Artificial intelligence in governance: experts and non-experts perspectives

Mesterséges intelligencia a kormányzásban: szakértők és nem szakértők nézőpontjai

ANGULO BAHÓN, Cecilio[1] – CAREGLIO, Davide[2] – DOMÈNECH ARGEMÍ, Miquel[3] – HERNANDO PERICAS, Francisco Javier[4] – SOLÉ PARETA, Josep[5] – VALLÈS-PERIS, Núria[6]

Abstract

Artificial Intelligence (AI) is today part of our lives as a powerful tool for life easiness. However, AI is not always an easy question and concern, especially when applied in governance issues regulating our lives, jobs, or social relationships. In the context of the Erasmus+ project HEDY - Life in the AI era, two focus groups have been conducted to discuss with experts and non-experts in AI the challenges, opportunities, risks, and expected impacts of AI in governance in our society. The main objective is to collect opinions and questions, concerns and debated ideas from different social actors, thus broadening the current debates in the academic literature.

Keywords

Artificial intelligence, governance, focus group, qualitative research

Absztrakt

A mesterséges intelligencia (MI) életünk része, mint annak megkönnyítésének hatékony eszköze. Az MI azonban nem mindig egyszerű kérdés és fogalom, ha az életünket, munkánkat vagy társadalmi kapcsolatainkat szabályozó kormányzási kérdésekben alkalmazzák. Az Erasmus+ projekt HEDY – Élet a mesterséges intelligencia korszakában keretében két fókuszcsoportot szerveztünk, hogy megvitassuk a mesterséges intelligencia szakértőivel és nem-szakértőivel a mesterséges intelligencia kihívásait, lehetőségeit, kockázatait és várható hatásait a társadalmunk kormányzásában. A fő cél az, hogy összegyűjtsük a különböző társadalmi szereplők véleményét és kérdéseit, aggályait és vitatott gondolatait, kiszélesítve ezzel a szakirodalom aktuális vitáit.

Kulcsszavak

Mesterséges intelligencia, kormányzás, fókuszcsoport, kvalitatív kutatás.

[1] cecilio.angulo@upc.edu | ORCID: 0000-0001-9589-8199 | full professor, IDEAI, Universitat Politècnica de Catalunya (UPC) | egyetemi tanár, IDEAI, Katalóniai Politechnikai Egyetem

[2] davide.careglio@upc.edu | ORCID: 0000-0002-7931-8147 | associate professor, IDEAI, UPC | egyetemi docens, IDEAI, Katalóniai Politechnikai Egyetem

[3] Miquel.Domenech@uab.cat | ORCID: 0000-0003-2854-3659 | full professor, Barcelona Science and Technology Studies Group (STS-b), Universitat Autònoma de Barcelona | egyetemi tanár, Barcelonai Tudományos és Technológiai Tanulmányok Csoportja, Barcelonai Autonóm Egyetem

[4] javier.hernando@upc.edu | ORCID: 0000-0002-1730-8154 | full professor, IDEAI, UPC | egyetemi tanár, IDEAI, Katalóniai Politechnikai Egyetem

[5] josep.sole@upc.edu | ORCID: 0000-0002-9411-6308 | full professor, IDEAI, UPC | egyetemi tanár, IDEAI, Katalóniai Politechnikai Egyetem

[6] nuria.valles.peris@upc.edu | ORCID: 0000-0003-4150-761X | postdoctoral researcher, IDEAI, UPC | posztdoktori kutató, IDEAI, Katalóniai Politechnikai Egyetem

INTRODUCTION

Artificial Intelligence (AI) is today part of our lives. We can be aware of its presence and interact with it for instance when we ask Siri to find a restaurant for us. But, in many other aspects, we are not fully conscious that AI is also there: for example, financial institutes leverage AI to identify potentially fraudulent activities in our accounts; AIs are used to track and predict environmental impacts in farm fields using data from satellite scanning and monitoring of crop and soil health; AI has become the main way that companies keep us safe from cyber-attacks. Those are only few examples and, according to several studies, the Covid-19 epidemic has expedited the adoption of AI throughout all sectors of the economy [1].

