or videos and leverage this to improve understanding of situations. This can build from identifying simple relations to more complex interactions and symbolic structures that most characterize the situation. The important aspect that will be explored is the co-learning of both the structures and the individual actors/identities. This compositional image labeling, moving from using ontologies as symbolic priors to relation-learning, and/or (additionally/alternatively) learning the ontological knowledge (the predicate vocabulary) from the images. The research would proceed from simple (learning relations between digits in handcrafted datasets) to complex (recognising relations between images in street scenes for city services, processes, and interactions). The outcome would be a prototype of an AI4EU microservice for labeling relations in images.

Deliverables (brief description and month of delivery)

M09 Contribution of work on Learning Reasoning and Planning with Human in the Loop to **D6.1** (First report on research results, their application significance and the resulting evolution of the research agenda).

M12 D1.1 First year microproject results (papers, tools, datasets) deposited for general use on the AI4EU platform (with contributions from WP 2 and 3)

M21 Contribution of work on Learning Reasoning and Planning with Human in the Loop to **D6.2 (**Second report on research results, their application significance and the resulting evolution of the research agenda).

M24 Contribution of work on Learning Reasoning and Planning with Human in the Loop to second year microproject results **D2.1** (papers, tools, datasets) to be deposited for general use on the AI4EU platform

(compiled by WP 2)

M36 Contribution of work on Learning Reasoning and Planning with Human in the Loop to **D6.3** (Final report on research results, their application significance and the resulting evolution of the European research agenda beyond project end).

M36 Contribution of work on Learning Reasoning and Planning with Human in the Loop to third year microproject results **D3.1** (papers, tools, datasets) to be deposited for general use on the AI4EU platform (compiled by WP 3)

WP 2 Lead INRIA	**Multi Modal Perception and Modeling**
M01-36	**1 (DFKI): 18.0 | 5 (ATHENA): 5.9 | 9 (CNR): 8.3 | 10 (CNRS): 5.9 | 12 (CU): 3.7 | 21 (INESC TEC): 9.9 | 23 (INRIA): 19.7 | 25 (JSI): 13.2 | 28 (ORU): 5.4 | 31 (SORBONNE): 3.3 | 37 (TUBITAK): 8.4 | 40 (TU WIEN): 16.7 | 42 (UCPH): 8.1 | 43 (UGA): 8.2 | 44 (ULEI): 2.5 | 46 (UNIBO): 3.4 | 47 (UNIPI): 4.7 | 48 (UCL): 10.3 | 49 (WARSAW): 9.4 | 50 (UOS): 5.8 | 52 (UVB): 6.6**

Objectives

Our ambition it to build on recent progress in discriminative and generative networks, to provide integrated multi-modal perception and modeling that combines fast real-time reaction for sensori-motor reflexes, with spatiotemporal and geometric reasoning, prediction of recurrent events and consequences for actions and dynamic processes and linguistic expressions for perceptual concepts to enable communication with and learning from humans. In prticuar we intend to develop systems that can understand complex human actions, motivations and social settings.

Description of work

This workpackage will provide enabling technologies for multi-modal perception and modeling of modeling of objects environments and processes (T2.1), individuals, actions, activities and tasks, (T2.2), awareness, emotion and attitude (T2.3), Social Signals and Social Dynamics (T2.4), and Distributed Perception and modeling (T2.5). Dedicated efforts will be devoted to assembling and publishing labeled training data (T2.6) and benchmark datasets (T2.7), for challenges described in a continuously maintained research agenda (T2.8) responding to requirements arising from research and innovation actions (T2.9).

T2.1: Learning of multimodal models grounded in physical reality (DFKI)

As described in section 1.3.1.3.1 this task addresses the problem of learning models that integrate perception from visual, auditory and environmental sensors to provide structural and qualitative descriptions of objects, environments, materials, and processes to provide context for perception of objects, events, and actions. An important challenge will be creation of perceptual concepts with linguistic labels under the guidance of a human tutor.

An example of a microproject would be development of techniques to model the layout and contents of a kitchen workspace including abilities to detect and recognize tools, food stuffs, work surfaces, and kitchen appliances, and to assign operational capacities and affordances to tools and appliances through interaction with a human tutor.

T2.2: Multimodal perception and modeling of actions, activities and tasks (INRIA)

This task will respond to the requirement of section 1.3.1.3.2 by providing techniques that can recognize human actions and place them in the context of an activity or task, with predictions of the intended and actual consequences of the action, and explanations for the purpose of the action. An important challenge will be recognizing and modeling actions in unstructured real world environments from visual, auditory and other perceptual channels, and interpreting actions when the task and context are not known apriori.

An example of a microproject would be a system that monitors activity in a natural human workspace such as a kitchen or dining area, completing the description of each action with a description of the intended

purpose of the action, likely consequences, and predictions of future actions.

T2.3: Multimodal perception of awareness, emotions, and attitudes. (INRIA)

This task will provide integrated tools that allow systems to perform real time perception and modeling of awareness, emotion and attitude. This task will build on recent progress in remote eye tracking and visual and auditory perception of valence, arousal and dominance to provide robust integrated perception and modeling of awareness and emotion to enable shared attention and collaboration (section 1.3.1.3.3).

An example of a microproject would be development of tools to observe and assist people with different levels of expertise engaged in solving problems in domains such as chess, math, or circuit design.

T2.4: Perception of Social Signals and Social Dynamics (SORBONNE)

This task will address parts of the research agenda devoted to making AI systems aware of subtle social aspects of human interactions, including the ability to model and reflect on their impact of such social aspects. More detailed description of the aims and approach is given in section (1.3.1.3.4).

T2.5: Distributed Collaborative Perception and Modeling (UNIPI).

This task will seek to advance research in the area of collaborative perception with different agents (some of them AI systems some possibly human cooperating in the interpretation of a situation and in building and enhancing their respective world models (see section 1.3.1.3.5)

T2.6: Dealing with lack of labeled training data (DFKI)

This task is dedicated to alleviating the training data problem with respect to perception of complex human activities and real life situations. It will follow the approach described in section 1.3.1.3.6 and not just work on methods for collecting data and reducing the need for training data but provide concrete datasets and tools for creating/augmenting datasets (see also T2.7)

T2.7: Assembling benchmark datasets (SUSSEX)

Sussex will guide the collation and curation of multimodal benchmark datasets for multimodal perception and modeling from a variety of sources. The purpose of benchmark datasets is to serve as data sources to challenge AI in specific scientific, technological or ethical aspects. The objectives of this task are:

i) When dedicated datasets are collected by project partners, to provide expertise to guide the creation of these datasets so that they are reusable by the wider scientific community (future-proofing datasets).

ii) Out of the datasets emerging from the consortium, to assemble them in an accessible way for the wider community (e.g. through publications describing these datasets, online visibility, public activities, etc).

iii) When particular AI aspects need to be evaluated and challenged, to identify datasets among those in the wider scientific community, or from within the "in-house" dataset archives of the project.

The University of Sussex (group of Prof Roggen) has a significant expertise establishing large scale multimodal benchmark datasets. Many datasets originating from this research group having become well established datasets in multimodal perception and modeling and HCI (see partner description).

T2.8: Consolidation and coordination of the research agenda (INRIA)

This task implements the consolidation and coordination function for the research agenda of this WP according to the approach sketched in section 3.1.1.1. WP specific issues are in particular the importance of curating, preserving and making available to the community datasets that most of microproject will produce and dea with in one way or the other. Another is the close connection to both WP 1 (in terms of respective learning methods) and WP 3

T2.9: Responsible Research and Innovation Assessment (RRIA) (INRIA)

This task provides WP specific RRIA support according to the approach sketched in described in section section 3.1.1.1.2. The WP specific aspects are in particular issues associated with data collection including data related to human subjects.

Example microprojects that could begin immediately after initiating the project

Microproject (challenge) (1): Narrative Description and Assistance for Kitchen Activities.

The kitchen provides the arena for a rich variety of human activities including cooking, cleaning and social interaction. Kitchens offer semi-structured environments with a well defined set of tools, appliances and workspaces. Common kitchen activities follow loosely defined scripts that are often performed as routines. Kitchens are the seen of creative interpretation for loosely defined menus that prescribe steps in food preparation that leave extensive latitude for improvisation. Thus kitchens make an ideal arena for microprojects for developing abilities for multimodal perception and modeling of environments, actions, awareness, emotions and collaborative services.

Challenges for microprojects include developing methods to construct narrative descriptions of food preparation, monitoring of cleaning activity, and enabling technologies for a collaborative agent that monitors cooking activities and offer warnings of problems and to respond to spoken requests for assistance or advice using online information.

An interesting challenge would be to derive information from textual HowTos that you can find online using NLP techniques, such as semantic structure analysis, semantic role labeling, discourse and coreference resolution. Set up an experiment (with a simple activity), evaluate the improvement in recognition rates when using the model derived from language (Sensor based activity recognition (Paul Lukowicz DFKI), NLP (Jan Hajic Prague), Knowledge representation (Thomas Eiter Vienna), Vision (James Crowley INRIA) Simulation-Based Training and Validation (Slusallek, DFKI).

Microproject (challenge) (2) : Challenge in computational behavioral analytics with AI and sensors. Computational behavioral analytics refers to the automated analysis and eventual understanding of human activities from the data originating from miniature sensors (such as those found in wearables or in smart environments), which are interpreted using advanced AI and machine-learning techniques.

A challenge will be organised within the cohort, where a state of the art dataset will be provided to the cohort (e.g. the SHL dataset developed by the University of Sussex, and which was already used in ML challenges in 2018 and 2019) and the challenge task will be to apply advanced ML techniques - selected by the participants - to analyse the data of sensors, over a period of one week.

Finally, a presentation and seminar will conclude this challenge, where each of the cohort members gets to present the approach they devised, their findings and challenges. This will be a moderated session where participants get to share their experience and learn from their colleagues.

Deliverables (brief description and month of delivery)

M09 Contribution of work on Multi Modal Perception and Modeling to **D6.1** (First report on research results, their application significance and the resulting evolution of the research agenda).

M12 Contribution of work on Multi Modal Perception and Modeling to first year microproject results **D1.1** (papers, tools, datasets) to be deposited for general use on the AI4EU platform (compiled by WP 1)

M21 Contribution of work on Multi Modal Perception and Modeling to **D6.2 (**Second report on research results, their application significance and the resulting evolution of the research agenda).

M24 D2.1 Second year microproject results on learning, reasoning, perception and interaction (papers, tools, datasets) deposited for general use on the AI4EU platform (including contributions from WP 1 and WP 3)

M36 Contribution of work on Multi Modal Perception and Modeling to **D6.3** (Final report on research results, their application significance and the resulting evolution of the European research agenda beyond project end).

