Human Centered Artificial Intellingence and Machine Learning

1st Hemant Rana

Fakultät für Mathematik und Informatik

Georg-August-Universität

Göttingen, Germany

hemant.rana@stud.uni-goettingen.de

Abstract—With the advent of Artificial Intelligence, AI systems have been involved in our daily lives. Hence it is quite important to understand the role of humans in such systems as the better understanding we have of these systems, better would be formulation of the problems they are solving and hence results from them would be more useful and definitive towards expected goals. This need for systems to better understand needs and goals of humans made Human Centered Artificial Intelligence(HCAI) an active research area. This report focuses on discussing aspects of HCAI primarily based on human-computer interaction that are effective to be employed in practice and also focus where problems arise in developing human-centered systems.

Index Terms—Artificial Intelligence, System Design, Human-Computer Interaction

I. INTRODUCTION

A. Artificial Intelligence

Artificial Intelligence deals with designing algorithms that perform tasks which when performed by humans require some form of critical thinking. Generally an AI system can be perceived as something that is indistinguishable from humans. Such intelligent systems are popularly known as agents with the ability to take decisions based on some final objective. Machine Learning being one of the sub fields of AI in which these decisions are largely based on data. Some popular examples being non-playable characters in computer games, voice assistants like Alexa, Cortana etc. which are more observable AI agents whereas some unobservable agents like loan-approval systems, recommendation systems on e-commerce sites etc. affect our daily lives without giving a sense that we have been interacting with them.

B. Human centered AI and ML

Having just improved algorithms doesn't necessarily mean that employing AI systems would result in greater good. Hence it is necessary to develop such systems based on the idea that there will be various stakeholders involved when this system will be deployed for use. Human stakeholders can be classified into two broad categories:

1) Operators: People responsible for developing AI systems, managing and maintaining it.
2) Users: could be the end-users interacting with the systems or clients for whom the system has been developed based on some specific use case.

HCAI primarily deals with designing systems that takes social responsibility of humans into account and takes following societal issues into consideration before making a decision

1) Fairness: refers to correcting both algorithmic and data bias while making automated decisions.
2) Accountability: the system developed will keep functioning properly and it works according to specific regulations and norms existing in the region where it is employed.
3) Interpretability: refers to understanding how the system reached it's decision.
4) Transparency: refers to the understanding of the system, the stakeholders associated with it, the algorithms employed, what data is used in the development etc.

C. Humans and AI

AI systems should be able to consider commonsense reasoning to interpret and predict actions of how people reach a specific decision. Common sense knowledge can be divided into following categories:

1) Declarative: some pre-defined set of rules that needs to be followed to reach a goal.
2) Procedural: following a specific set of actions in order to reach a goal.

[1] has defined following characteristics that should be fulfilled by a system in order to be human-centric.

1) Have an understanding of human socio-cultural norms.
2) are capable of producing explanations that non-expert end-users can understand.

Understanding aforementioned requirements can help design systems that are fair and avoids prejudices in situations which might not be anticipated by developers [1].

D. Assisting Humans in understanding a system

Even for expert users understanding black box systems like neural networks can be quite challenging. Understanding how a system reaches a result is necessary in order to make necessary changes to the system so it works in the intended way or provide some sort of compensation in case the outcome have severely affected a person or a group of people. For non-expert stakeholders, they may be more interested in achieving remedy. *Remedy* refers to the concept that a user should be

able to correct or seek compensation for a perceived failure. The primary step for achieving remedy is to get enough information in order to decide an appropriate remedial course of action. For eg. inspecting sensors, dataset bias etc. in order to understand why a system behaved in a way it did. Remedy can be achieved using the following

1) Explanation: provides a description of how a system came to a given conclusion. It can be further divided into following
   - Visualizations: a graphical representation of how algorithm works. For eg. Fig.1 shows the playground feature provided by Tensorflow deep learning framework that let user fine tune hyperparameters like learning rate, activation function, regularization function, ratio of training data and testing data etc. Additionally it provide option to design neural network architecture and highlights which neuron is activated for a specific epoch and an output.
   - Description of algorithm: providing a description how an algorithm processes data and reaches an outcome.
2) Generating rationales: For complex systems providing a rationale for a specific output/action just as humans do. But rationales fails to provide a reasoning for in depth working of a system.

