Interactive Human Centered Artificial Intelligence: A Definition and Research Challenges

Albrecht Schmidt
albrecht.schmidt@ifi.lmu.de
Ludwig-Maximilians-Universität München
München, Germany

Figure 1: Is AI and ML a new material to name an epoch? Will it impact live as the "stone age" or "bronze age" did?

ABSTRACT

Artificial Intelligence (AI) has become the buzzword of the last decade. Advances so far have been largely technical with a focus on machine learning (ML). Only recently have we begun seeing a shift towards focusing on the human aspects of artificial intelligence, centered on the narrow view of making AI interactive and explainable. In this paper I suggest a definition for "Interactive Human Centered Artificial Intelligence" and outline the required properties. Staying in control is essential for humans to feel safe and have self-determination. Hence, we need to find ways for humans to understand AI based systems and means to allow human control and oversight. In our work, we argue that levels of abstractions and granularity of control are a general solution to this. Furthermore, it is essential that we make explicit why we want AI and what are the goals of AI research and development. We need to state the properties that we expect of future intelligent systems and who will benefit from a system or service. For me, AI and ML are very much comparable to raw materials (like stone, iron, or bronze). Historical periods are named after these materials as they fundamentally changed what humans can build and what tools humans can engineer. Hence, I argue that in the AI age we need to shift the focus from the material (e.g. the AI algorithms, as there will be plenty of material) towards the tools and infrastructures that are enabled which are beneficial to humans. It is apparent that AI will allow the automation of mental routine tasks and that it will extend our ability to perceive the world and foresee events. For me, the central question is how to create these tools for amplifying the human mind without compromising human values.

AVI '20, September 28-October 2, 2020, Salerno, Italy

ACM ISBN 978-1-4503-7535-1/20/09.
https://doi.org/10.1145/3399715.3400873

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Ubiquitous and mobile computing theory, concepts and paradigms**; • **Computing methodologies** → **Artificial intelligence**; **Machine learning**.

KEYWORDS

Interactive Human Centered Artificial Intelligence, Human Computer Interaction, Artificial Intelligence

ACM Reference Format:
Albrecht Schmidt. 2020. Interactive Human Centered Artificial Intelligence: A Definition and Research Challenges . In *International Conference on Advanced Visual Interfaces (AVI '20), September 28-October 2, 2020, Salerno, Italy.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3399715.3400873

1 INTRODUCTION

Artificial Intelligence (AI) has become the buzzword of the last decade. With strong advances in Machine Learning (ML), a range of new applications have become possible and many others appear to be within reach in the near future. Speech interaction, intelligent assistants, and autonomous driving are just some examples. In conferences, exploding in size, researchers in AI and ML have demonstrated technical feasibility (e.g. fast and robust face recognition techniques) and celebrated major scientific advances with regard to gains in speed, recognition performance and precision, sometimes even under real world conditions. The applications that were enabled by these advances, received a mixed response from society. On one side there is great excitement about new applications (e.g. personal image search including names), but also fear about the use of this technology against individuals and society (e.g. face recognition for surveillance). This has stirred up the *old question* of whether a technology is good or bad and raises concerns about the bigger picture of responsibility.

There is an understanding that the advances seen in AI and ML are only the beginning and that there will be more to come as our

lives become increasingly digital. From a technical perspective, this is truly exciting. From an individual as well as societal perspective, there is a fear of competition between humans and AI. This fear may be very concrete (and very realistic): is the job I am good at easily be replaced by an intelligent agent? Or it may be very abstract (and in my view not well founded at this point in time): will AI take over control and dominate humankind. In general, there is the hope that if we get it right, the majority of people could benefit from the advances driven by AI. This is well summarized by the following statement from the title page of [1]:

> "Experts say the rise of artificial intelligence will make most people better off over the next decade, but many have concerns about how advances in AI will affect what it means to be human, to be productive and to exercise free will".

To focus on the human aspects of artificial intelligence, there are several initiatives working towards a human centered AI as well as towards making AI interactive and explainable. Examples include the European Humane AI Project[4] and the Stanford Institute for Human-Centered Artificial Intelligence[3]. These initiatives bring together researchers from different backgrounds, in particular from artificial intelligence and human computer interaction but also from social science, philosophy, and ethics. With the following definition of "Interactive Human Centered Artificial Intelligence" we hope to start a more in-depth discussion on the goals of AI research and the properties that we expect of future systems.

2 INTERACTIVE HUMAN CENTERED ARTIFICIAL INTELLIGENCE

The term "Artificial Intelligence" is used in many different ways and there is no clear definition. The term "Machine Learning" is more technical and more specific, but here too finding a clear definition is difficult. Even going one step back and looking for a definition of "intelligence" it becomes clear that this also opens a philosophical discussion.

Hence that the term "Interactive Human Centered Artificial Intelligence" is in no way self-explaining. However, attempting to defining this term can help to uncover what challenges we will face when working towards an Interactive Human Centered Artificial Intelligence. In the following I suggest a definition as a starting point for further discussion.

2.1 Definition

An **Interactive Human Centered Artificial Intelligence** is an Artificial Intelligence that enables interactive exploration and manipulation in real time and is designed with a clear purpose for human benefit while being transparent about who has control over data and algorithms.

2.2 Properties

An "Interactive Human Centered Artificial Intelligence" has the following **properties**:

- Individuals can interact in real time with the algorithms, models, and data and can manipulate and control all relevant parameters.
- The impact of changes and manipulations made by the user can be observed in real time.
- In fast processes the speed can be reduced to allow interactions, interventions, and manipulation.
- Individuals can interactively explore why and how specific decisions are made and find out how changes in the parameters, data, and models impact outcomes.
- It states how humans can benefit from the artificial intelligence.
- It explains what risks the artificial intelligence poses for individuals as well as on societal level.
- It is visible who has control of the artificial intelligence, in particular who has the power over data, models, and algorithms.
- It is visible what data, knowledge base, and information is used or has been used to create and inform the artificial intelligence.

With this definition and list of properties we aim to be inclusive with regard to different technologies and to ensure that humans can effectively explore and control how an AI works. We also ask that the underlying goals that motivated the design be made transparent. To verify that an AI is human centric it is essential to know the design goals and who has control of the data and algorithms.

3 LEVELS OF ABSTRACTION AND GRANULARITY OF CONTROL

Working with different levels of abstractions is one of the most powerful methods in computer science. For creating an Interactive Human Centered Artificial Intelligence abstractions and granularity of control play a major role in understanding and controlling systems.

3.1 Understanding Systems

Understanding an algorithm on a very abstract level is in many cases straight forward and does not required in-depth knowledge. For example, the explanation that a chat-bot has "learned" all the responses people gave on Reddit to a certain question and compiles its response based on this "knowledge" can be understood by virtually anyone. In most cases, this is explanation is good enough to understand the behavior of the chat-bot. In specific cases, individuals may want to dig deeper into the behavior, e.g. by looking at knowledge representation or learned weights.

Arguing for open sourcing of code and models may be misleading as this limits comprehension to a very small set of people. To understand systems that are built on current ML and AI techniques, looking at the implementation of the algorithm, the models, or the parameters of a trained network may provide little insight to even an experienced AI researcher. An algorithmic level of understanding, which we see in current AI explainability research, seems to focus again on technological curiosity of the experts who really want to understand how and why their neural networks works. But this will add little to making AI understandable for end-users.

Looking at the design goals and the functionality that was implemented with these techniques provides insights as to what outputs can be expected. Hence we ask that not only the source code be made publicly available, but also the design goals.

3.2 Staying in control

Control of our environment is of great importance for humans. Being able to determine what should happen is strongly related to self-determination and freedom of choice, and is ultimately a basis for feeling safe. The granularity of control is however of lesser importance. When going on a train or a plane we want to decide where we want to go, we want to have control of the target destination. In this case we are less concerned with the exact route taken or from which runway we will take off. In a conventional car there is a very fine-grained control. We have many choices, e.g. where we drive on the road (e.g. more to the middle or closer to the edge) or how fast we accelerate. Most of these choices are not relevant in general and we may be happy to hand them off to an AI that can handle the details of driving for us. As long as these choices are not relevant to what we want to do (typically arriving safely and in time at a certain place) we may not feel a loss of control – as we still stay in control on a different level of abstraction.

To achieve this, we argue that we must fundamentally rethink how we create systems that support humans. Requiring humans to interact does not scale with ubiquitous computing technologies. To overcome this problem, we have suggested a new interaction paradigm: Intervention User Interfaces [8]. Here the idea is that systems are well designed and run autonomously, but humans can intervene operation at any time. In this paradigm, systems must make their state observable and provide ways for humans to see how to intervene and predict the outcome of their interventions.

4 A NEW MATERIAL FOR TOOLS?

Is AI and ML just another material for creating new tools? Is there something fundamentally new when we can build systems based on AI? It seems AI is in a tradition of other materials that have changed human abilities and massively changed societies.

4.1 AI as basis for new tools?

Human history is an impressive account how tools have changed abilities. The invention of the wheel has changed our ability to transport things; changes in record keeping and writing tools fundamentally changed the human ability to remember things. We could consider AI and ML as a new technological basis – as a material for new tools – and that changes what tools can be made, see figure 1.

In history we have name these epochs by the base material (e.g. "stone age", "bronze age" or "iron age") and not by the tools that were created. AI could therefore be a new material that advances humankind. What tools we make out of it (good or bad) – an iron shovel to dig large irrigation canals or a highly effective battle axe – is obviously not a new question. It is interesting to see that epochs are name on materials: "stone age", "bronze age", or "iron age". Whenreferring to oral cultures and societies that could read and write we look at the impact of tools. With AI and ML the technical and the societal aspects may be combined.

Also interesting in this historic reference is that once a material was understood, the focus of the majority of people was on what to create and achieve with the material. Only a few concentrated on how to improve the materials. With regard to AI this would mean that the next focus is not on a new algorithm or model, but on the human ability to make use of it.

4.2 Automation vs. Augmentation vs. Amplification

Automation vs. augmentation is an old but still highly relevant discussion. In our work we added a further term: Amplification, see [5] and [6]. This is to highlight that we aim for improving and amplifying abilities at larger scale. With current technologies it becomes harder to make a clear discrimination. In a simplistic view we can see automation that serves a human need as augmentation, potentially on a larger scale of granularity.

Smart tools are one example. A broom augments our ability to sweep the floor. A manual vacuum cleaner is just advancing this tool to augment the ability even further. What is a vacuum cleaning robot? Is it just an advanced broom, where my need for interaction and control is further reduced or is it an automation? Here it becomes clear that this discrimination is in many cases no longer helpful. The questions shift to how agency and control are implemented and how this is manifested in the interaction between humans and smart tools. In these cases, we off-load larger and larger tasks to tools that are somehow intelligent.

We see similar developments in the space of cognitive and perceptual support. AI technologies provide means for augmentation or automation of cognitive and perceptual processes. Here too the border between augmentation and automation is blurred. Already Douglas Engelbart made a strong argument for keeping the human in the loop, for Augmenting the Human intellect [2] rather than for automation.

In the following we have several themes that show that there are opportunities for more general benefits for humankind through advances in AI. To highlight that this is an advancement we use the term amplify rather than augment.

5 HOW DOES HUMANKIND BENEFIT FROM AI?

Who benefits from AI? This question needs to be at the very center of any technological advancement. So far, the novelty and technological progress provide enough excitement to put this question to the back – even on a political level where large national investments are made in AI. At the same time, it is apparent that some individuals and many companies have a clear understanding how they can benefit from using AI and how their profits can be increased using ML-based technologies. Related to this question is also who will lose out from AI, and this too must be discussed and addressed.

5.1 Artificial Intelligence to Automate Mental Routine Tasks

Digital technologies have provided new tools that are based on automating routine tasks that are based on data. Examples are searching in big data bases for specific content or matching data from different sources. With AI there is a hope to automate routine mental tasks, e.g. automating decision making or classification tasks. In many areas this step is gradual rather than revolutionary. Text processing tools provide new smart corrections, a search engine

adds new features predicting what you look for, or a spreadsheet suggests graphs you may want to create. One central question that arises here is, what happens if we are not involved in routine mental tasks anymore? Will we lose skills, or what skills will we lose? Is this a bad thing or is this rather positive to free mental capabilities for more important and relevant tasks? Another issue is whether automation lead to more consistent behavior. Will it lead to less variety (e.g. texts become more and more similar as keyboards suggest next words or provide template texts for responses)? Research here is still in its infancy and we have to address these questions experimentally.