M36 Contribution of work on Multi Modal Perception and Modeling to third year microproject results **D3.1** (papers, tools, datasets) to be deposited for general use on the AI4EU platform (compiled by WP 3)

WP 3 Lead Aalto	**Human AI Interaction and Collaboration**
M01-36	**1 (DFKI): 9.0 | 2 (AALTO): 19.4 | 5 (ATHENA 11.7 | 6 (BRNO U): 4.8 | 9 (CNR): 8.3 | 10 (CNRS): 6.8 | 12 (CU): 11.1 | 16 (FBK): 10.7 | 23 (INRIA): 19.7 | 24 (IST): 7.3 | 27 (LMU): 13.3 | 28 (ORU): 5.4 | 31 (SORBONNE): 13.3 | 32 (STICHTING): 12.0 | 35 (TILDE): 3.0 | 38 (TU DELFT): 16.2 | 40 (TU WIEN): 10.0 | 42 (UCPH): 16.1 | 43 | 44 (ULEI): 1.7 | 45 (UMU): 9.3 | 48 (UCL): 15.4 | 49 (WARSAW): 9.4 | 50 (UOS): 2.9 | 51 (UPF): 4.9**

Objectives

This work package aims to establish new methodological and conceptual basis for human-AI collaboration. As described in Section 1.3.1.4, the goal is to develop 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 **main objective** of this WP is machine-learning methods and suitable interaction techniques based on theoretically grounded models of human-human communication, which can drive the inference and planning of an AI agent in a more human way and with less training data. These models include models of multimodal communication, for grounding, theory of mind, and emotion. They work with interaction histories collected over a longer timespan and over a richer set of sensors than previously.

Description of work

Intelligent systems 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 WP 2. As cases we will look at human-robot interaction and interactive agents.

T3.1 Foundations of Human-AI interaction and Collaboration (INRIA)

Following the approach outlined in section 1.3.1.4.1 this Task will advance the research agenda with respect to basic questions underlying the HumanE AI Net human-AI collaboration and interaction concept including emotion expression and group cognition.

T3.2 Human-AI Interaction/collaboration paradigms (LMU)

As described in section 1.3.1.4.2 this task will leverage the the basic understanding involved in our new approach to Human AI collaboration to design, implement and evaluate specific interaction paradigms as described in section 1.3.1.4.2.

T3.3 Reflexivity and Adaptation in Human AI collaboration (UCL)

As sketched in section 1.3.1.4.3 this task will focus on computational rationality modeling, which can be used to infer human reward functions, emotions, beliefs, in order to enable AI to estimate human reactions to events and development over time.

T3.4 User Models and Interaction History (STICHTING)

As described in section 1.3.1.4.4 this task develops user models that allow the AI to understand the behavior of a human partner in the light of a longer history of collaborative actions. This is necessary for the inference of more stable, but latent, traits affecting behavior. In particular, we will develop user models for understanding emotion, as linked to cognition and social behavior. These models will help AI to infer and express emotions with a human partner. We will secondly develop models for group cognition, especially grounding and theory of mind, which will help an AI partner infer and express beliefs to a human partner. These models can take the form of, for example, bounded rationality models, Bayesian belief models, or similar. We pursue high transparency in the AI's model of the user, as well as principled ways to deal with uncertain, multimodal and ambiguous signals in communication with humans.

T3.5 Visualization Interactions, and Guidance (TU Wien)

As outlined in section 1.3.1.4.5 visual interaction and guidance remains an important paradigm for helping users handle complex issues and systems. We will Visual Analytics, interaction, and Guidance techniques ease and support to interact, guide, and enrich the HumanE AI net vision of Human-AI collaboration and co-creation interfacing to advanced learning and reasoning systems and allowing humans to not just understand but also influence and guide such processes,.

T3.6 Language-based and Multilingual Interaction (CU, BUT, DFKI)

This task will focus 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. It will be closely connected to the more technology- and platform-oriented multilingual technologies task in WP 6 (T6.4). It will involve close cooperation with the European Language Grid (see T9.7).

T3.7 Conversational, Collaborative AI (IST)

This task develops models for human-human communication that can be used in collaborative turntaking

situations, such as in conversation or in cooperative activities. As described in 1.3.1.4.6 this is a key research topic in order to enhance human reflection on the actions they are carrying out (e.g. decision-making, problem solving) through having a small dialogue with the AI system

T3.8 Trustworthy Social and Sociable interaction (Warsaw)

This task develops methods that build on the previously listed tasks by exposing, communicating, and making transparent the beliefs of the AI about the human (see section 1.3.1.4.7). This goes beyond the state of the art by allowing explicit communication of beliefs about the human partner in a manner understandable in the social setting.

T3.9: Consolidation and coordination of the research agenda (Aalto)

This task implements the consolidation and coordination function for the research agenda of this WP according to the approach sketched in section 3.1.1.1. A specific challenge is the fact that the work in this Wp needs properly relate to and leverage the results of WP 1 and 2. At the same time much of the involved partners are from the HCI community rather then the core AI community making collaboration and coordination more challenging.

T3.10: Responsible Research and Innovation Assessment (RRIA)

This task provides WP specific RRIA support according to the approach sketched in described in section 3.1.1.1.2. The spec ific challenge within WP 3 is the fact that this WP needs to work with human subjects in many cases preferably in real world settings which always involves considerable ethical and regulatory challenges.

Example microprojects that could begin immediately after initiating the project

Microproject (1) : Interactive Reflective human-AI systems

This project entails meta reasoning between the human and the AI system, where they can ask together or each other “are we doing the right thing?”, “Is it ethical what we are suggesting?” The goal is to enhance reflection through having a small dialogue at particular times. The main outcome will be a new interaction technique that can be used for a variety of contexts and application areas, from bike GPS systems, education agents to hospital AI workflow systems. A number of papers could be published from studies conducted using this interaction technique versus existing baselines.

To begin we will develop a chatbot that can support the switching between the two levels of reflection. The agent could be reflecting about the dynamic with the human and also how it has learned. It could prompt the human to do this, too. And the human can prompt the AI system.

Microproject (2) : Empathy in human-AI systems

This project considers how AI systems can exhibit computational empathy toward a human user and what are the effects of such system behavior. Specifically we will consider a scenario in which an AI system is part of the daily life of a chronically ill patient (e.g. diabetes). Having access to personal information and advising the patient in the day-to-day management of the chronic disease, the question is raised if the collaboration with the AI system would benefit from an empathetic human-AI system. Being empathetic in this context refers to an understanding of how the other person is feeling. Key research questions will be (1) how to design an AI system that is empathic, which micro-interactions are needed for implementing (a sense of) empathy in the AI system and (2) what effect does interacting with the empathic AI system have on the human. Concrete results to be achieved will include (1) proof-of-concept demonstrator of an empathetic AI system, quantitative measures of the empathetic AI systems’ effect and qualitative measures of adherence, acceptance, and how the user perceives the system.

Microproject (3) : Minutes of Multi-party Multi-lingual Meetings

This project will involve a set of traditional language and multimodal interaction tasks, such as automatic speech recognition, speaker identification, language detection, face recognition, speech segmentation, machine translation, summarization, reasoning and inference, question answering, named entity recognition and linking, etc., interconnected in a way that enables to produce coherent, unbiased, noise-free minutes of such complex meetings. We envisage a setting where the meeting is being recorded by a spatial microphone grid and possibly also videorecorded from several angles. While preceptory tasks will provide basic attributes, such as who speaks when (and appropriate transcripts of what they said, using advanced speech and face/identity/language recognition and detection techniques), we will focus here on (possibly multi- and cross-language) understanding on what has been said, assigning arguments to statements, creating argument flow, and relating entities and events to real world knowledge, such as relevant ontologies, meeting documents, etc. As a proof of such understanding, the system will produce not only comprehensive but “edited” minutes, but also connect the minutes to conclusions and explanations, as much as it is able to

produce from its internal understanding.

Deliverables (brief description and month of delivery)

M09 Contribution of work on Human AI Interaction and Collaboration to **D6.1** (First report on research results, their application significance and the resulting evolution of the research agenda).

M12 Contribution of work on Human AI Interaction and Collaboration to first year microproject results **D1.1** (papers, tools, datasets) to be deposited for general use on the AI4EU platform (compiled by WP 1)

M21 Contribution of work on Human AI Interaction and Collaboration to **D6.2 (**Second report on research results, their application significance and the resulting evolution of the research agenda).

M24 Contribution of work on Human AI Interaction and Collaboration to second year microproject results **D2.1** (papers, tools, datasets) to be deposited for general use on the AI4EU platform (compiled by WP 2)

M36 Contribution of work on Human AI Interaction and Collaboration to **D6.3** (Final report on research results, their application significance and the resulting evolution of the European research agenda beyond project end).

M36 D3.1 Third year microproject results on learning, reasoning, perception and interaction (papers, tools, datasets) deposited for general use on the AI4EU platform (including contributions from WP 1 and WP 2)

WP 4 Lead UNIPI	**Societal AI**
M01-36	**6 (BRNO U): 4.8 | 9 (CNR): 16.5 | 24 (IST): 7.3 | 30 (SAP): 1.5 | 31 (SORBONNE): 3.3 | 44 (ULEI): 4.2 | 45 (UMU): 9.3 | 46 (UNIBO): 5.4 | 47 (UNIPI): 23.4 | 49 (WARSAW): 28.2 | 50 (UOS): 1.9**

Objectives

This work package aims at shaping the research on the **societal dimension of AI**, as increasingly **complex socio-technical AI systems** emerge, made by (explicitly or implicitly) interacting people and intelligent agents as described in section 1.3.1.5. It aims to address the **undesired emerging network effects** of social AI systems, as well as the **design of transparent mechanisms for decentralized collaboration** and **decentralized personal data ecosystems** that **help toward desired aggregate outcomes**, i.e., toward **the realization of the agreed set of values and objectives at collective level**, such as accessible and sustainable mobility in cities, diversity and pluralism in the public debate, fair distribution of economic resources, environmental sustainability, a fair and inclusive job market.

Description of work

The conceptual approach to the study of the societal dimension of interactive AI is explained in detail in section 1.3.1.5. The partitioning of this WP into tasks follows the structure of the conceptual approach (as given by subsections 1.3.1.5.1-1.3.1.5.4) adding tasks for the coordination and consolidation of the research agenda and RRIA.