II. Designing efficient Systems

[2] has conducted a review of interdisciplinary published research work in order to define competencies that should be possessed by the stakeholders and design considerations that must be employed during the process of development. These aspects can be undertaken for designing efficient AI systems.

A. Competencies of the stakeholders

1) **Recognizing AI**: AI can be defined as an activity that is devoted to making machines intelligent. Although definitions of AI can differ from source-to-source but having an understanding of what AI generally means in a specific context should be known by the user and operator.
2) **Understanding Intelligence**: Analyzing and highlighting features that makes an entity intelligent and comprehending difference between intelligence definitions for humans and machines. A possible definition can be *"thinking or acting either humanly or rationally"* [3].
3) **Interdisciplinarity**: There can be various interpretation and hence development procedures of intelligent machines based on type of technology employed or one's field of study. Hence identifying a variety of technologies that uses AI spanning over various fields including robotics, Machine Learning and cognitive systems is crucial.
4) **General v/s Narrow**: Distinguishing between whether an AI system is general or narrow. Narrow AI is developed for a specific field or use case whereas a general AI is something that can rival human intelligence across multiple domains.
5) **AI's strength and weaknesses**: [4] suggests people's trust in AI is strictly influenced by the task they perform. Hence having accurate knowledge of AI can help people in forming informed decisions about how to use and when to use AI for solving a specific problem. Hence identifying whether AI really is effective in solving a problem can help in planning the human's intervention in the system or how much intervention is required and what human skills can be employed in the system.
6) **Future Implications**: Understanding what AI can do in future can help individuals to explore novel ideas and critically evaluate the long-term effects a technology can have on the world.
7) **Knowledge representation**: stakeholder must be familiar with how knowledge is represented by computers and how are they processed by an algorithm. For eg. images are represented as 3-d matrices consisting of float values.
8) **Decision making**: Cognitive systems use various kinds of strategies to form decisions, although it is not necessary to understand all these strategies in detail but an abstract understanding of these strategies can augment in decision making by helping interpret and understand algorithms.
9) **Understanding machine learning**: understand the steps involved in machine learning, the practices involved in the field and challenges that occurs while employing this field. There can be certain misconception associated with what machine learning actually means based on the expertise of a user, their knowledge background or sources from which they collect information. In order to tackle misconceptions certain techniques for tackling them has been proposed including engaging in embodied interactions or enacting algorithms in order to understand them better.
10) **Role of humans in AI**: recognizing various roles a human can take in the development of AI systems. People can be involved in the development process by taking up several roles including programming, choosing models, fine tuning systems, garnering data, designing algorithms etc.
11) **Data Literacy**: [5] has put forward the following data literacy competenencies in information literacy standards