5.2 Artificial Intelligence to Amplify our Perception

Using artificial intelligence, we can change what we see and hear. This ranges from providing images that are processed to increase the human ability to see specific things, be it a tumor in a tissue or an enemy in the woods. AI can help to focus on the "right" things by directing attention based on information extracted from the environment. Another example is the use of sensors that are more sensitive than human senses (e.g. hearing) or capture information human senses can't (e.g. thermal imagining) and processing these signals for human perception. We have explored numerous research probes in the context of the European project AMPLIFY [7].

On a higher level, perception can be aided by providing the right amount of information to not overload the user or to provide it in ways that can be more easily understood (e.g. reading interfaces).

This amplification of perception is in the tradition of optical instruments (in particular the microscope and telescope), but with using AI as the material to create the new tools. By literally making things visible or perceivable that could not be seen before a new level understanding is provided. Once things can be perceived (e.g. bacteria in the example of the microscope), they fundamentally change how we understand the world.

6 ARTIFICIAL INTELLIGENCE TO AMPLIFY OUR MIND

Iron improved the strength of mechanical tools. Engines amplified the physical strength of humans through powered machines. Writing and reading had a strong impact on the memory capabilities of humans. With artificial intelligence we have the chance to create a change for improving the abilities of the mind, similar to what engines and motors have done for the body.

The ingenuity in amplifying physical strength was and is rooted in the devices that are built around the motor. In many cases the motor is very much exchangeable as the innovation lies in the system or device. Relating this back to AI and ML, we can see that it is important to have good algorithms and methods, but in order to make it useful for human kind we have to find the innovations that advance the abilities of individuals as well as of society.

6.1 Human Centered Training of Artificial Intelligence

To truly advance what we can do with AI it is not enough to observe the current world and learn from it. Scraping existing databases, document collections, or online forums is not a good idea, as we have seen from several examples. Observing the world as it is, is not enough. The question of how to transfer skills and knowledge to a system is not as simple as collecting data, as each data collection is biased. When training systems, humans will need to define goals and aims, which may be on very high and abstract level. We will always encode our values – explicitly through the algorithms and methods, or implicitly through the selection of data. It is important to make transparent exactly which values have been encoded as this aids in understandability.

6.2 No Opting Out of Artificial Intelligence

I would expect that individuals as well as societies will realistically have no opportunity to opt-out of using devices and interaction with systems that are driven by AI. Not using AI will create major disadvantages, as others will be faster, see things you do not see, and will be able predict events long before you encounter them. Opting out is like putting your head in the sand. Since opting out is not a viable option, we need to ensure that AI is working for humans and not against them.

Societies that will be effective in using AI and ML – by creating products and services around them – will have major advantages (and likely to dominate the others). The important part of this statement is that it is not just the societies that have AI, but those who are able to use it effectively to advance their cultures and economies who will succeed.

ACKNOWLEDGMENTS

This work was partly conducted within the Amplify project which received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 683008).

REFERENCES

[1] Janna Anderson, Lee Rainie, and Alex Luchsinger. 2018. Artificial intelligence and the future of humans. *Pew Research Center* 10 (2018).
[2] Douglas C Engelbart and William K English. 1968. A research center for augmenting human intellect. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I.* 395–410.
[3] John Etchemendy Fei-Fei Li. 2020. *Advancing AI research, education, policy, and practice to improve the human condition.* Retrieved August 20, 2020 from https://hai.stanford.edu/
[4] Paul Lukowicz. 2020. *European Project Humane AI: Human-Centered Artificial Intelligence.* Retrieved August 20, 2020 from https://www.humane-ai.eu/
[5] Albrecht Schmidt. 2017. Augmenting human intellect and amplifying perception and cognition. *IEEE Pervasive Computing* 16, 1 (2017), 6–10.
[6] Albrecht Schmidt. 2017. Technologies to amplify the mind. *Computer* 50, 10 (2017), 102–106.
[7] Albrecht Schmidt. 2020. *European Project ERC AMPLIFY: Digital Technologies to Amplify Human Perception (ERC Consolidator Grant no.683008).* Retrieved August 20, 2020 from https://amp.ubicomp.net/
[8] Albrecht Schmidt and Thomas Herrmann. 2017. Intervention user interfaces: a new interaction paradigm for automated systems. *interactions* 24, 5 (2017), 40–45.