T4.1: Graybox models of society scale, networked hybrid human-AI systems (CEU)

This task addresses the question of modeling and understanding large (society) scale complex hybrid systems consisting of a mix of AI agents and humans. As described in section 1.3.1.5.1 our aim is to develop modeling methodologies that combine complex systems models AI approaches (in particular data driven ones) into so called gray box models (midway between data-driven “blackbox” and mathematical “whitebox” methods.

T4.2: Individual vs. collective goals of AI systems (CNR)

This task addresses the question of how to approach social dilemmas that occur when there is a conflict between individual and public interest. As described in section 1.3.1.5.2 on which this task is based such problems may appear in hybrid Human-AIs society scale systems with additional difficulties due to the relative rigidity of the trained AI system on the one hand and the necessity to achieve social benefit and

keeping the individuals interested on the other hand.

T4.3: Societal impact of AI systems (UNIBO)

This task deals with the evaluation of the societal impact of competing AI technologies as outlined in section 1.3.1.5.3. this will include in-vitro experiments, and mathematical and simulational models. In particular we will consider the impact of societal scale AI on governance, social cohesion, conflicts and conflict resolution.

T4.4: Self-organized, socially distributed information processing in AI-based techno-social systems (Warsaw)

As described in section 1.3.1.5.4 this task addresses the part of the research agenda devoted 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). Key results will be methods for enhancing distributed information processing in socio-technical systems so that they provide a platform for common action toward both individual and collective benefit.

T4.5: Consolidation and coordination of the research agenda (UNIPI)

This task implements the consolidation and coordination function for the research agenda of this WP according to the approach sketched in section 3.1.1.1. A key concern will be integrating the different communities involved in this WP: complex systems, social science, Ai and HCI.

T4.6: Responsible Research and Innovation Assessment (RRIA) (UNIPI)

This task provides WP specific RRIA support according to the approach sketched in described in section 3.1.1.1.2. WP specific challenges are related to the profound potential social impact of the work and the related ethical questions.

Example microprojects that could begin immediately after initiating the project

Microproject (1): Network effects of mobility navigation systems.

Study of emergent collective phenomena at metropolitan level in personal navigation assistance systems with different recommendation policies, with respect to different policies for navigation recommendations and different collective optimization criteria (fluidity of traffic, safety risks, environmental sustainability, urban segregation, response to emergencies, …). Modeling self-organizing decentralized mobility systems to explore the balance between individual and social benefit. Result: big data-driven simulations, research papers. Participants: UNIPI, CNR, Generali, ETHZ, Volkswagen

Microproject (2): Characterize the behavior of a distributed AI system on top of a social network. What is the effect of the topology of an underlying social network on the outcome of specific distributed AI problems (e.g., distributed classification)? result: scientific paper(s) Partners: CNR Pisa (decentralised learning, social networks), CEU (modeling complex systems), Univ. Pisa (social networks), University of Warsaw (social networks).

Microproject (3): Social norms to counteract misinformation in human-AI hybrid systems. Misinformation, fake news and opinion polarization may lead people to become insensitive to information that contradicts their own existing position. But what if this effect is not only due to conformity and individual preferences, but is also due to the pressure of complying with the social norms of the group? We propose to 1) investigate the extent to which existing norms are "obstacles" to consensus formation (e.g., pluralistic ignorance) and 2) provide cases where social norms can be integrated in AI system as "catalyst" of behavior change (norm-based interventions).

Deliverables (brief description and month of delivery)

M09 Contribution of work on Societal AI to **D6.1** (First report on research results, their application significance and the resulting evolution of the research agenda).

M12 D4.1 First year microproject results on societal, ethical and responsible AI (papers, tools, datasets, best practice guidelines) deposited for general use on the AI4EU platform (including contributions from WP 5)

M21 Contribution of work on Societal AI to **D6.2 (**Second report on research results, their application significance and the resulting evolution of the research agenda).

M24 Contribution of work on Societal AI to **D5.2 (**Second year microproject results on societal, ethical and responsible AI (papers, tools, datasets, best practice guidelines) to be deposited for general use on the AI4EU platform, coordinated by WP 5)

M36 Contribution of work on Societal AI to **D6.3** (Final report on research results, their application significance and the resulting evolution of the European research agenda beyond project end).

M36 D4.2 Final microproject results on societal, ethical and responsible AI (papers, tools, datasets, best practice guidelines) deposited for general use on the AI4EU platform (including contributions from WP 5)

WP 5 Lead UMEA	**Ethics, Law and Responsible AI**
M1-M36	**7 (BSC): 6.7 | 8 (CEU): 10.1 | 9 (CNR): 16.5 | 10 (CNRS): 2.5 | 30 (SAP): 3.0 | 38 (TU DELFT): 16.2 | 45 (UMU): 18.7 | 46 (UNIBO): 2.7 | 47 (UNIPI): 7.0 | 50 (UOS): 1.9 | 52 (UVB): 29.3**

Objectives

This WP is dedicated to ensuring that AI systems operate within a moral and social framework, in verifiable and justified ways as elaborated in section 1.3.1.6). 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 ethical and societal principles. This concerns legal, ethical, trustworthy aspects but need to be combined with robustness, social and interactivity design. The focus here is the prioritization of ethical, legal, and policy considerations in the development and management of AI systems to ensure responsible design, production and use of trustworthy AI. This requires integration of engineering, policy, law and ethics approaches. This topic is thus about understanding, developing and evaluating ethical agency and reasoning abilities as part of the behavior of artificial autonomous systems (e.g. artificial agents and robots). We will focus on explanation aspects and core data protection principles of fairness, transparency, accountability and responsibility.

Description of work

Ethics by design' methods will be investigated that aimed at understanding how can values be 'wired' into socio-technical systems and what it means to do so. These may include (but are not limited to) Fairness, non-Discrimination, Compliance, Security, Data Protection and Privacy by Design, and how to implement these in combination with AI techniques and algorithmic governance through formal analysis and representation of regulatory principles, allocating rights, distributing liability, and ensuring legal protection by design.

Even though AI systems are increasingly able to take decisions and perform actions that have moral impact, AI systems are artefacts and therefore are neither ethically nor legally responsible. Individual humans or human corporations should remain the moral (and legal) agent. We can delegate control to purely synthetic intelligent systems without delegating responsibility or liability to them. To this effect, computational and theoretical methods and tools will be investigated, that support the representation, evaluation, verification, and transparency of ethical deliberation by machines with the aim of supporting and informing human responsibility on shared tasks with those machines. Research is needed to understand what suitable constraints on system behavior are, and to elicit desiderata on the representation and use of moral values by AI systems.

T5.1: 'Legal Protection by Design' (LPbD) (VUB)

This task will address the question of incorporation of fundamental rights protection into the architecture of AI systems including (1) checks and balances of the Rule of Law and (2) requirements imposed by positive law that elaborates fundamental rights protection. A key result of this task will be a report on a coherent set of design principles firmly grounded in relevant positive law, with a clear emphasis on European law (both EU and Council of Europe), part of D5.3. It will contain

- A sufficiently detailed overview of legally relevant roles, such as end-users, targeted persons, software developers, hardware manufacturers, those who put AI applications on the market, platforms that integrate service provision both vertical and horizontal, providers of infrastructure (telecom providers, cloud providers, providers of cyber-physical infrastructure, smart grid providers, etc.);
- A sufficiently detailed legal vocabulary, explained at the level of AI applications, such as legal subjects, legal objects, legal rights and obligations, private law liability, fundamental rights protection;

- High level principles that anchor the Rule of Law: transparency (e.g. explainability, preregistration of research design), accountability (e.g. clear attribution of tort liability, fines by relevant supervisors, criminal law liability), contestability (e.g. the repertoire of legal remedies, adversarial structure of legal procedure).

T5.2: Empirical study of LPbD aspects of real life projects (VUB)

VUB (LSTS) will visit 3-5 of the microprojects, to engage in a kind of constructive technology assessment, interacting with the developers of the projects, teasing out potential risks for the rights and freedoms of natural persons who may suffer the consequences of implementation. One result of those studies will be a detailed ending in a set of recommendations on how to integrate legal protection by design into the architectures developed in the microprojects (D5.2). This will include: (1) an overview of the research design in terms of training & validation data, feature space, hypothesis space, machine readable tasks, performance metrics, out of sample testing; (2) an overview of the types of bias that may occur due to the type of training data used, the labeling (if relevant), the models trained and the way they are employed; (3) an assessment of the types of risks to fundamental rights and freedoms that may occur due to the implementation of AI research in real world situations; (4) an overview of types of legally relevant explanations that enable individual persons to object against algorithmic decision-making that could infringe upon their rights and freedoms; (5) a set of mitigating measures, such as e.g. data protection by design, to reduce infringements and to prevent violations.

Part of the funds for the "dynamic microprojects" will be assigned to supporting microprojects in participating in work related to this task. Also each microproject will have as part of its obligations the availability to devote time to LPbD assessment.

T5.3: 'Ethics by design' for autonomous and collaborative, assistive AI systems (CNR Pisa)

This task deals with understanding how values can be *'wired'* into socio-technical systems including issues related to **Compliance, Security, Data Protection and Privacy by Design, Fairness, Explainability** and how to implement these in combination with AI techniques and algorithmic governance. A special focus will be devoted to link research in WP1, and WP2 with : (1) methods to design principles for meaningful human control over autonomous AI systems, (2) methods for evaluating and measuring explicability in high-stakes AI-decision making based on human-machine interaction capable of revealing *causality and counterfactuals*. (3) methods for discrimination and segregation discovery as well as protection of novel vulnerabilities, (4) feedback methods to inform policy-makers and regulators on missing elements in current regulations and practices.

Examples of microprojects include: derivation of quantifiable criteria from high-level ethical values to be used as non-functional requirements to design human-AI complex systems; investigation of how complex interactions in the human-AI ecosystem are shaped by the specification of those values as non-functional requirements; formal specification of values that would allow for the automatic detection of conflicts between values and verify the system's abidance to values.