    - Ability to identify the context in which data is produced or recycled.
    - Ability to recognize source data value, types and formats.
    - Ability to know when data is needed.
    - Ability to access data sources appropriate to the information needed.
    - Ability to critically assess data and their sources.
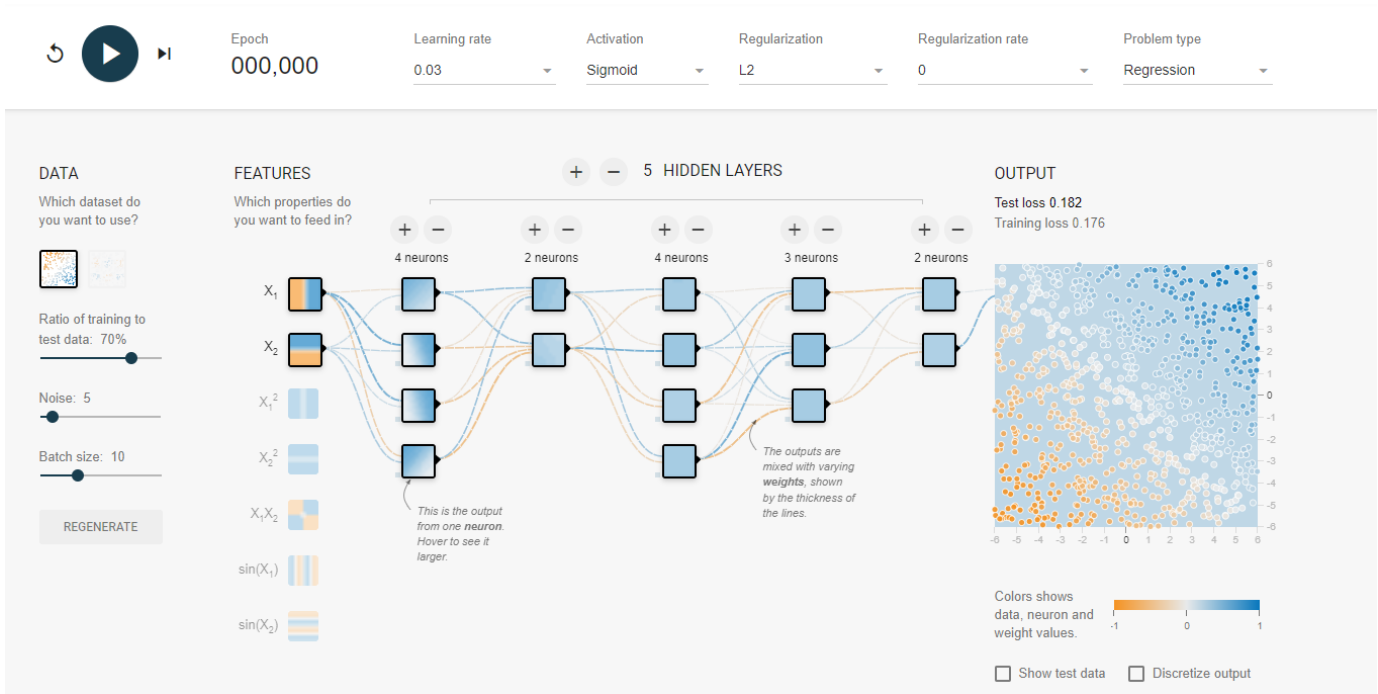   - Ability to determine and use suitable research methods

Fig. 1. Tensorflow playground User Interface

- Ability to handle and analyze data
- Knowing how to select and synthesize data and combine them with other information sources and prior knowledge
- Ability to present quantitative information (specific data, tables, graphs, in reports and similar)
- Using data ethically
- Ability to apply results to learning, decision-making or problem-solving
- Ability to plan, organize and self-assess throughout the process

12) **Learning from data**: recognizing that computer learns from data.
13) **Critically interpreting data**: data cannot be taken just based on it's importance but it requires interpretation. Determine how the training data provided in the initial dataset can affect the results of an algorithm.
14) **Action & Reaction**: acknowledging whether the AI system has the ability to act physically in an environment. The action can be a result of either reasoning or reactive stimulus like jumping.
15) **Sensors**: understanding what sensors are, recognize how it collects data, in which format it collects data and transmits it. It is necessary to understand as sensors can be present in heterogeneous environment, collect heterogeneous data and hence it is crucial in order to understand how data will be processed in a reconciled form for the system.
16) **Ethics**: having a fair understanding of the ethical issues associated with designing AI systems. The following aspects are related to ethics in AI:

- Privacy/surveillance: due to collection of large amount of data, there has been various concerns regarding privacy.
- Employment: although AI based filtering system has made business operations more productive but it poses concerns regarding bias.
- Misinformation: especially fake news on social media has been responsible for influencing masses by polarising their ideologies.
- Harm to people: the negative effects employing an AI system can have on humans.
- Diversity: in CS workforce diversity is quite a concern with majority of professors and AI based job applicants being male. Lack of diversity can cause certain types of biaseness that may lead to mistreatment of some groups, genders etc.