Nonetheless, AI is not all puppy dogs and rainbows. Many academics point out that the way AI tools are produced must change due to limitations in collaboration and inaccurate data assumptions, such as the unreasonable expectations that drive the usage of AI systems not robust enough. For example, inaction on AI prejudice has resulted in many injustices against entire groups of people, racial profiling, and other disturbing incidents. Deepfakes and the ability to create realistic videos, pictures, text, speech and other form of (social) communication have raised many ethical and legal concerns lately about the use of AI and its capability of manipulating human perceptions. In cybersecurity, bad actors have also access to AI tools, so the cat-and-mouse game continues. Video surveillance based on AI to recognise persons through their face, speech, walk or movement have also raised some concerns regarding privacy. The Amazon Alexa has recently suggested to a 10-year-olf girl to touch live plug with penny after the girl asked for a challenge to do [2].

In this scenario of pros and cons when dealing with AI, the implementation of a *governance* becomes fundamental. Governance refers to all governmental procedures: the formation, maintenance, and regulation of rules or activities, as well as the assignment of accountability. It is usually a collective problem done by the government of a state, by a market, or by a network [3]. In a nutshell, AI governance should close the gap that exists between accountability and ethics in technological advancement [4] and make sure that reliable boundaries within technology are set, so it does no harm and further aggravate inequalities incidentally while it operates.

In the context of the Erasmus+ project HEDY - Life in the AI era [5], two focus groups have been conducted to discuss with experts and non-experts in AI the challenges, opportunities, risks, and expected impacts of AI governance in our society. The objective is to collect opinions and questions, concerns and debated ideas from different social actors, thus broadening the current debates in the literature. This work summarises the outcomes of these focus groups.

The rest of the paper is organised as follows. Section 2 introduces the AI governance and its principles. Section 3 describes the adopted methodology adopted for the conduction of the focus groups. Section 4 presents the main findings and highlights the key ideas of the focus groups. Section 5 concludes the paper.

AI IN GOVERNANCE VS AI GOVERNANCE

When AI is included in the term governance, two different interpretations can be found: i) The use of systems based on AI in the governance, meaning the adoption of AI in service provision, policy-making, and enforcement in government practices and public-sector ecosystems [6]; ii) The governance of the AI, meaning the promotion of a proper institutional and legal framework for the development and use of AI [7].

Despite both are considering different topics, it is not possible to maintain a discussion about AI in governance without considering AI governance, because they work as communicating vessels. Thus, governance is understood here in reference to what is known as "AI governance", an idea composed of three components related to: a) the infrastructure - obtaining, storing and processing data; b) the application - the management of data; c) the utilisation – the decision-making and evaluation processes based on data.

Many other definitions can be found in the literature. For instance, AI governance is referred as "*a variety of tools, solutions, and levers that influence AI development and applications*" in [8], as "*the structure of rules, practices, and processes used to ensure that the organisation's AI technology sustains and extends the organisation's strategies and objectives*" in [9], as "*a set of processes, procedures, cultures and values designed to ensure the highest standards of behaviour*" in [10]. Probably, the most complete definition is available in [11], standing that "*AI governance is a system of rules, practices, processes, and technological tools that are employed to ensure an organization's use of AI technologies aligns with the organization's strategies, objectives, and values; fulfils legal requirements; and meets principles of ethical AI followed by the organization*".

Nonetheless, to be effective and provide the correct trade-off between company's strategies and objectives, legal requirements and ethics, many actors work on identify the main principles. For instance, Harvard University [12] created a visualisation map of 32 sets of AI principles. KPMG [13] provides four guideposts to help organizations ensure the proper governance of algorithms. Google [7] highlights five specific areas where precise, context-specific guidance from governments and civil society would help to advance the legal and ethical development of AI. In our work, a set of six AI principles are considered for AI in governance which are functionally algorithm-agnostic, technology-agnostic and sector-agnostic:

- **Accountability** requires a clear identification of who hold responsibility for decisions and actions when designing, developing, operating, and/or deploying AI system. It must be people or organizations that are ultimately accountable for the acts of AI systems, no matter how complex the AI system is.
- **Transparency** regards the ability to explain why an AI system behaves in a certain way in order to boost people's confidence and trust in the accuracy and appropriateness of its predictions.
- **Fairness** must ensure that AI systems are ethical, free from bias, free from prejudice and that protected attributes are not being used.
- **Safety** regards taking measures against both inadvertent and intentional abuse of AI that poses a threat to humans.
- **Human control** means that people need to be in one or more points in the decision-making process of an otherwise automated system.
- **Universality** principle recommends the definition and application of technical, clinical, ethical and regulatory standards during algorithm development, evaluation and deployment in order to have interoperability, cooperation and given level of quality, safety and trust.