T5.4: "Ethics in design": methods and tools for the responsible development of AI systems. (TU Delft)

This task is devoted to methods and tools for the value-based design and development of AI systems that ensure (a) the analysis and evaluation of ethical, legal and societal implications; (b) the participation and integrity of all stakeholders as they research, design, construct, use, manage and dismantle AI systems; (c) the governance issues required to prevent misuse of these systems, and (d) means to inspect and validate the design and results of the system, such as formal verification, auditing and monitoring. This will include: (1) methods to elicit and align multi-stakeholder values and interest and constraints, (2) methods to integrate and validate a combination of different possibly conflicting values and integrate them into the computational solutions, (3) tools to support the contextual definition and the verification and validation of system's properties: robustness, accountability, explainability, responsibility and transparency

T5.5: Support of RRIA of Tasks 2, 3, 4, 6, 8 (UMEA)

This task will provide ethical and legal support of the RRIA tasks of the other research WPs. This will include a 2-day tutorial (VUB (LSTS) attended by a senior researcher of each partner. The following themes will be explained: difference between law and ethics, difference between legal norms and computer code, introduction to fundamental rights that are relevant in the context of AI, introduction to impact assessments of potential infringements of fundamental rights by AI applications and infrastructure and mitigating measures, introduction to legal obligations to implement impact assessments and legal protection by design, liability and enforcement measures in case of violation of fundamental rights due to AI applications. It will also include best practice guidelines and interactive counseling when required. This Task will work closely with the **Board for Operational Ethics and Legality (BOEL, see section 5)**

T5.6: Consolidation and coordination of the research agenda (UMEA)

This task implements the consolidation and coordination function for the research agenda of this WP according to the approach sketched in section 3.1.1.1.. A challenge is the combination of direct research in Task 5.1 with the microproject oriented approach of tasks

Example microprojects that could begin immediately after initiating the project

Microproject (1): AI based assistive technologies. What are the moral limits of nudging by moral AI assistive technologies? For the system to verify the adherence to values as well as detect conflicts between values (in this case, privacy versus well-being), we will evaluate a formal specification of values and verify the system's adherence to these values. We will further analyse means of 'algorithmic recourse': tools for public consultation and contestation.

Microproject (2): Ethical games. Here we will focus on how to address the pitfalls of crowdsourcing ethical decisions and using ML on this data. Aim is to design ethical games and engage citizens to play with them. This will generate data about people's hypothesis and choices and allow to choose rules accordingly and provide material to think about how to embed them in political and policy decision making.

Microproject (3): Explanatory Tool for Clinical Analysis of Patients. The explanation of decision-making systems is particularly important in health. Testing the confidence level of AI decision systems such as *Doctor AI,* where predictions may be multi-label and explanations should be multimodal (text, images, narraatives) and targeting a variety of stakeholders (practitioners and patients), with a particular focus on specific applications and classes of diseases.

Microproject (4) Formal Specification of Values. Say a person refuses to share their data on social networks with other parties (privacy value), but the ML algorithm learns that this person is suicidal (well-being value). In this microproject, we will study the formal specification of values, which allows for the automatic verification of the system's adherence to values and the detection of conflicts between values.

Deliverables (brief description and month of delivery)

M02 D5.1: Tutorial on Legal Protection by Design (LPbD) to be given at the project plenary meeting.

M09 Contribution of work on AI Ethics and Responsible AI to **D6.1** (First report on research results, their application significance and the resulting evolution of the research agenda).

M12 Contribution of work on AI Ethics and Responsible AI to **D4.1 (**First year microproject results on societal, ethical and responsible AI (papers, tools, datasets, best practice guidelines) to be deposited for general use on the AI4EU platform, coordinated by WP 4)

M21 Contribution of work on AI Ethics and Responsible AI to **D6.2 (**Second report on research results, their application significance and the resulting evolution of the research agenda).

M24 D5.2 Second year microproject results on societal, ethical and responsible AI (papers, tools, datasets, best practice guidelines) deposited for general use on the AI4EU platform (contributions from WP 4)

M30 D5.3 Report on Ethical, Legal and Responsible AI concepts, design principles, best-practices and tools, including a report on for LPbD (impact assessment and mitigation measures) and a set of lessons learnt

M36 Contribution of work on AI Ethics and Responsible AI to **D6.3** (Final report on research results, their application significance and the resulting evolution of the European research agenda beyond project end).

M36 Contribution of work on AI Ethics and Responsible AI to **D4.2 (**Final microproject results on societal, ethical and responsible AI (papers, tools, datasets, best practice guidelines) to be deposited for general use on the AI4EU platform, coordinated by WP 4)

WP 6, Lead DFKI	**Applied research with industrial and societal use cases**
M01-M36	**1 (DFKI): 13.5 | 3 (AIRBUS): 8.8 | 4 (Algebraic AI): 2.7 | 5 (ATHENA): 8.8 | 7 (BSC): 4.0 | 9 (CNR): 8.3 | 12 (CU): 3.7 | 15 (ETHZ): 15.0 | 16 (FBK): 10.7 | 17 (FORTISS): 10.0 | 18 (FRAUNHOFER): 6.2 | 19 (Generali): 6.9 | 22 (ING): 9.0 | 29 (PHILIPS): 15.0 | 30 (SAP): 7.5 | 34 (TID): 15.9 | 35 (TILDE): 12.0 | 37 (TUBITAK): 8.4 | 39 (TUK): 32.5 | 47 (UNIPI): 4.7 | 50 (UOS): 1.9 | 53 (VW AG): 7.5**

Objectives

This WP is dedicated to synchronising the research agenda and network activities with industrial and social needs. The three main objectives are (1) ensuring that the needs of important European industry are adequately then into account within the research agenda, (2) making sure that key results are evaluated in industrially (and socially) relevant use cases and (3) making sure that the knowledge created by the microprojects of WP1-5 reaches key European industrial players.

Description of work

This WP consists of three groups of tasks. T6.1-6.4 address horizontal issues important for our vision of a European brand of human-centric AI applications (security, multilinguality) and practical platform related concerns (software in cooperation T9.1 on AI4EU and hardware/HPC platforms). Tasks T6.5 to T6.10 focus on vertical application domains, each driven by a European industrial champion. They will be in charge of running the stake holder workshops (together with WP 7) to define the respective domain specific research agenda and will be conducting microprojects related to concrete use cases. Tasks

T6.1 Security Issues (SAP)

This task aims defining the security and privacy challenges for AI enabled applications, as well as the directions for addressing them. It includes investigating new attacks to the AI algorithms (e.g., poisoning attacks on training data), and on the productive AI applications (e.g., inference or membership attacks on the privacy of training data), and the corresponding security countermeasures. It will complement, from a technical perspective, the ethical and legal principles and methods, described in WP5.

6.2 Hardware platforms and resources (BSC, TUK)

This task aims to support HumanE AI Net consortium to bring their implementations to HPC resources. We will support, when necessary, the set-up of HPC efficient environments for the Pilots created in WP1-WP5 (sic). This implies supporting AI components orchestration and its deployment in HPC environments as the future MareNostrum V. We will support the connection with PRACE. T6.2 aims to take the best of HPC for AI, parallelization of the work and orchestration must be adapted and optimized for proper scalability."

Furthermore we will consider and advise the applied research how special purpose hardware and platforms can leveraged to make real life deployment practicable.

T6.3 Software platforms and frameworks (DFKI, BSC, SAP)

In close cooperation with task T9.1 (interface to AI4EU) this task will assist the applied research in selecting the right tools and platforms for the different types of applications and different types of AI methods. I tiwll also investigate the requirements and architectures related to such platforms.

T6.4 Language technology and multilinguality (Tilde, CU,DFKI)

This task will develop and privode key technologies needed to use speech, in particular multi-lingual speech in industrial AI applications. It will be conducted by Tilde , Europe's premier provider of translation and language technologies in close cooperation with WP 3 (in particular T3.6) but also T9.7 (interface to ELG) defining the research agenda for applied language technologies and conducting application related microprojects.

T6.5 Health related research agenda and industrial use cases (Philips)

Taking a holistic view of people's health journeys, starting with healthy living and prevention, precision diagnosis and personalized treatment, through to care in the home, healthcare is positioned as a care continuum. In this task we will explore (in microprojects) AI driven concepts for supporting consumers, patients and healthcare staff throughout the care cycle. These AI concepts contribute to enhanced patient and staff experience, better health outcomes and lower costs of care.

This task will aim to leverage the research in WPs 1-5 to further the above vision in the insurance industry within specific use case microprojects.

T6.6 Mobility/automotive related research agenda and industrial use cases (Volkswagen)

AI clearly has a key role in the development of automated and connected driving. Although it seems unlikely that there will be nothing but a large deep neural network between sensors (radar, Lidar, etc.) and actuators (longitudinal and lateral control), deep learning will be indispensable for environmental perception and maneuver planning. More classical symbolic AI (reasoning, knowledge representation) will also make a significant contribution, especially in the verification and validation of AI modules. Key challenges for this

domain include explainable, verifiable sub-symbolic AI and robust, flexible architectures that can utilize AI modules safely.

In addition to its relevance aboard connected vehicles, in the next few years, AI will play an important role in the evolution of the automotive sector and the field of mobility in general (e.g., drivers, pedestrians, city managers, etc.). Indeed, AI will be a key factor in analyzing and facilitating changes in cities and citizen behaviors related to mobility: as soon as vehicles connect with each other and with road/city infrastructure in real time, interesting opportunities for optimizing the efficiency of the entire mobility system will emerge. For example, in a city environment, nowcasting and optimized route-planning techniques can support drivers and city managers by controlling an entire fleet of vehicles and reducing traffic jams and stationary vehicles, with important advantages related to transport of goods, public transport and personal mobility

T6.7 FinTec related research agenda and industrial use cases (ING)

Financial institutions and insurance companies already heavily rely on AI technology, from automated trading systems to various automated valuation methods to credit-rating tools. Over 70% of all transactions at stock exchanges are traded automatically. The mitigation of systemic risk in the banking system is a central challenge in today's financial system, as the ongoing financial, economic and debt crisis is a major source of instability in many societies (most recently in Turkey). The HumanE AI Net vision can have a profound influence on the use, usefulness and socio-economic impact of AI technologies in the financial sector.

This task will aim to leverage the research in WPs 1-5 to achieve such an impact through specific use case related microprojects.

T6.8 Insurance related research agenda and industrial use cases (Generali)

For an insurance company like Generali, the quality of the customer experience is vital in order to keep a relationship of trust between the company and the customer. Without trust, it's impossible to create the virtuous circle that helps the company not only to better serve its customers, but also to spread the culture of protection in the society. Assicurazioni Generali wants to overtake the 'one policy fits all' vision and is focused in developing new insurance products that are tailored to each customer needs. To achieve the goal of putting the customer at business' center, is essential to adopt technologies that prevents privacy issues by design. Also, to build and maintain a strong relationship of trust with each customer and the company, the application of AI must be transparent to humans.

This task will aim to leverage the research in WPs 1–5 to further the above vision in the insurance industry within specific use case microprojects.

T6.9 Aerospace related research agenda and industrial use cases (Airbus)

Current automation in aviation, such as the autoflight system, is fully deterministic and provides its functionality based on a pre-defined parameter set of system and/or external conditions. Accordingly, the scope of automation offered to the human user is virtually always identical for any given system. Increasingly complex autonomous functions involving AI technology create various challenges with respect to human-machine-interaction, with the well-documented automation ironies and automation awareness aspects as key issues. Even the most complex autoflight systems today are essentially rigid, closed systems with a finite number of functions and modes, and thus fully "learnable" by pilots, whereas particularly future cockpit automation providing an aid in pilot decision making will entail an open and potentially non-deterministic system that is not limited in its proposals. This calls for a completely novel approach toward human-automation interaction and human-machine teaming. Automation will need to provide an understandable rationale for the decisions it proposes to keep the human operator in the loop, and the distribution of tasks between pilot and automation may no longer be static in future single pilot cockpits.

This task will aim to leverage the research in WPs 1-5 to further the above vision within specific use case microprojects. It will also consider further aerospace related use cases including in particular in aircraft production and quality control.

T6.10 Telco related research agenda and industrial use cases (Telefonica)

Within the telco domain value creation increasing moves from the mere provision of connectivity toward value added services, in particular services connected to data and data usage. Such services are only viable if provided within stringent ethical and legal boundaries and in a way that involves and empowers the user. This task will investigate in concrete use cases how to leverage the HuamanE AI Net vision and technology to achieve this in concrete use case related microprojects.

T6.11 AI for Education (TUK)

The European educational system must deal with an increasingly heterogeneous structure of diverse

backgrounds (for example, through migration and inclusion), professional qualifications (through more flexible, evolving careers) and objectives (from regular studies to part-time continuing education). Therefore, future education must be much more individualized and personalized. The question of how digital technologies and AI can help with individual education is currently being widely explored (AI in Education; AIED; reviews and state-of-the-art papers, e.g., [duBoulay2016; duBoulay2018; Luckin2017; Luckin2018; Rosé2018])

This task will aim to leverage the research in WPs 1-5 to develop and demonstrate new solutions for personalized education within specific use case microprojects.

T6.12 AI for social good (ETHZ)

AI (and, more generally, digital transformation) is having a tremendous impact on society and the political process. Dealing with this transformation poses many challenges to which the vision of HumanE AI can make a major contribution. The ability to understand complex social settings and human motivations and feelings as well as the capability of quickly adapting to new situations are all key ingredients needed, e.g., for detecting fake news and bot campaigns against political processes, a problem that democracies are especially vulnerable to. The HumanE AI Net focus on participation and regulation will support social acceptance with ethically acceptable systems, i.e., one that aligns with human values and principles. Society is increasingly more complex, leading to an over-polarized political debate. The notion of value-and-ethics-based AI can help detect hate speech and alleviate the problem of political polarization through information bubbles. It can also facilitate the creation of tools for informed, constructive political debate and help both citizens and policy-makers better understand the complexity of a networked globalized world.

This task will aim to leverage the research in WPs 1-5 to develop and demonstrate how AI can further the above vision within specific use case microprojects.

T6.13: Consolidation and coordination of the research agenda (DFKI)

This task implements the consolidation and coordination function for the research agenda of this WP according to the approach sketched in section 3.1.1.1. The specific concern for WP 6 is to coordinate the setting of the research agenda with the microprojects and to make sure that the relevant advances from the microprojects of WPs 1-5 find their way into follow up evaluation in industrial use cases within WP6

T6.14: Responsible Research and Innovation Assessment (RRIA) (BSC)

This task provides WP specific RRIA support according to the approach sketched in described in section 3.1.1.1.2. The close involvement of industrial use cases will pose a particular challenge but also opportunity. As a consequence there will be especially intensive cooperation with WP 5.

Example microprojects that could begin immediately after we initiate the project

Microproject (1): AI coach for behavioral change. The use of AI-based solutions for coaching people within the healthcare domain significantly increased in the last year. The challenging goal of this project is to induce an attitude and behavioral change toward healthy living style, also exploiting the ability to construct narrative-based systems for healthy living. Addressing the problem of behavior change requires the ability to build persuasive architectures combining techniques tailored to the gathering and analysis of the necessary data, the managing of data and knowledge of all the involved domains, multiturn interactions and more in general conversations, intention recognition, content personalization. The research will be focused on techniques capable of inducing an attitude and behavioral change based on the continuous learning of the observed behaviors and the current narrative integrated with prior knowledge and narratives for health; on producing benchmark datasets; and metrics to measure the effectiveness of the solutions.

Microproject (2): Improving air quality in large cities using mobile phone data and AI.

Mobile phone data analysis can provide actionable insights about traffic and crowd mobility patterns to help the authorities measure and predict pollution in a more cost-effective way, and therefore give valuable information about how to make more efficient public transit system which has the benefit of improving citizen happiness through reducing commuter stress.

In the context of this project, Telefónica plans to work with relevant partners from WPs 1–5 to develop and extend the current prototype applied to the cities of Madrid and Sao Paulo, infusing it with new AI technology from the partners where appropriate, and testing it in other cities in Spain and Germany. If successful, this can be an example for other telecommunications operators in Europe to improve the air quality across Europe.

Deliverables (brief description and month of delivery)

M09 D6.1 First report on research results, their application significance and the resulting evolution of the

research agenda (integrating contributions from WPs 1–5).

M21 D6.2 Second report on research results, their application significance and the resulting evolution of the research agenda (integrating contributions from WPs 1–5).

M36 D6.3 Final report on research results, their application significance and the resulting evolution of the European research agenda beyond project end (integrating contributions from WPs 1–5)

WP 7 Lead GE	**Innovation Ecosystem and Socio-Economic Impact**
M01-36	**7 (BSC): 2.7 | 17 (FORTISS): 3.3 | 18 (FRAUNHOFER): 3.1 | 20 (GE): 13.5 | 22 (ING): 3.9 | 27 (LMU): 17.8 | 50 (UOS): 1.0 | 53 (VW AG): 1.5**

Objectives

The objective of this work package is to maximize the socio-economic impact of the research roadmap of the consortium. This is twofold, (1) providing means and mechanisms to transform basic and applied research results into ventures and businesses that are provide value to European citizens, and (2) to ensure that applied research is guided by real world challenges and steered toward domains that are beneficial for society. In this work package we provide research and provide mechanism that supports the creation of start-ups, the transformation of traditional (non-digital) SMEs into high-tech companies, and to push agile innovation in major industries. A range of dedicated mechanisms in envisioned and will be created, that creates leaders in AI technologies and applications.

Description of work (where appropriate, broken down into tasks), lead partner and role of participants

T7.1 Multiply Successful Mechanisms (FORTISS)

Catalog, describe, and learn from existing and successful support structures and formats for innovation in Europe and around the world. Understand the requirements and how they impact innovation and transformation in start-ups, SMEs, and industries. Replicate, adapt and multiply successful mechanism.

T7.2 Platform for Matching People, Ideas, Research, and Resources (GE)

Bringing people, researchers, developers, customers, and investors together is a key for sustained innovation. Matching people with ideas and research results, with companies looking for specific solutions, and bringing in required resources is essential when moving from ideas and research results to successful ventures. In this task, a platform will be created that allows effective matching on different dimensions for innovation in AI.

T7.3. Innovation Infrastructure, Ecosystem, and Support Formats (GE)

Flexibility, speed, and agility are key when transforming research into products and services. Especially in the areas that are strongly dominated by software and AI this is critical. In this task we will assess what are the appropriate means to create an effective and efficient innovation environment and what components are required to make the innovation eco-system economically successful and societal relevant.

T7.4. Learning, Teaching, and Inspiration for AI Innovation (GE)

AI is changing how innovation is created and assessed. AI is changing what innovations are possible. In this work package we will design educational tools and offer events, that are designed to educate on how AI transforms innovation around the world. Formats include online resources, data bases, but also schools where people learn together at a specific side. The education is targeted at students as well as professionals that are concerned with the creation of new products and services.

T7.5. Regulation for AI innovation (LMU)

As AI innovation in safety critical fields is particularly difficult because the regulations are still unclear. At the same time is difficult to move regulations forward as AI is typically not fully explainable and it is difficult to prove whether a product complies to regulations or not. In this task we aim to strengthen collaboration between industry, startups and research on the one hand and European regulatory authorities on the other. Stakeholders in different industries like in Aviation or the Finance industry will be working with researchers and lawyers in regulatory sandboxes co-creating new products, better AI, meaningful regulations and automated reporting.

T7.6. European Data Hub (LMU)

Meaningful AI application and use-cases require quality data. Getting access to high quality labeled data especially in industry is still a big challenge for most industry players, SMEs and startups, and also for researchers. In this task we aim to make access to data for all easier. We moderate joining forces to create a common European Data Hub. Universities as well as Startups, SMEs and industries will develop means and protocols to share their Data and in return receive full access to Data in the Hub. The Hub will collect open Source Data as well as Data from the partners and will process and manage the Data to make it useable for all stakeholders.

T7.7 AI Innovation Networking Events (GE)

In this task we will co-organize a range of events that will foster an AI Innovation community. Examples of such events include short inspirational events, such as "meet the geek", where researchers with successful tech entrepreneurs. Another example of a longer event is a European Entrepreneurship Summer School. It takes place over seven-day. It brings a diverse group of students together to develop entrepreneurial solutions that meet the world's as well as Europe's biggest challenges. In the spirit of promoting "cooperation, innovation and development", we work on connecting young people in Europe, on the exchange of technological knowledge and on the implementation of future-oriented and sustainable ideas.

T7.8 AI Innovation Accelerators (ING)

In this task we will highlight Lighthouse accelerators, as examples of how to promote human centric AI startups. From existing accelerators with an excellent track record an exclusive set will be selected and show-cased as best practice. A further aim of this is, to network the Lighthouse accelerators across Europe. Furthermore an European Accelerator Program is designed to incorporates the idea to support start-ups in addressing new markets in Europe. The program will have a time limitation between 3 to 6 month and offer free office space as well as support in local hiring processes. Mentors from industrial leaders plus successful entrepreneurs will give guidance to start-ups during the program.

T7.9 AI Innovation Prize (GE)

A European AI Innovation Prize will be established. This prize will have different categories, for start-ups, SMEs, and corporations; there may be further categories for students and pupils. One specific focus is on AI innovation that address major challenges and that have a positive societal impact.

Deliverables (brief description and month of delivery)

M9 D7.1 Report on the concept and time plan AI Innovation networking events and the AI Innovation price.
M12 D7.2 Report on innovation methods and their applicability to AI. First draft of the concept for the matching platform, the envisioned eco-systems, the learning approach, and the initial specification of the European Data Hub. Providing a list of Lighthouse Accelerators.
M18 D7.3 Initial report on how regulation for AI can be brought forward in different critical domains including a time plan for further event.
M36 D7.4 Final concept and implementation of the innovation platform. Report on the networking and education events, as well as on the AI innovation price. Recommendations for AI innovation accelerators and innovation infrastructure and ecosystem. Results and recommendation of the work on regulations for AI innovation

Work package number	**8**	**Lead beneficiary**	**Sorbonne**
M01-36	**1 (DFKI): 9.0 | 6 (BRNO U): 3.2 | 9 (CNR): 8.3 | 13 (CINI): 16.4 | 18 (FRAUNHOFER): 21.6 | 26 (K4A): 20.6 | 28 (ORU): 13.4 | 30 (SAP): 2.3 | 31 (SORBONNE): 13.3 | 33 (THALES SIX): 5.1 | 47 (UNIPI): 2.3 | 50 (UOS): 1.0**		

Objectives

This WP is devoted to the aim of fostering excellence, increasing the efficiency of collaboration, disseminating the latest and most advanced knowledge to all the academic and industrial AI laboratories in

Europe and making HumanE AI Net the center of a vibrant AI network in Europe. This includes implementing and operating the the Virtual Laboratory, integration of the Virtua Laboratory with the AI4EU platform, running the industrial Ph.D. postdoc and internship program, running the dissemination events to all relevant target groups (from scientific summer schools to workshops for policy makers and participation in public festivals) and the creation and distribution of relevant dissemination and knowledge spreading materials (from MOOCs, through policy brochures to general public facing YouTube videos).

Description of work

The work in this WP is centered around the implementation and operation of the Virtual Laboratory closely interwoven with the AI4EU platform. The Laboratory will not only be a one stop access point for outside the consortium, but also the main collaboration platform within the consortium (leveraging the respective AI4EU mechanism where possible), means to organize and manage a project generated materials and means for the coordination of various events and the Ph.D./postdoc and internship programs where appropriate.

T8.1 Virtual Laboratory Infrastructure (Fraunhofer, Orebro)

This WP will implement the infrastructure needed for the concept of HumanE AI Net Virtual Laboratory as described in section2.2.1.3. This will include the Virtual Laboratory website with all relevant components (e.g., the Blog) as well as the embedding within the AI4EU platform. The embedding will on one hand make sure that the Virtual Laboratory can be accessed as a resource from the AI4EU platform. It will on the other hand make sure that the Laboratory can seamlessly access the materials that the project will deposit within the platform (see T8.3). To ensure smooth integration the task is run by Joachim Köhler From Fraunhofer who also works with platform integration tasks with the AI4EU project.

T8.2 Virtual Laboratory Operation (Fraunhofer)

Maintaining and running the virtual Laboratory including making any upgrades needed to maintain compatibility with changes in the AI4EU platform and ensuring continuous availability.

T8.3 Challenge, Benchmarking and scientific material sharing infrastructure (Orebro, Fraunhofer)

This task will implement the infrastructure needed to use the AI4EU platform for scientific collaboration in particular run community challenges (part of many of the tasks in WP 1-5) and make available datasets,benchmarks and publications preprints. The infrastructure will also be integrated within the Virtual Laboratory (Task 8.1). The integration within AI4EU will be done in close cooperation with task 9.1. It will be lead by Prof. Alessandro Saffiotti who is the Scientific Manager of the AI4EU project and is thus well familiar with the AI4EU platform, and will furthermore help to integrate the research aspects of the AI4EU project (see also 9.1).

T8.4 Industrial Ph.D./postdoc and internship program (Sorbonne)

This task will be in charge of setting up and running the HumanE AI Net industrial Ph.D. and postdoc program as described in section 1.3.2.8. It will deal with the organization, operate the brokerage platform within the Virtual Laboratory, synchronize the work on the curriculum guidelines and help the Ph.Ds link and network with various other components of the project (e.g. summer schools). The brokerage and networking efforts will also support personnel exchange between research partners and industry (within and outside the consortium) with respect to internships. This will be closely coordinated with the overall microprojects management as the microproject will be an important personnel exchange mechanisms through which academics will come to spend time at industrial sites and the other way round.

T8.5 Knowledge dissemination events for the European AI community(and beyond) (CINI)

This task will organise/coordinate the scientific summer schools, tutorials and workshops described in section 1.3.2.7 as a key instrument of knowledge spreading to the scientific and industrial R&D community. The annual summer school on human-centric AI (D8.3a,b,c) will be directly driven and organized by this task. Other summer schools, tutorials and workshops will be driven by the tasks/WPs responsible for the respective topics, but coordinated and supported by this task.

The Cini National lab AIIS (AI and Intelligent Systems) running this task organizes every year some national events such as ITal- IA (www.ital-ia.it) with several ws on AI applications with a very dynamic pitch format to connect researchers startuppers and industrial experts. This will be extended at EU level.

As well we plan to include in the program of the network the "Advanced summer schools in AI" scheduled for 2020 to 2022 in the universities of Unimore unibo and unife (Modena Bologna and Ferrara) co-funded by Emilia Romagna region under EU FESR program in 2019. They will have an international board.

T8.6 Scientific and technical knowledge dissemination materials (incl. MOOCs) (K4All)

This task will produce the planned MOOCs on human-centric AI which will be a key knowledge spreading material of HumanE AI Net (beyond scientific papers) as described in section 1.3.2.7. The MOOCs will be made available through the AI4EU platform and the virtual laboratory (T8.1). It will also support making available online multimedia materials from the summer schools, tutorials and workshops (T8.7). Finally it will coordinate the writing of the Handbook of human-centric AI This task is responsible for D8.4a,b[24]

T8.7 Dissemination to policy and decision makers (DFKI)

This WP will implement the dissemination measures aimed at European policy makers. It will leverage the extensive political role played by the involved research centers (e.g. DFKI, Fraunhofer in Germany, CNR, FBK, CINI in Italy, INRIA, CNRS in France, CU in Czech Republic, to name just some) and the involvement of both individual and organizations in various European bodies (e.g. 5 researchers from the consortium are members of the Eu High Level Experts group on AI). It will also use the political networks of CLAIRE, EuRAI and ELLIS who are all represented in the consortium with key members (e.g. Barry O'Sullivan the president of EurAI, Holger Hoos and Philipp Slusallek founding members of CLAIRE. This WP will coordinate the dissemination effort, the participation of the consortium in relevant EU events, production of the dissemination material and the organization of the Brussels policy meeting (D8.6).

T8.8 Engaging the general public and fostering pubic debate (K4All)

This task will coordinate the general public facing dissemination efforts. In addition to coordinating the work of the PR departments of the various partners into a coherent strategy and the usual instruments (social media, press releases, web site) considerable effort will into innovative concepts described such as the youtube competition, creating Reddit communities and explicitly targeting schools, producing VR materials. This task will also actively try to mediate the participation of Key HumanE AI Net scientists in public events, and debates.

Deliverables (brief description and month of delivery)

M06 D8.1 Initial Version of the Virtual Laboratory including benchmarking and challenge infrastructure

M06 D8.2 Initial Version of the Humane AI multimedia dissemination package (from tasks 8.7 and 8.8)

M12, 24, 32 D8.3a, b, c First, Second, Third Humane AI Net summer school on human-centric AI

M18 D8.4 Final Version of the Virtual Laboratory including benchmarking and challenge infrastructure

M18, M30 D8.5a, b First/second version of the HumanE-AI-Net scientific Knowledge Dissemination Package (MOOCs, summer school recordings etc.)

M30 D8.6 HumanE AI Net Policy dissemination event in Brussels

WP 9 Lead Cork	**Synergies with AI on demand platform(s) and the Broader European AI Community**
M01-36	**1 (DFKI): 9.0 | 9 (CNR): 8.3 | 13 (CINI): 1.8 | 17 (FORTISS): 3.3 | 26 (K4A): 8.8 | 28 (ORU): 2.7 | 30 (SAP): 0.8 | 33 (THALES SIX): 20.3 | 41 (UCC): 8.3 | 44 (ULEI): 5.0 | 50 (UOS): 1.0**

Objectives

The aim of this WP is to embed HumanE AI Net within the landscape of relevant European and national initiatives. This includes, as primary concerns, the interaction with the AI on demand Platform (AI4EU), Digital Innovation Hubs and other ICT 48 networks as specified in the call. However, given the interdisciplinary nature of our work, we believe that HumanE AI is important for a wide range of initiatives around AI such as SoBigData, the European Language Technology community (European Language Grid) and the broader European AI networks (EurAI, CLAIRE, ELLIS). The WP is lead by Barry O'Sullivan from Cork who is the President of EuRAI, the main European AI association and well connected within virtually all relevant European AI communities.

Description of work

Our approach to interaction with other initiatives is rooted in the fact that the consortium members are all well interconnected within those initiatives with many key people in the HumanE also being important members of the respective initiatives (e.g., Prof. Fosca Gianotti is Coordinator of the SoBiGData consortium, Prof. Barry O'Sullivan is the President of EuRAI, Prof. Holger Hoos is one of the 3 co-initiators and in the 5 person core-team of CLAIRE, Prof. John-Shaw Taylor and Prof Samual Kaski are leading members of ELLIS, Dr. Georg Rehm is the coordinator of ELG and General Secretary of META-NET, Patrick Gatellier is the coordinator of AI4EU, Prof. Alessandro Saffiotti is scientific manager within AI4EU and Prof James Crowley, Prof. Virginia Dignum and Prof. Michela Milano all task leaders within AI4EU.).
We rely on such key individuals to lead the respective tasks below, devise the optimal strategy for the interaction with each initiative making sure that there is continuous flow of information and that synergies can be leveraged. In addition to customized strategies for each initiative we will organize two community workshops to which members and representatives of all initiatives will be invited.

T9.1. Synergy and coordination with European AI on demand Platform (AI4EU) (Thales)

Collaboration with AI4EU has a special role within HumanE AI Net as the AI4EU platform will be the key channel for bringing tools, datasets and publications to the European AI community. We will also be running challenges through the Al4EU platform (see T8.3T) and the virtual laboratory (T8.2) will be made available through as a resource on the AI4EU platform. The infrastructure needed for challenges, benchmarks and sharing resources between researches will be a key contribution of the project to the platform (see T8.3). The collaboration strategy a whole is outlined in section 1.3.2.4.1. This task, run by Thales who is the coordinator of the AI4 Eu project will coordinate all AI4EU related activities, provide advice on the processes and conditions that must be observed when contributing to the platform, and help disseminate information about Humane AI Net within the Ai4EU consortium (and the other way round).

T9.2 Collaboration with the Digital Innovation Hubs (Fortis)

We will build on the fact that Fortiss is a member of the CSA of the DIH that is most relevant for this proposal (SmartAnythingEverywhere) to build an intensive interaction with the DIH including cross presentations at relevant meetings, distribution of information materials, leveraging the DIH industrial network and expanding the collaborators network. The second DIH most relevant for HumanE AI net, DIH network-cluster on Robotics, will be interfaced by DFKI DFKI who is both member of the RIMA consortium and the association of the European robotics community euRobotics (who is a partner in the corresponding CSA).

T9.3 Collaboration with other ICT 48 Networks (ICT 48 CSA) (ULEI)

This task is devoted to the interaction with the other projects of the ICT 48 call and the corresponding CSA and will coordinate the communication, attendance of meetings and events and the exchange of knowledge.

T9.4 Synchronization with the planned AI, BigData and Robotics PPP (Cork)

The HumanE AI FET preparatory action in which the consortium of this proposal has its roots has been an active participants in the initiative to establish an European AI, Robotics and Big Data with several of the consortium members being part of the consultations (Barry O'Sullivan, Philipp Slussalek, Holger Hoos, Paul Lukowicz). We will build on this connection to actively engage with that initiative.

T9.5 Interface to CLAIRE (ULEI)

The interaction with CLAIRE (Confederation of Laboratories for AI Research in Europe) will be led by Prof. Holger Hoos from Leiden University, who is one of three co-initiators of CLAIRE and a member of the 6-person team leading CLAIRE. CLAIRE has played an important role in establishing the current HumanE AI consortium, and many members of the consortium are also member of the CLAIRE Research Network, which encompasses over 300 AI research groups and institutions, spanning all areas of AI, across all of Europe, with a strong focus on human-centered, trustworthy AI. The focus will be on leveraging the CLAIRE network for interaction with AI experts not represented in the RIA networks in areas where such expertise might be beneficial to HumanE AI Net, via participation in regular CLAIRE events and use of the CLAIRE communications platform. Of particular interest are CLAIRE theme development workshops, such as the one organised in March 2019 with the European Space Agency, that present valuable opportunities to interact with AI stakeholders from industry and other organisations.

T9.6 Interface to ELLIS/PASCAL (UCL/K4All)

The interface with the ELLIS network (and the PASCAL network partners) will be managed through UCL and K4All. K4All was created as a legacy organisation for the PASCAL network and has been a founding supporter of the ELLIS network, while John Shawe-Taylor at UCL was the scientific coordinator of PASCAL and has been identified as the liaison by ELLIS. ELLIS has a range of themes that link with the Humane AI network and collaborations will enable synergies to be created that can help progress the

research of both networks.

T9.7 Interface to ELG and META-NET (DFKI)

The interaction with the European Language Grid (ELG; 2019-2021) project and the European Network of Excellence META-NET, established in 2010, will be lead by Dr. Georg Rehm (DFKI) who is the Coordinator of ELG and the General Secretary of META-NET. The joint goal of ELG/META-NET and HumanE-AI-Net is to set up a close and fruitful collaboration of the two initiatives, bringing the language-centric AI community even closer to the wider AI community. In addition, the emerging European Language Grid will provide its language processing and generation services and the available datasets to those HumanE-AI-Net microprojects that focus on language. Furthermore, ELG is organising two more annual conferences in the autumn of 2020 and also in 2021. For HumanE-AI-Net we plan to add a third day to the two-day ELG event in 2020 and 2021 as an annual HumanE-AI-Net conference and forum for the HumanE-AI-Net community to meet and to discuss interim results. In 2022 we will organise a two-day HumanE-AI-Net conference, adding a one-day ELG conference.

T9.8 Interface to SoBigData.eu (CNR)

(2015-2024, H2020-Excellent Science, n. 871042) (www.sobigdata.eu), will be lead by Fosca Giannotti (CNR) who is the coordinator. SoBigData is a multi-disciplinary research community, that aggregates 32 partners of 12 EU Countries aimed at realising large-scale social mining experiments to understand the complexity of our contemporary, globally-interconnected society. SoBigData provides open and responsible access (FACT: Fairness, Accuracy, Confidentiality and Transparency) to more than 200 social mining resources: curated datasets, algorithms, training material, needed to observe and measure social phenomena at individual, collective and community scale. The e-infrastructure (aligned with European Open Science Cloud - EOSC) has over 2,500 registered users, with daily peaks of accesses and executions in the millions. All this provides a fertile ground for expanding Humane-AI community and set-up fruitful collaborations, and realizing microproject over the available social datasets (mobility data, social media data, demographic data etc, survey data) already collected and curated, in synergy with the Transactional Access program of SoBigData. The SoBigData "data challenges" planned for 2020 and 2021 "social data for well being and information disorder" will be launched jointly with HumanE-AI-Net. In 2022 we will organise a two-day HumanE-AI-Net conference, adding a one-day SoBigData conference.

T9.9 Coordination with national initiatives (INRIA, CINI, DFKI)

Many of the HumanE AI Net partners are instrumental insetting their respective countries' AI strategies. This is true for example for the project coordinator DFKI in Germany, for INRIA in France, for CU and BUT in the Czech Republic and for CNR and CINI in Italy. This task will coordinate the activities of the partners on national levels to ensure the the project vision is communicated in a coherent way. It will support the national dissemination with appropriate materials. Finally, it will ensure that all partners are aware of relevant national activities in all countries.

T9.10 Global Outreach (K4All)

This task will work toward ensuring global visibility of HumanE I Net as a center excellence in AI. To this and it will help the partners leverage each other's international networks (eg to help students get interesting international internships or to help with recruiting), coordinate dissemination efforts beyond Europe and collect and produce dissemination material. The effort will not be restricted to academia but will include broader outreach to for example NGOs and other international organizations (eg. within UNO or WHO).

Deliverables (brief description and month of delivery)

D9.1 M04 HumanE AI Net Community Kickoff Workshop for all relevant initiatives

D9.2 M06 HumanE AI EU and national networking strategy plan

D9.3 M30 HumanE AI Net Community Outlook Workshop for all relevant initiatives

D9.4 M36 HumanE AI EU and national networking strategy report and outlook beyond the project

WP 10 Lead DFKI	**Management and Governance**

M01-36	1: DFKI 36, all with 5% of their PMs

Objectives

WP 10 globally aims at ensuring a strong and efficient day-to-day management in order for the project to meet its objectives on time and within budget constraints. More specifically, WP10 focuses on ensuring administrative and financial management of the project; developing a cooperation spirit between partners; enabling smooth work progress; ensuring project reporting; enabling an interface with the European Commission; contract managing and assuring compliance with the Commission's reporting requirements

Description of work (where appropriate, broken down into tasks), lead partner and role of participants

T10.1 Project administration and financial management Task leader: DFKI

This task encompasses essential leadership and project management aspects required for successful planning, undertaking and finalising of any project. These aspects are especially relevant, given that the project will deliver eleven (10) work packages, many of which will operate in synergy with each other. DFKI will manage the project. It will establish and maintain the central project management team and undertake the full range of managerial, administrative, supporting and co-ordinating activities necessary and appropriate for a project of this size. Humane AI Net coordinator Paul Lukowicz is an experienced infrastructure organizer and project coordinator directly supported by the DFKI administration, which administers a total annual volume of more than 30 Million Euro in various publicly funded projects (EU, national industry). In addition, Lukowicz will be supported by George Kampis; Kampis has outstanding eresearch management in EU project management, along with a robust history of supporting and managing 5 EU and other research projects. Kampis' specific work within this and other projects includes: establishing a high-profile Scientific Advisory Board; managing funds and their transfer to partners; dealing with all contractual matters; preparing and leading project meetings and reviews (proposing agendas, preparing minutes), and providing efficient financial management and timely payment procedures.

T10.2 Consortium and Project Management Task leader: DFKI. Participants: ALL.

This task will ensure that all partners share the consortium's collective mission and that all partners are integrated inthe general decision and scientific process. Task T10.2 will coordinate the overall project regarding networking, scientific and dissemination issues to ensure highest quality of service and completing deliverables according to the timeframe and foreseen budget. It will prepare all project's management bodies meetings and control the progress ofall project work bodies, ensuring that they keep to schedule. Regarding external reporting and communication, the task will gather and compile all reports as requested by the Commission and all other eternal bodies as required. General Partner Meetings include all partners and control work carried out at strategic and content level. There will be annual consortium meetings, involving all beneficiaries and directly managed by the Coordinator. The advisory board, the ethics board other external consultants (when deemed relevant) will be periodically invited, on an 'as-needed' basis. Regular Project Management Board meetings will take place either aligned to the general partner meetings or via teleconferences. Moreover, on a more frequent basis, meetings of the Steering Board will take place. We will review all the project deliverables and reports in order to guarantee textual and presentation quality.

Deliverables (brief description and month of delivery)

M03 D 10.1 Project handbook containing the key procedures (e.g. on microprojects)

M03 D10.2 All Relevant Boards established. Definition of the composition and installation of the boards involved in the project's management.

(M12 M24 M36) D10.3a,b,c Project periodic report (Type: R). The periodical management and financial report as required by the European Commission, namely Report 1, Report 2, and the final Management and Financial Plan.

M03 D10.4 Risk Identification and Evaluation. Will address different kinds of risk (external, internal,

strategic, operational, other) and will be updated periodically from the Management Board.

Table 3.1c: List of Deliverables

No	Deliverable name	WP	Lead	Type	Dissem. level	Due
D1.1, 2.2, 3.1	**First. Second, Third year microproject results** (papers, tools, datasets) deposited for general use on the AI4EU platform	**1, 2, 3**	**UCL, INRIA, Aalto**	R	PU	**12, 24, 36**
D4.1, 5.1, 4.2	**First, Second, Final year microproject results** on societal, ethical and responsible AI deposited for general use on the AI4EU platform	**4, 5, 4**	**UNIPI, UMU, UNIPI**	R	PU	**12, 24, 36**
D5.2	Report on Ethical, Legal and Responsible AI concepts, design principles, best-practices and tools, including a report on for LPbD (impact assessment and mitigation measures) and a set of lessons learnt	**5**	**UMU**	R	PU	**30**
D6.1, 6.2, 6.3	**First, Second, Final report** on research results, their application significance and the resulting evolution of the research agenda	**6**	**DFKI**	R	PU	**09, 21, 36**
D7.1	Report on the concept and time plan AI Innovation networking events and the AI Innovation price.	**7**	**GE**	R	PU	**9**
D7.2	Report on innovation methods and their applicability to AI.	**7**	**GE**	R	PU	**12**
D7.3	**Initial** report on how regulation for AI can be brought forward in different critical domains including a time plan for further event.	**7**	**GE**	R	PU	**18**
D7.4	**Final** concept and implementation of the innovation platform.	**7**	**GE**	R	PU	**36**
D8.1, 8.4	Initial Version of the Virtual Laboratory including benchmarking and challenge infrastructure	**8**	**Sorbonne**	Other	PU	**06, 18**
D8.2	Initial Version of the Humane AI multimedia dissemination package	**8**	**Sorbonne**	Other	PU	**06**
D8.3 a) b), c)	First, Second, Thirds Humane AI Net summer school on human-centric AI	**8**	**Sorbonne**	Other	PU	**12, 24, 32**
D8.5 a), b)	First, Second version of the HumanE-AI-Net scientific Knowledge Dissemination Package	**8**	**Sorbonne**	Other	PU	**18, 30**
D8.6	HumanE AI Net Policy Dissemination event in Brussels	**8**	**Sorbonne**	DEC	PU	**30**
D9.1	HumanE AI Net Community Kickoff Workshop for all relevant initiatives	**9**	**UCC**	DEC	PU	**04**
D9.2	HumanE AI EU and national networking strategy plan	**9**	**UCC**	R	PU	**26**
D9.3	HumanE AI Net Community Outlook Workshop for all relevant initiatives	**9**	**UCC**	DEC	PU	**30**
D9.4	HumanE AI Net EU and national networking strategy report and outlook beyond the project	**9**	**UCC**	R	PU	**36**
D10.1	Project handbook containing the key procedures	**10**	**DFKI**	R	PU	**03**
D10.2	All Relevant Boards established. Definition of composition and installation of the boards involved in the project's management.	**10**	**DFKI**		PU	**03**
D10.3 a), b), c)	Periodic Report	**10**	**DFKI**	R	PU	**12, 24, 36**
D10.4	Risk Identification and Evaluation	**10**	**DFKI**	R	PU	**04**

Table 3.2a: List of milestones

No.	Milestone name	WP	Due	Verification
M1	Micro-project procedures in place, first 5 micro-projects operating,	**all**	**M3**	**existance**
M2	Content can be placed and accessed through the Virtual Laboratory and AI4EU	**8**	**M7**	**existance**
M3	Stake holder workshops for agenda setting have been run	**6**	**M6**	**report**
M4	Micro-project procedures for external researchers in place first 5 external researchers participate in micro-projects	**all**	**M6**	**document**
M5	Procedures for conducting challenges in place, first challenge run	**1-5**	**M9**	**document**
M6	Concept and responsibilities for scientific dissemination package defined	**8**	**M12**	**document**
M7	Concept and authors for the Handbook of Human Centric AI defined	**8**	**M24**	**document**

M8	Concept or establishing the Human AI summer school beyond project end ready	**8**	**M30**	**document**
M9	10 successful industrial use cases demonstrated	**6**	**M24**	**presence of use cases**
M10	Having given a HumanE AI Net talk at at least 1 events of every relevant initiative from WP 9	**9**	**M18**	**reprots**
M11	Concept for sustaining the community beyond the project in pace	**10**	**M30**	**document**

Table 3.2b: Critical risks for implementation

No.	Description of risk	WP	Mitigation
R1	Problems with the process for establishing and running micro-projects (Low)	**1-6**	Process has been discussed within the consortium and will be addressed immediately with the help of the administration
R2	Problems with the interface to the AI4EU Platform (Med)	**8,9**	key AI4Eu persons are involved in the implementation, alternative methods for collaboration
R3	Lack of cohesion/cooperation in the consortium (Low)	**all**	many members of the consortium have worked together before, stringent management
R4	Key person organization dropping out or not performing (Low)	**all**	For all key functions there is a high degree of redundancy in the project (including the coordinator level)
R5	Inability to mobilize enough support and interest in the community	**8, 9**	The consortium is extremely well embedded in the community, reach out will be done early on
R6	Disruptive developments in the field of AI that would make parts of the agenda irrelevant	**1-5**	The dynamic micro-project-oriented concepts allows us quickly respond to new developments
R7	Organization/management problems	**10**	The management structure has been designed to be appropriate for the consortium's size and project's type. The coordinator together with the EC will monitor how well the management structure functions and changes will be introduced swiftly

Table 3.4a: Summary of staff effort

Participant		WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	WP10	Total PM
1	**DFKI**	0.0	18.0	9.0	0.0	0.0	13.5	0.0	9.0	9.0	36.	**95,5**
2	**AALTO**	12.9	0.0	19.4	0.0	0.0	0.0	0.0	0.0	0.0	1.7	**34**
3	**AIRBUS**	0.0	0.0	0.0	0.0	0.0	8.8	0.0	0.0	0.0	0.5	**9.3**
4	**Algebraic AI**	10.8	0.0	0.0	0.0	0.0	2.7	0.0	0.0	0.0	0.7	**14.2**
5	**ATHENA**	2.9	5.9	11.7	0.0	0.0	8.8	0.0	0.0	0.0	1.5	**30.9**
6	**BRNO U**	0.0	0.0	4.8	4.8	0.0	0.0	0.0	3.2	0.0	0.9	**13.77**
7	**BSC**	0.0	0.0	0.0	0.0	6.7	4.0	2.7	0.0	0.0	0.7	**14.2**
8	**CEU**	0.0	0.0	0.0	0.0	10.1	0.0	0.0	0.0	0.0	0.5	**10.6**
9	**CNR**	8.3	8.3	8.3	16.5	16.5	8.3	0.0	8.3	8.3	4.4	**87**
10	**CNRS**	1.7	5.9	6.8	0.0	2.5	0.0	0.0	0.0	0.0	0.9	**17.8**
11	**CSIC**	10.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	**10.6**
12	**CU**	0.0	3.7	11.1	0.0	0.0	3.7	0.0	0.0	0.0	1.0	**19.4**
13	**CINI**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.4	1.8	1.0	**19.2**
14	**ELTE**	16.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	**17**
15	**ETHZ**	0.0	0.0	0.0	0.0	0.0	15.0	0.0	0.0	0.0	0.8	**15.8**
16	**FBK**	0.0	0.0	10.7	0.0	0.0	10.7	0.0	0.0	0.0	1.1	**22.5**
17	**FORTISS**	0.0	0.0	0.0	0.0	0.0	10.0	3.3	0.0	3.3	0.9	**17.6**
18	**FRAUNHOFER**	0.0	0.0	0.0	0.0	0.0	6.2	3.1	21.6	0.0	1.6	**32.5**
19	**Generali**	0.0	0.0	0.0	0.0	0.0	6.9	0.0	0.0	0.0	0.4	**7.3**
20	**GE**	0.0	0.0	0.0	0.0	0.0	0.0	13.5	0.0	0.0	0.7	**14.2**
21	**INSEC TEC**	6.6	9.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	**17.3**
22	**ING**	0.0	0.0	0.0	0.0	0.0	9.0	3.9	0.0	0.0	0.7	**13.6**
23	**INRIA**	0.0	19.7	19.7	0.0	0.0	0.0	0.0	0.0	0.0	2.1	**41.4**
24	**IST**	0.0	0.0	7.3	7.3	0.0	0.0	0.0	0.0	0.0	0.8	**15.3**
25	**JSI**	5.6	13.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	**19.8**
26	**K4A**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.6	8.8	1.6	**31**
27	**LMU**	13.3	0.0	13.3	0.0	0.0	0.0	17.8	0.0	0.0	2.3	**46.8**
28	**ORU**	0.0	5.4	5.4	0.0	0.0	0.0	0.0	13.4	2.7	1.4	**28.3**
29	**PHILIPS**	0.0	0.0	0.0	0.0	0.0	15.0	0.0	0.0	0.0	0.8	**15.8**
30	**SAP**	0.0	0.0	0.0	1.5	3.0	7.5	0.0	2.3	0.8	0.8	**15.8**
31	**SORBONNE**	0.0	3.3	13.3	3.3	0.0	0.0	0.0	13.3	0.0	1.8	**35**
32	**STICHTING**	18.1	0.0	12.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	**31.7**
33	**THALES SIX**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.1	20.3	1.3	**26.7**
34	**TID**	0.0	0.0	0.0	0.0	0.0	15.9	0.0	0.0	0.0	0.8	**16.7**

35	**TILDE**	0.0	0.0	3.0	0.0	0.0	12.0	0.0	0.0	0.0	0.8	**15.8**
36	**TUB**	13.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	**14.2**
37	**TUBITAK**	0.0	8.4	0.0	0.0	0.0	8.4	0.0	0.0	0.0	0.9	**17.7**
38	**TU DELFT**	0.0	0.0	16.2	0.0	16.2	0.0	0.0	0.0	0.0	1.7	**34.2**
39	**TUK**	0.0	0.0	0.0	0.0	0.0	32.5	0.0	0.0	0.0	1.7	**34.2**
40	**TU WIEN**	6.7	16.7	10.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	**35.2**
41	**UCC**	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.3	0.6	**12.5**
42	**UCPH**	2.7	8.1	16.1	0.0	0.0	0.0	0.0	0.0	0.0	1.4	**28.3**
43	**UGA**	0.0	8.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	**8.6**
44	**ULEI**	3.3	2.5	1.7	4.2	0.0	0.0	0.0	0.0	5.0	0.9	**17.5**
45	**UMU**	0.0	0.0	9.3	9.3	18.7	0.0	0.0	0.0	0.0	2.0	**39.3**
46	**UNIBO**	2.0	3.4	0.0	5.4	2.7	0.0	0.0	0.0	0.0	0.7	**14.2**
47	**UNIPI**	4.7	4.7	0.0	23.4	7.0	4.7	0.0	2.3	0.0	2.5	**49.2**
48	**UCL**	25.7	10.3	15.4	0.0	0.0	0.0	0.0	0.0	0.0	2.7	**54.2**
49	**WARSAW**	0.0	9.4	9.4	28.2	0.0	0.0	0.0	0.0	0.0	2.5	**49.4**
50	**UOS**	1.9	5.8	2.9	1.9	1.9	1.9	1.0	1.0	1.0	1.0	**20.2**
51	**UPF**	4.9	6.6	4.9	0.0	0.0	0.0	0.0	0.0	0.0	0.9	**17.3**
52	**UVB**	0.0	0.0	0.0	0.0	29.3	0.0	0.0	0.0	0.0	1.5	**30.8**
53	**VW AG**	6.0	0.0	0.0	0.0	0.0	7.5	1.5	0.0	0.0	0.8	**15.8**
Total Person Months		**181.5**	**177.1**	**241.8**	**105.8**	**114.7**	**213.1**	**46.7**	**116.5**	**69.3**	**95.2**	**1361.7**

References: Literature list is attached as annex to Part 4