B. Design Considerations

1) **Explainability**: including graphical visualizations, simulations, explanations to provide a better understanding of the system and it's decision making. Additionally providing some interactive modules can help users to learn about the system.
2) **Interactivity**: enabling user to participate in the decision making process can help them understand the reasoning process of the system. Interactivity may include embodied simulation of algorithms or physical interaction with the system.
3) **Contextualizing Data**: providing information about how data was collected and who collected it and limita-

tions that are associated with it. Some aspects related to dataset in order to be comprehended by the user are relevant to real lives, are low dimensional and not perfectly categorizable.

4) **Promote Transparency**: primarily eliminating black box functionality. More importantly improving documentation, adopting explainable AI, contextualizing data and using following feature designs:
   - Interpretative affordance: characteristics of an educational resource that indicate if and how a particular learning behaviour could possibly be enacted within the context.
   - Sim city effect: systems shape's their outermost knowledge according to user's interest so that a path can be constructed to make user understand that concept in more depth.

5) **Unveil Gradually**: to prevent information overload at user's end, the exploration and explanation should be carried out in stages by providing user the option to inspect and learn about different system components starting with few components at first. [6] conducted a study among school students in Germany, Denmark, Sweden and United States, who first evaluated an intelligent agent and were asked questions regarding it. This work highlights the following guidelines in order to make current students inclusive of the AI development community and help them better understand what it really means
   a) Try experimentation with the currently existing technologies and preventing concepts or theory and while development debugging should be made quite intriguing.
   b) Design systems that deals more with logical reasoning and decision making rather than something that tries to mimic humans.
   c) Make the agent transparent and provide a provision to students with which they can explore the decision making process of the agent and allow them to raise questions.
   d) Provide ways in which students can program or instruct or customize the agent.
   e) The agent should provide an appropriate outcome for every input from the student so they can track the learning progress of the agent.
   f) Allow collaboration by letting other students to interact with the agents of other students.

6) **Programmability**: providing ways to stakeholders to program or provide some form of instructions to the agent. It is recommended to keep a low barrier in terms of programming and providing more emphasis on visual/auditory elements by employing strategies like Parsons problems and fill-in-the-blanks code.

7) **Development in stages**: consider the development of system in stages based on the user's demographics such as age, prior experience etc. as these characteristics affect the perception of AI.

8) **Critical Thinking**: encourage learners to ask questions about the intelligence and trustworthiness of the system.

9) **Identifying cultural, temporal and other differences**: it must taken into consideration that people from different cultural, temporal, value etc. backgrounds can have different interests and perceptions of AI. Recognizing such differences can encourage learners and motivate them to study more about the system.

10) **Monitoring Learning**: The operators or other involved stakeholders should have a provision for tracking and personalizing the learning experiences of the end users or their own progress.

11) **Social Interaction**: design attributes of a system that encourage learners to interact and collaborate for achieving common goals and solving shared problems.

12) **Leverage Learner's interest**: leveraging interest of users by employing entities in the system that user come across every day eg. current issue, common pastimes like games,music etc.

13) **Acknowledging Preconceptions**: taking into account that user's knowledge can be based on various sources that may be vague or faulty sometimes, hence the system should take this into consideration and address them in the best way possible so that users can be well informed about the topic.

14) **New Perspectives**: introducing different perspectives in the learning phase that may not be well represented in popular media eg: not so popular AI subfields, deeper discussions about major subfields of AI etc.

15) **Minimal or no barrier**: ensuring that people not from math or CS backgrounds can interpret the system by reducing the prerequisite knowledge required for studying or using the agent.

III. HUMANS IN HCAI: A SURVEY

[7] conducted a survey in order to get a better understanding about how people perceive AI and the gap that exists between the users and agents. The group size consisted of people from following backgrounds
- 75 AI developers
- 130 Users

Their study highlights various interesting results such as aspects that are necessary from the viewpoint of users are not necessarily crucial from the perspectives of developers and improving the agent from the perspective of user is quite crucial in order to make it more human-centric.

A. Motivation

There are profound negative effects of AI on the lives of people. Some concerning incidents that have impacted lives of common people are, a family losing their benefits after government's system failed to identify the members, An amazon employee lost his job after a performance algorithm predicted his performance to be below a threshold but the employee claims that his performance degraded due to aspects

he had no control over and having personal information leaked by a chatbot.

The problem lies within HCAI guidelines and design principles as it tends to focus on generally abstract and broad aspects like human values, ethics and privacy. They can be overly abstract and difficult to implement as in practice only aspects that are specific can be implemented. [8] discusses where AI lacks in providing concrete specifications that are required to develop a system

- Aims and assigned duties
- Personal history and norms
- Some strict guidelines to translate principles into practice
- Legal and widely accepted accountable mechanisms

whereas [9] exclaims that human centered AI systems should focus on the following aspects

- Enhancing human experiences
- Satisfying requirements
- Amplifying, augmenting and enhancing human performance.
- System being reliable, safe and trustworthy
- Support successful operations
- Encourage creativity
- Clear assignment of roles
- Encourage collaboration

And many more propositions have been present for quite a while which indicates that ethical guidelines are the underlying solution for making a system more human-centered but they are often too vague to be understood from the perspective of software engineers as clear requirements are needed in order to develop a software program.

A possible solution to the aforementioned problems can be achieved by formulating frameworks or guidelines based on the following classification of AI

1) Specific type of AI: agents that have been employed for a specific purpose. For eg: customer support chatbots etc.
2) AI in a specific context: agents that are employed in a specific field. For eg: machine learning algorithms used for medical imaging etc.

B. Study

The survey has attempted to answer the following research questions

1) Do developer priorities align with user experience?
2) Do developer priorities align with HCAI in principle including guidelines, principles and theories?
3) Does popular interpretations of HCAI align with user experience?

C. Method

The researchers distributed a survey to 205 participants online through a website. The participants belonged to United States and Australia. The participants had a background of psychology, cyber security, human-computer interaction, software engineering and machine learning students along with professional software developers.

The collected responses for the surveys were analyzed and 37 different labels were assigned majorly by authors while the remaining ones by other members of the team and coders. They employed a strategy called Light's kappa with $\kappa = 0.34$. After assigning labels all the assigners conducted a meeting discussing their results and refining it further.

D. Results of the survey

Table I depicts the labelling results from the conducted survey [7]. Most popular keywords being the following:

- **Functionality**: defines usability, usefulness and performance of an AI agent or system. As functionality has been top keywords for developer reviews and bad experience for users whereas being the second most prominent keyword in good experience for users. Hence it can be concluded that functionality is important for HCAI when good experience is considered.
- **Social Impact**: refers to the impact of AI on users and other stakeholders. relative to users, developers considered social impact to be less of a priority when developing the system is concerned. Whereas social impact has been an important label for both good and bad user experience , hence it can be concluded that it is an aspect that should be considered seriously while developing the system.
- **Understandability v/s understanding the user**: the agent should be understandable, explainable and intelligible. Making a system understandable has been not so relevant for developers and can sometimes be considered a negligible feature for user experience. For users a system understanding them is more important than understanding the system.
- **Ethics, privacy and security**: Some developers pointed out that there were some racial biases in some AI applications. Also they exclaimed that it is crucial to identify values that must be included in an AI that represents human values and make it work in a legal way. Also developers referred to security but didn't provide much descriptions on it.

Following are the key findings that have been concluded from the study

- Functioning was a defining feature of user experience with AI and was a key priority for developers.
- Social impact was a defining feature of user experience with AI particularly positive experience.
- Understanding was less important for users than being understood by AI.
- Ethics, privacy and security were important developer priorities but were not important from the perspective of user experience.
- Many users failed to come up with a positive and negative experience of AI during the survey.

Following are the implications for HCAI as mentioned by [7]

- Adapting an AI system to user's requirements is a vital part of making them human centered.

TABLE I
Labelling results from survey

Developers (important considerations)	Users (good experience)	Users (bad experience)
Functionality (31%)	Social impact (42%)	Functionality (39%)
Ethics (20%)	Functionality (35%)	Unsure (21%)
Social impact (20%)	Unsure (18%)	Understanding user (16%)
Data (13%)	User experience (10%)	Social impact (13%)
Creation process (11%)	Understanding user (7%)	User-AI interaction (11%)
Privacy (9%)	Impressive (7%)	User experience (7%)
Efficiency (6%)	Easy to use (5%)	Emotion (6%)
Security (6%)	Safety (4%)	Privacy (5%)
Understandability (6%)	Efficiency (4%)	Ethics (4%)
Coding (4%)	Emotion (2%)	Understandability (4%)

- HCAI requires a greater focus and clear conceptualization of impact of AI on all stakeholders.
- Although understandability, explainability and interpretability are important for HCAI, it is also important that AI systems make user feel understood.
- AI policies tend to reflect on avoidance of harm which is although crucial but is not sufficient to fulfill the goals of HCAI.
- More research is required to get an idea of how much people are familiar with the concept of AI.

E. Future Scope

The paper discussed the following points regarding future research

- Before conducting a study, proper definitions of terms should be shared with the users so they have a better idea of what kind of evaluation is required from them and hence they can better formulate their answers.
- This paper considered people from relevant educational background and countries like Australia and United States whereas future studies should be conducted including people of marginalized communities.

IV. Conclusion

This report focused on some preliminary definitions in section I and then described the roles of humans in an AI system and introduced possible problems that occurs in developing more comprehensible systems. For which section II dealt with proposing some possible design considerations and competencies that needs to be possessed by users in order to better understand a system. Section III, a survey attempted on defining the problems that are still associated with making the system human centric and proposed some key points for making these systems more human centric. Although a lot of research has been conducted on developing principles and policies for HCAI systems but there still lacks a provision for defining policies into definitive requirements that developers can program into.

References

[1] Riedl, Mark O. "Human-centered artificial intelligence and machine learning." Human Behavior and Emerging Technologies 1.1 (2019): 33-36.

[2] Long, Duri, and Brian Magerko. "What is AI literacy? Competencies and design considerations." Proceedings of the 2020 CHI conference on human factors in computing systems. 2020.

[3] Russell, Stuart J., and Peter Norvig. "Artificial intelligence. A modern approach. With assistance of Ernest Davis, Douglas Edwards. Global edition. Boston, Columbus." (2016).

[4] Pegasystems. "What Consumers Really Think About AI: A Global Study." (2018).

[5] Calzada Prado, Javier, and Miguel Ángel Marzal. "Incorporating data literacy into information literacy programs: Core competencies and contents." Libri 63.2 (2013): 123-134.

[6] Druga, Stefania, et al. "Inclusive AI literacy for kids around the world." Proceedings of FabLearn 2019. 2019. 104-111.

[7] Bingley, William J., et al. "Where is the human in human-centered AI? Insights from developer priorities and user experiences." Computers in Human Behavior 141 (2023): 107617.

[8] Mittelstadt, Brent. "Principles alone cannot guarantee ethical AI." Nature machine intelligence 1.11 (2019): 501-507.

[9] Shneiderman, Ben. "Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems." ACM Transactions on Interactive Intelligent Systems (TiiS) 10.4 (2020): 1-31.