Proactive governance measures are becoming more widely recognized as a differentiating feature for firms seeking to establish a reputation for trustworthiness. There are a number of worldwide frameworks on AI governance and ethics concepts. European Union issued the General Data Protection Regulation (GDPR) which includes a special set of rules that relate to a consumer's right to explanation when corporations employ algorithms to make automated choices. Nonetheless, it attracted also some controversial as does not afford a right to explanation of automated decision-making [14]. In this regard, the EU is likely to be the first to enact AI regulatory legislation [15]. The Algorithmic Accountability Act [16] in the US requires major companies with access to large amounts of data to audit AI-powered systems for fairness, privacy, accuracy, and security risks. A notable initiative is the Singapore AI Governance Framework. It is the first model developed in Asia and its strength is that it translates principles into a practical, operational framework for immediate action, decreasing the entry barriers to AI adoption. This framework is based on two factors: i) AI solutions should be human-centric, and ii) decisions made or assisted by AI should be transparent, explainable and fair.

METHODOLOGY

A focus group is a type of qualitative technique of data collection, in which a group of people, guided by a moderator, have a conversation and discuss around a common topic. It normally consists of a group of 7-9 persons who do not know one another. These individuals are chosen because they have particular traits that are relevant to the focus group's subject. By fostering a tolerant and caring environment that fosters many perspectives and points of view, without pressing members to attain consensus, the moderator uses the group and its interaction to learn more about a particular subject [17].

In our case, the subject under study is AI in general and AI governance in particular. In order to gather different opinions and contrasted ideas, during February 2022, we prepared and conducted two different focus groups: one focus group with only experts in AI and one focus group with only non-experts in AI.

For the case of experts in AI, 9 persons (7 males, 2 females) aged 35-70 years attended the focus group. We define the term expert as a person with a university degree, working for at least 5 years in the area of AI, digital society, human-robotic interaction or Industry 4.0, and at least 3 published scientific or professional articles.

For the case of non-experts in AI, 10 persons (7 males, 3 females), aged 22-70 years attended. These persons had no previous knowledge on AI and came from different sectors of the civil society, but with personal interest in technology advancements.

It is clear that this methodology has some limitations. Firstly, it is an analysis whose conclusions make it possible to identify the different interpretations and arguments socially available on an issue, but unlike quantitative analysis, its conclusions are not representative, but significant. Moreover, there exists the limitation of the heterogeneity of the focus groups since most of the experts were academics and the non-experts had a university degree; hence the outcomes may not represent the general population’s views on the topic. However, it is worth mentioning that we contrasted people’s opinions with the available literature and vice versa, so our findings are valuable and other similar works are likely to reach the same conclusions.

RESULTS OF THE FOCUS GROUPS

Structuring and mapping the topics

The structure of this article follows the structure of identified topics during the analysis of focus groups on the use of AI systems in governance. The content of the term discussion during the sessions has been analysed using the qualitative methodology of the Thematic Analysis (TA). TA is a qualitative method to identify, analyse and explain patterns (or fears) in data. It organizes and describes them in detail and also aims to interpret some aspect of the research topic. It is flexible in the theoretical approach, it is not associated with data processing technology and it is useful for analysing data generated through various techniques: interviews, focus groups, case studies, documentary texts.

As explained in [18], the TA depends on a series of decisions about the method. In the current phase of the research, due to the analysis of the topics for both focus groups, decisions taken about the method have been made based on the relevance in relation to the research question, not based on the prevalence or space in the discussion. In this way, the selected topics for articulating the identification of priority questions for citizens in the use of AI systems in governance have been organized into systems and a series of sub-topics, such as those shown in the scheme in Fig. 1.

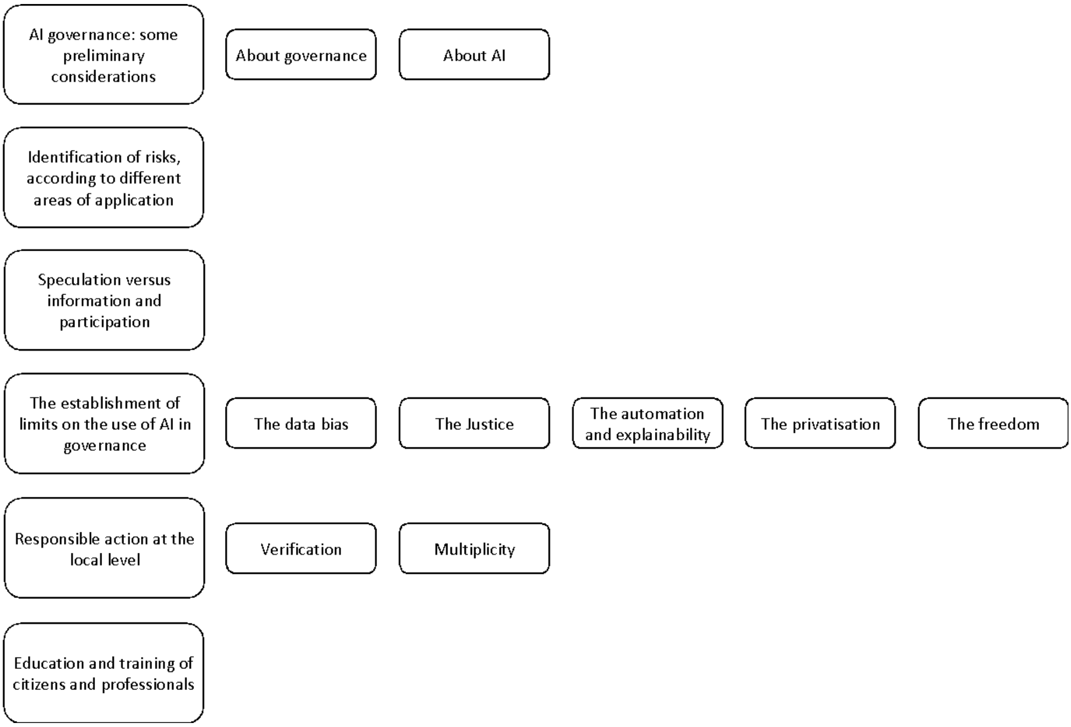
AI governance: some preliminary considerations
About governance
About AI
Identification of risks, according to different areas of application
Speculation versus information and participation
The establishment of limits on the use of AI in governance
The data bias
The Justice
The automation and explainability
The privatisation
The freedom
Responsible action at the local level
Verification
Multiplicity
Education and training of citizens and professionals

Figure 1.: Mapping of topics that have articulated the debate of the use of AI in governance

Some preliminary considerations

Prior to the analysis of the criteria that would guide the use of AI systems in governance, some general considerations should be introduced in order to contextualize the debate. These considerations refer to the very notions of "governance" and "artificial intelligence" that are going to mobilize the majority of focus groups.

- **About governance**: throughout the focus group, a definition of what is meant by the use of AI in governance has not been articulated. Therefore, a broad definition has been assumed that refers to the use of AI-based systems in decision-making processes, whether governmental, global or private.
- **About AI**: AI is conceived in two different ways in both focus groups:
    a) **Restrictive vision**: AI is another technology and, therefore, it can be treated as any other technology.

"AI is technology and a technology is not for everything, it is for what it is".
Participant in the expert focus group

b) **Disruptive vision**: AI is a different technology, which marks a turning point in human society and the relationship between technologies.

"It has come to change society and we will not be able to go back".
Participant in the expert focus group

These two ways of conceiving AI appeared alternately at the end of the debate, allowing the focus of attention to be placed on different problems and proposals. Therefore, it is considered that, even though the apparently restrictive and disruptive vision is understood as contradictory, in fact they are complementary visions that make it possible to cope with the complexities of opinions, concerns and proposals regarding the use of AI systems in the governance.

Identification of risks, according to different areas of application

From a restrictive point of view of AI, the dilemmas posed by this issue are limited and solved by establishing a very clear line as to why AI can be used or not. In this sense, it is considered that AI can be very useful for data management and analysis, or for information support for decision making and evaluation, but instead should not be used to make automated decisions. In this sense, it is considered that those decisions that directly affect people must be made by people.

In contrast, from a disruptive view of AI, in contemporary societies any form of governance integrates or will integrate AI. This ability of AI to be used in decision-making processes is applicable to several areas. It is in these specific areas that the risks of using AI systems need to be assessed. At the end of the focus groups, there has been a discussion especially about the risks in the business, communication or medical field.

- In the **business** field: The risk is related to finding a balance between the economic interests of companies and the non-violation of the rights of citizens in matters related to privacy and individual freedom.

- In the **communication** field: It refers to the proliferation of fake news or the aggravation of certain harmful behaviours for young people. In this sense, special attention is paid to the "loop" mechanism of social networks, which provide content and information related to the history of searches and interests expressed by users in their use of social networks. A group that is considered particularly vulnerable in this regard are young people and children, who are highly influenced by this “loop” effect of communication made possible by AI-based systems.
- In the **health** field: This is an area in which AI is considered as a technology that enables an intense improvement in diagnostic processes, an area in which benefits are preferred to risks.

Speculation versus information and participation

A shared concern, which is mostly associated with a restrictive view of AI, and which appears both explicitly and implicitly throughout the focus groups, is related with the relationship of AI systems to science fiction imaginaries and with the idea that AI can solve all problems of any kind. Many applications have been developed in the field of AI, and they can be applied to many fields, but there is a significant gap between the current technical capabilities and functionalities and the narrative about what AI could do in the future.

This type of narrative around AI, which does not correspond to current developments, is considered to have two types of negative effects:

a) On the one hand, the difficulty in articulating a contrasted public debate on responsibility when forms of AI are used in the decision-making process and;

b) On the other hand, the emergence of a series of catastrophic imaginaries that generate reluctance towards AI among public opinion and citizens.

In order to avoid this type of narrative and its effects, actions related to information and citizen participation are proposed:

- **Information**: Ensure that the mass media report in an ethical and honest manner when talking about AI systems, which allows a clear differentiation between speculative futuristic visions and current developments and possibilities. Develop an educational task that allows citizens to learn how AI works and what applications are being developed.
- **Participation**: Involve citizens in the establishment of AI development priorities, at the service of needs. It is considered that the participative dimension can be the added value of the European strategy for the development of AI, with respect to other strategies that may be more advanced in terms of technology or implementation, such as the case of China or the USA. It is considered that the European strategy can incorporate as an added value to its AI the integration of citizens in the establishment of priority areas in which to develop or apply it.

> *"With artificial intelligence, citizenship is needed. And I think that AI technicians are still not clear about this... either they don't realize that citizens are very important in various aspects of AI research and implementation, or it's not valued.”*
>
> *Participant in the expert focus group*

The establishment of limits on the use of AI in governance

There is a widespread consensus on the need to discuss the limits in the development of AI systems, because their use can have very important negative consequences for people's lives, or reproduce social models that are considered morally reprehensible.

> *"A research [project] to recognize a person based on the iris was financed through tax haven funds, to identify women with burqa and to know whether or not they were with their husband. I was very surprised [...]. How should it be done? Get here, yes? Get this far, right? What limits?"*
>
> *Participant in the non-expert focus group*

The limits, however, are not clear, and it is difficult to establish or agree on an ethical, political or regulatory framework that can regulate the development of forms of AI that can then have a high impact on social decisions. One of the difficulties that emerges in this regard, especially from a disruptive view of AI that understands more problematically everything to do with limitations to the development of AI, is the tension between a series of guarantees for the citizens and, at the same time, competitiveness in research and innovation.

In order to organize the definition of limits, especially in the focus group of experts, throughout the discussion the ability to intervene in decisions is considered in three different stages or stages:

1. In the management of the data that allow the decision to be taken.
2. In the evaluation of the decisions taken.
3. In the decision itself – a stage that, at the outset, is considered to be exclusive to humans.

> *"At the end they are algorithms and we shouldn't let them decide for us".*
>
> *Participant in the expert focus group*

In order to limit the use of AI systems in decision-making processes and/or to establish how this use should be carried out, in both focus groups issues related to: data bias, justice, automation of decisions and privatization.

The data bias

As specified at the beginning of the paper, the analysis of the social and ethical considerations of AI governance is inherent in the analysis of the use of AI in governance, an idea that captures the concept of data governance.

For this reason, in any decision-making process in which AI systems are used, participants from both groups emphasized the need to ensure that the data collected is not biased by gender, socio-economic level, ethnicity, etc. Guarantee of data diversity and its composition refers to the use of AI in all stages of the process, data collection, the decision itself or the evaluation.

> *"Humans make many decisions based on an ideology (...) A machine will also make a biased decision. Biased by whom? Because of the data, because of the engineer*

> *who designed it or the company behind it, or the ideology of the state that financed it"*
>
> *Participant in the expert focus group*

As illustrated in this quote, concern about how databases are built responds to the idea that any decision-making process, more or less automated, is biased, so there is ideology. Despite the apparent supposed neutrality of AI and other artifacts, the use of machines for decision-making is not exempt from this ideology underlying any decision. These ideologies can represent interests of various actors, being them of a political, technical or economic nature. This is an important issue to be solved in order to guarantee that collected data and their use respond to the objectives for which they are designed.

The justice

AI systems mainly work based on data compilation and statistics relationships. Beyond the data used, automated decision-making, regardless of whether or not the data is biased, poses a problem of justice, because the criterion of justice prevails over efficiency.

> *"[The AI] decides based on statistics. I am a fan of Rafa Nadal. If we were to pay attention to the statistics, he would not have won and he won. It is not fair that, in a case of conditional freedom, statistics are applied. It should be banned. We are forgetting the human factor, which AI does not take into account. AI is only the rational part, everything else, emotional intelligence, where is it? This is very important".*
>
> *Participant in the non-expert focus group*

Using the ability to handle large volumes of data and make statistical predictions is seen as an important value of AI. This is information to be taken into account when making contrasting decisions. However, this information cannot be used to make automated decisions that affect aspects directly related to people's lives.

From a more disruptive view of AI, it is assumed that even if we do not want AI to participate in numerous aspects of our daily lives, it is necessary to make an assessment of the costs and benefits, based on valuing what if the decisions made by AI systems were wrong. If the decisions affect non-substantive issues for people's lives, this error in the AI's decisions can be considered a minor issue and therefore, the AI could be used to make decisions on that particular issue. On the other hand, whether decisions affect substantive issues of people's lives, a wrong decision could have terribly unfair effects that would condition the person's life and, therefore, in that matter the decisions should not be made by systems day.

> *"Over the years we have built an important judicial system, which we want to maintain. There are areas in which the impact [of decisions made by AI] on the person is very important. AI should not enter this area."*
>
> *Participant in the non-expert focus group*

The automation and explainability

Decisions are currently being automatically taken in several areas, even though AI systems are not used. There exist numerous processes in public administration that are already highly standardized involving a significant volume and time of work. Continuing with the example of the legal field:

> *"In justice, a large part of a judge's time is spent issuing very standard sentences. Less complicated decisions, for example on commercial issues, can be delegated to algorithms. 90% of the sentences are very simple."*
>
> *Participant in the non-expert focus group*

During the discussion, the participants point to a process of automating processes that goes beyond the development and use of AI systems. In other words, in relation to automation, a restrictive view of AI is assumed, because what is considered truly disruptive is the introduction of automated systems in more and more areas of our lives. This process, which has to do with the definition of standardized indicators and the difficulty of negotiating some processes, is prior to the popularization of AI systems. Therefore, the debate about limiting the automation of decision-making processes cannot be limited to AI, in the same way that AI cannot be considered solely responsible for the automation of decisions.

The problem with AI is when those who design an algorithm are not able to explain its decisions, as well as when users do not know criteria that AI designer has implemented into the algorithm. Regardless of the final decision or prediction, guaranteeing the transparency and explainability of the entire process is essential in order to be able to use AI systems in governance.

The privatization of governance

One of the main concerns in the use of AI systems in general and especially in the field of governance, which has appeared especially in the panel of experts, is the important control of data and the accumulation of knowledge that some large companies or corporations currently have. Given the high economic and technical capacity increasingly necessary to make intensive use of data, this phenomenon poses a threat to democratic decision-making.

Certain companies or corporations are accumulating a lot of algorithmic knowledge and about the behaviour of the population, which implies a lack of guarantees that these data or this knowledge is carried out respecting principles or agreed ethical values. In this sense, the accumulation of data and knowledge in AI by entities outside the scope of government supervision means the privatization of governance, an issue that should be corrected.

> *"We have to think carefully about the part of the relationship with humans and how we organize ourselves in a different way to favour AI for the benefit of people, not for the benefit of companies."*
>
> *Participant in the expert focus group*

Faced with this situation, and in order to guarantee an AI that makes fair decisions and that respects democratic values, it is necessary to align the three legs that are considered to make up the governance of AI (citizenship, technology and administration). With this intention, apart from developing legal regulations, it is proposed to carry out data and algorithm audits on private companies.

> *"I believe that regulatory institutions should be created, in the same way that there are institutions that regulate banks and audit them to see what they do with the money. You should audit these companies like Google, Netflix and such, to see what their algorithms are really doing."*
>
> *Participant in the expert focus group*

The freedom

Freedom is one of the topics of most concern in both focus groups, as AI is considered to be a very powerful instrument for social control.

> *"It is a very powerful tool for control"*
>
> *Participant in the expert focus group*

The threat to freedom posed by the use of AI systems in decision-making processes can be understood from two different levels. The first dimension refers to the strategies that use AI to achieve greater advertising or visualization, based on algorithms that make users enter loop-type processes, which are used by Meta or Twitter-type companies. This type of process can mean a significant manipulation of some groups of people who are more influenceable or less educated, such as the youth. In this dimension, it is considered necessary to legislate the operation of these loops to avoid harm to people.

> *"I have teenage children, who believe what they see: the fake news, the bleach they drank to cure themselves of covid. I have a 12 years old daughter. I see that the information they see is a brutal danger. People are impressionable and this is very complicated. When you start to see a content, when we are young, we look for news that is what you expect, we are more influenceable. If you see a video that comes out... Well, you say 'I want to go to Malibu', 'I want a Prada bag'. The algorithm moves you."*
>
> *Participant in the non-expert focus group*

The second dimension, related to the first but taken to the extreme, has to do with a very disruptive vision of AI. In this sense, it is alerted to the ability of AI to control emotions and regulate feelings. Taking into account the digital trail that all citizens leave in all their daily movements, obtaining and using these data for commercial or authoritarian purposes can be very dangerous. According to this view, the problem is not the predictive capacity of AI systems in governance processes, but the use that can be made of these predictions. Faced with this situation, the solution proposed by the participants starts from questioning the supposed objectivity of the predictions and, therefore, proposes a use of the predictions based on subjective and contextual criteria, which can be known, negotiated and discussed.

Responsible action at the local level

AI changes the scale of decisions, has global effects, and therefore global control measures are also needed. This nature of AI transforms the way we understand governance and the ability we can have to govern its effects. Global control and regulation mechanisms are needed, but at the same time, there is a need to develop local mechanisms that favour responsibility.

Verification

In the focus group of experts, the idea is raised that decisions about whether we should use AI in governance in one area or another and in what way they cannot be definitive ones, because we do not have sufficient knowledge about its effects and their consequences. One of the great difficulties in order to introduce ethical and responsible criteria in the use of AI systems in decision-making processes is their global and intertwined scale. Faced with this situation, a response based on the development of small-scale forms of experimentation and monitoring is proposed. In this way, the responsible decisions must come from the result of the application of testing processes implemented in a controlled manner in very limited local areas. These controlled tests make possible to know the repercussions of the use of these technologies in specific areas and different cases. Since AI has global effects, it is difficult to think on a local scale, and it is precisely this scale that must be introduced in governance.

Multiplicity

This issue is also related to the different forms of technological development that AI is adopting. Just as in politics there is more than one model (different parties with different ideologies proposing different actions), AI for governance must also represent this diversity. There is no single technological answer. This proposal developed during the discussion represents a powerful alternative to the technocratic determinism that often accompanies AI: Technology gives us tools to find the best solution, but there are always many better possible solutions. In this sense, it is considered essential to accompany the emergence of open-source experiences, experimental techniques, etc. that allow the development of bottom-up strategies that represent this multiplicity of possibilities that AI can offer in governance.

Education and training of citizens and professionals

AI is a technology that in its design and development is so far removed from everyday life, that among experts it is considered that the population is not sufficiently educated to be able to make decisions about how AI should be used. Although, at the same time, it is considered that the public needs to make decisions and decide the course of AI. For this reason, the group of experts and non-experts points out the need to train citizens in the operation, potential and possible effects of AI.

> *"We must have an educated population"*
>
> *Participant in the expert focus group*

> *"Rules must be put in place and citizens must be at the centre... and these citizens must be educated. There must be ethics in AI. And engineers don't have to do it"*
>
> *Participant in the expert focus group*

In the same way, AI experts themselves consider that they too do not have sufficient knowledge to be able to decide on ethical and social issues, a knowledge that should be integrated in an interdisciplinary manner.

> *"I think that we lack more technical people, more knowledge about the evolution of society [...]. And also on the other side, to the people who are more in the field of governance [...] who also understand this new colleague that they have on the way everywhere ... At an educational level, we must try to make an effort to integrate this AI into the entire knowledge base out there."*
>
> *Participant in the expert focus group*

CONCLUSIONS

In the framework of the Erasmus+ HEDY – Life in the AI era project, we have conducted two focus groups with experts and non-experts in AI to discuss the impact of AI governance on our society. Focus groups are unique tools in qualitative research where the interaction of participants allowed the organisers to collect different social actors' opinions and questions, concerns and debated ideas, thus providing complementary information to that available in the literature.

We have identified for instance that there are two different yet complementary visions that we called restrictive and disruptive that make it possible to cope with the complexities of opinions, concerns and proposals regarding the use of AI systems in the governance.

From a restrictive point of view, the dilemmas posed by the utilisation of AI can be limited and solved by establishing a very clear line as to why AI can be used or not. There is also common association of AI with science fiction imaginaries and the idea that AI can solve all problems of any kind. In contrast, from a disruptive view, AI marks a turning point in contemporary societies with no possibility to go back.

There is however a prevalent agreement for both visions that it is vital to talk about the boundaries of AI system development since their use may have gravely detrimental effects on people's lives or may replicate ethically dubious societal paradigms. The boundaries are vague, and it is challenging to come to an agreement on an ethical, political, or legislative framework that can control the growth of AI. The conflict between a number of guarantees for the citizens and, at the same time, competitiveness in research and innovation is one of the challenges that arises in this regard.

For example, one of the main concerns in the field of governance is that certain large companies or corporations are accumulating a lot of knowledge about the behaviour of the population, which implies a lack of guarantees that these data or this knowledge is carried out respecting principles or agreed ethical values.

In conclusion, AI is a technology that in its design and development is so far removed from everyday life that the experts believe that the population is not trained enough to make decisions about how to use AI. The experts themselves consider they do not have enough knowledge to decide on ethical and social issues alone. For this reason, it is considered necessary that citizens should be able to make decisions and decide on the course of AI. Teachings, courses and trainings in schools and higher education institutes are needed to train citizens in the operation, potential and possible effects of AI and to facilitate the use and adoption of AI for young people and future generations.

REFERENCES

[1] M. Malluk Batley, "AI adoption accelerated during the pandemic but many say it's moving too fast: KPMG survey", *Thriving in an AI World*, KPMG study, March 2021, Accessed on August 2022, https://info.kpmg.us/news-perspectives/technology-innovation/thriving-in-an-ai-world/ai-adoption-accelerated-during-pandemic.html.

[2] BBC news, Alexa tells 10-year-old girl to touch live plug with penny, December 28, 2021, accessed on August 2022, https://www.bbc.com/news/technology-59810383.

[3] M. Hufty, "Investigating policy processes: the governance analytical framework (GAF)", U. Wiesmann, H. Hurni, et al. eds. *Research for Sustainable Development: Foundations, Experiences, and Perspectives*, Geographica Bernensia, