HumanE AI:

Toward AI Systems that Augment and Empower Humans by Understanding Us, our Society and the World Around Us

Grant Agreement Number: 761758
Project Acronym: HumanE AI

Project Dates: 2019-03-01 to 2020-03-30
Project Duration: 14 months

D4.2 Wrap-up of the Humane AI vision

Author(s): George Kampis
Contributing partners:
Date: May 28, 2020
Approved by: Paul Lukowicz
Type: Report (R)
Status: Final report
Contact: George.Kampis@dfki.de

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

DISCLAIMER

This living document contains material, which is the copyright of *HumanE AI* Consortium parties, and no copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the *HumanE AI* Consortium as a whole, nor a certain party of the *HumanE AI* Consortium warrant that the information contained in this document is suitable for use, nor that the use of the information is free from risk, and accepts no liability for loss or damage suffered by any person using this information.

This document reflects only the authors' view. The European Community is not liable for any use that may be made of the information contained herein.

DOCUMENT INFO

0.1 Authors

Authors	Institution	e-mail
George Kampis (GK)	DFKI	George.Kampis@dfki.de

0.2 Document History

Revision		
Date	**Lead Author(s)**	**Comments**
28.05.2020	GK	Initial draft
28.05.2020	GK	Final draft

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This is a summary statement of the HumanE AI vision. It provides an overview of the essentials but instead of introducing new material it overviews work done in other Deliverables, as explained next.

1. INTRODUCTION

The HumanE AI vision has been formulated and published at several places, including some other Deliverables such as D2.2 or D1.3., and repeated in D7.4.

Albeit still developing in its details (e.g. in the ethics aspects, which are continually worked on by our high-level experts Profs. Jeroen van den Hoven of TU Delft, NL, and Virginia Dignum of Umea University, Sweden, or the legal issues, which are tackled by Prof. Mireille Hildebrandt, VU Brussels – a member not of the current CSA but of the upcoming HumanE AI Net ICT48 project), the vision shown in the current Document has largely lost its significance as a stand-alone file, after the original project goals had to be realigned after the cancellation of the Flagship programme. The resulting HumanE AI vision is now more like a platform statement for a movement and the motivation behind future project proposals in response to various (mostly FET and ICT) Calls.

2. WRAP-UP OF THE VISION

Albeit there can be no Flagship proposal arising from the current project, still the HumanE AI vision remains to be interesting. In particular, the concept of **micro-projects including 3rd parties,** funded by a main project (such as a HumanE AI derivative) and administered by that, a central idea of the current vision's *organizational side*, has proved to be a valuable tool and a chosen way to go.

The idea of micro-projects, pioneered by the current proposal, was received well by the EC**,** and the idea has been put to practice in the upcoming ICT48 project HumanE AI Net.

The *conceptual side* of the current vision is best formulated in D2.2 (mistakenly submitted under the name D2.1, but corrected later). The Introduction (as well as the Executive Summary) essentially says the following (text adopted and edited):

We discuss the context and motivation for an initiative to organize the HumanE AI community. describes the Humane AI Vision to create the foundations for AI systems that empower people and society. We describe enabling technologies and list required innovations in a broad spectrum of areas including Machine Learning, Computer Vision, Robotics, Human Computer Interaction, Natural Language Processing and Conversational AI.

We detail the research challenges that must overcome to meet this vision. These include new approaches to learning, reasoning, and planning, new theories and techniques for multimodal perception and modeling to enable intelligent systems to

perceive and model humans, and human behaviours, human language, and social interaction, new theories and methods for combined human-machine intelligence, where AI and humans interact and collaborate. We also describe the challenge of assuring that the symbiosis of humans and AI systems, and the need for new technologies that enable AI systems that are ethical, legal and responsible by design, in accordance with European ethical, cultural and social values.

We go on to describe methods by which HumanE AI can best pursue its vision in the current socio- economic environment. We put forward a program for challenge-based research in which the community documents research challenges and with published performance targets using benchmark problems and data sets. We describe an approach centered on dynamically defined collaborative mini-projects undertaken by small groups of researchers to opportunistically respond to research problems and opportunities as they emerge, without having to wait for the multi-year funding cycles.

HumanE AI represents a community of researchers and innovators who seek to create the conditions for AI technologies that empower humans and human society to vastly improve quality of life for all. This community has set as its goal to develop the scientific and technological foundations needed to shape the AI revolution in a direction that is beneficial to humans and humanity on both an individual and a societal level. The aim is to facilitate AI systems that enhance human capabilities and empower people as individuals and while assuring evolution of a healthy and nurturing society.

The HumanE AI community brings about the mobilization of a research landscape far beyond the direct project funding and create a unique innovation ecosystem that will provide manifold return on investment for the European economy and society. The HumanE AI community was originally formed to create a Flagship initiative to meet the challenge of Human Centered AI. We report preliminary results from a one-year community-wide initiative to organize the community to meet this vision.

HumanE AI must develop the scientific and technological foundations for artificial intelligence that is beneficial to humans and humanity, in accordance with European ethical, social, and cultural values. The core challenge is the development of robust, trustworthy AI systems that understand humans, adapt to human environments, and behave appropriately in social situations. Our overall goal is to develop a technology for artificial intelligence that enhances human abilities and empower individuals and society.

Achieving this vision requires new solutions to fundamental scientific questions—not just within narrow classical AI silos, but at the intersections of AI disciplines such as learning, reasoning, and perception with scientific domains including human-computer interaction, cognitive science, and the social sciences. This will require substantial research in areas such as:

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 modelling****. 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 interaction and collaboration****. Developing paradigms that allow humans and AI systems including service robots and smart environments to interact and collaborate in a way that enhances human abilities and empowers people.*

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.*

3. CONCLUSION

We have summarized the HumanE AI vision and characterized it briefly from the organizational and conceptual point of view. For a detailed account, we refer to the original documents cited at the appropriate places.