HUMAN- CENTERED ARTIFICIAL INTELLIGENCE (HCAI): AN ETHICAL APPROACH

Stafylas Demetris

The department of Inventions of Informatics and Communications at the Hellenic Industrial Property Organization (OBI), Athens, Greece

Human-centered Artificial Intelligence (HCAI) systems designed with ethical machine algorithms is an auspicious prospect for systems focusing on human's needs and motivations. They will play a vital role at the safeguard of human's creativity and self-effectiveness, protection of privacy and equality, social justice and human rights. This article argues the importance of humanistic design in combination with rationalistic perspective for achieving high levels of human control and automation and helping simultaneously the mutual understanding between systems and humans. It is further argued the adaptation of an algorithmic fairness framework assuring individual's rights to achieve their goals free from prejudices of their social background and economic status.

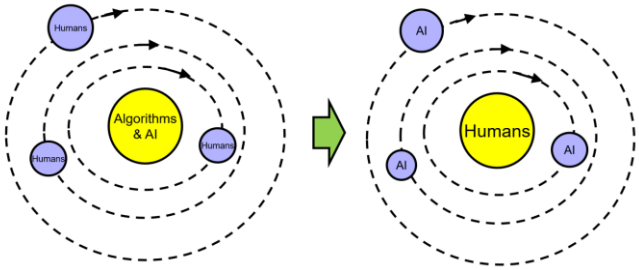
1 INTRODUCTION

Artificial Intelligence - A.I has already invaded in all aspects of life with the aim to change the way we work and live up to now. Advanced computational systems fed with digitized data and equipped with sophisticated algorithms come to support humans and provide solutions to their problems. Such a profound impact on the world has been caused by A.I that has given rise to the greatest fears at how an A.I enhanced future will look. A dearth of legal and ethical questions are linked to human's hesitation and uncertainty to trust A.I. The new automated machines which most of the times emulate human's behavior provoke a lot of concerns about what will happen at the worst hypothetical scenario where machines might exceed the ability of human's brain in general intelligence

2 HCAI DESIGN

The solution to all these threatening questions is given by HCAI which dynamically help people realize that there are other ways of thinking about future technology.

A different approach including fresh ideas that enable researchers to create new possibilities for future products and services has been inserted by Shneiderman, (2020). The main concept is revolved around the notion of changing HCAI's design thinking approach and putting the human users at the center focusing on user needs. It is claimed that only by breaking the shackles of old beliefs is it possible to have a two- dimensional HCAI framework where high levels of human control can be performed in combination with high level automation. What is more, it is necessary to stop making sophisticated machines emulating people but to empower people to take the control of said machines. Users who associate with autonomous, intelligent computers want to be in control of these technologies with the aim to improve their self-efficacy and their creativity.



A step in the right direction for implementing the above-mentioned ideas is to adopt a more humanistic design perspective. As is well presented by Auernhammer,(2020), Human-Centered Design puts at the center of design-thinking humans and their needs, motivations, behaviors and mainly their well-being. Some interesting design practices with a view to taking into consideration all the crucial factors for designing ethical HCAI systems are proposed as follows:

- Human-centered systems and social design. Human-centered systems examine the changes on the social systems such as organizations resulting from the A.I implementation and use. The critical aspects of the designer's ideologies as well as the cultural context in the design of the work system are covered by social design. In this approach, the designer is interpreted as the translator of the societal needs into A.I design system. Any design preference or choice made that impact society is included into the designer's responsibility.
- In addition to this, Participatory design and inclusive design point out the importance of the democratic procedures at the design and development of A.I systems. The participation of users from all walks of life and excluded communities can contribute to preventing machine biases in the design of the A.I systems. Developing diverse design teams, a meaningful relationship of trust can be created between people and A.I systems irrespective of

their gender, race and socieconomic status.

- Furthermore, Interaction design, Persuasive technologies and Human-centered computing enable the examination of the interaction and behavior of people with A.I. More specifically, interaction design approach examines the behavior of people and potentially harmful experience in the interaction with A.I system. It is identified the usefulness of the A.I system and how easily accessible is by recording people's meaningful interaction and attitude towards them. Persuasive technologies explain how human's attitude towards A.I systems can be altered by interacting with the advanced technology. Moreover, being examined the motivational strategies that are incorporated in A.I technology with the aim to persuade users to an intended behavior. Human-centered computing approach integrates people's lifestyles and the system design by viewing intelligence as an attribute of the composition of human's everyday life around A.I systems. The goal of this philosophy is the enhancement and support of human capabilities through computer systems.
- Finally, the Need-design response approach identifying human needs in the early phases of the design project through practices such as need-finding.

3 ETHICAL MACHINE LEARNING

Apart from the humanistic approach, Human-centered A.I needs also to focus on the rationalistic perspective. Discussing for HCAI it must be clear that the used machine learning algorithms consist a part of a longer human-centered system. Simply explaining, machine learning is a process based on collected sample data set with a view to training said data, in order to create that accurate model which allows computers, having acquired their own knowledge, to make decisions. As analyzed by Ried, (2020) is essential the decision-making algorithms to be able to understand the sociocultural norms of human behavior and simultaneously to help humans understand these algorithms. A conflicting point of view highlights the hazard of algorithmic decision-making by criticizing its potential to lead to privacy violation, information asymmetry and lack of transparency and discrimination.

In the publication by Lepri, Oliver, Pentland, (2021) is stressed that there is a dire need to build HCAI by ensuring privacy and data proprietorship, lucidity, and responsibility as well as fairness in A.I-driven decision-making processes. Only by reversing the risks of A.I when used for decision-making could societies exploit machine-learning decision-making procedures to generate a HCAI. A new user-centric model called data cooperative

has been introduced to preserve the fundamental right to privacy. Individuals take the absolute control of their personal data by handling their own data's life cycle and controlling the data collection to their data repository. On top of everything they have the right to inhibit access to the data repository. Transparent and accountable algorithms should be deployed by placing humans at the center of interest of the decision made via algorithmic means. The introduction of explanations could be a possible solution for overcoming the previous-mentioned drawbacks. Explanations either visual or through natural language processing should be used for providing enough information to users in order to understand the decision-making algorithms and indicating them to take a course of actions so as to remedy possible faulty situations. Noteworthy explanations regarding the behavior of a machine-learning model are the determination of the internal component's role of the model or the estimation of a complex model by means of a simpler one.

A safe way to maximize algorithmic fairness is to apply the individual fairness framework based on resemblance metric between persons. People who bear similarities should be ranked in the same way regardless of their social background and economic status. All people entitle to achieve their desired results through their selections, actions, and efforts.

Looking trends in A.I techniques, machine learning predominates according to the available patent data. Our society is witnessing an unprecedented momentous occasion where complex problems can be tackled through the use of algorithmic decision-making process. Due to the combination of machine learning algorithms and the availability of infinite quantities of human behavioral data, advanced A.I enables people to analyze and solve complicated issues.

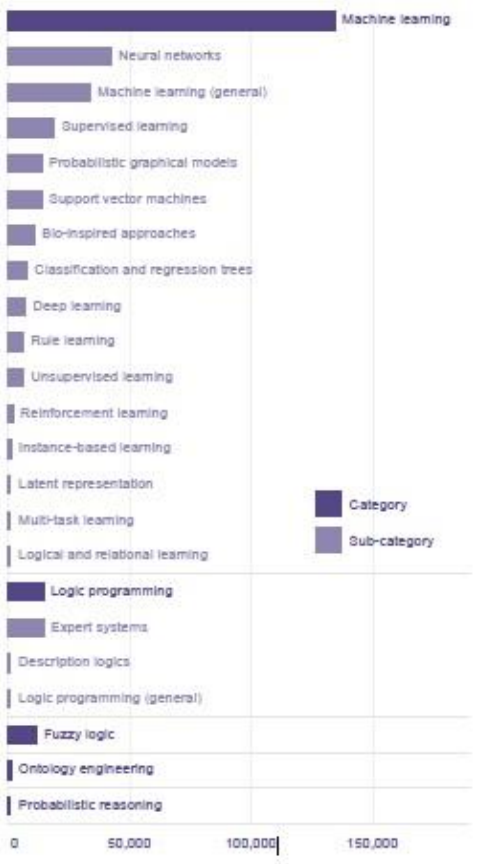


Figure Machine learning as a dominant A.I technique representing 89 percent of patent families

4 CONCLUSIONS

HCAI systems are essential to put humans at the center of system design thinking understanding human's expectations and needs. As a result, making easily accessible and understandable A.I systems by people, they will help users understand said systems in return. Furthermore, human-centered A.I needs to focus on both two integrated perspectives when designing A.I rationalistic and humanistic systems. Multi-disciplinary teams working together with expertise in areas such as machine learning, decision theory, ethics and law hold the opportunity to examine the impact of A.I on society and to build human-understandable A.I systems. Our scientific understanding of fundamental A.I and machine learning will be improved by overcoming fears related to the use of algorithmic decision- making processes including computational offence of privacy, information asymmetry, lack of accountability, injustice and bias.

REFERENCES

Mark O. Riedl (2019), Human-centered artificial intelligence and machine learning. https://doi.org/10.1002/hbe2.117

Ben Shneiderman (2020), Human-Centered Artificial Intelligence: Three fresh ideas. https://doi.org/10.17705/1thci.00131

Jan Aurenhammer (2020), Human-centered AI: The role of Human-centered Design Research in the development of AI. DOI:10.21606/drs.2020.282

Bruno Lepri, Nuria Oliver, Alex Pentland (2021). Ethical machines: The human-centric use of artificial intelligence. https://doi.org/10.1016/j.isci.2021.102249

World International Property Organization (WIPO). Artificial Intelligence. Technology Trends (first published 2019).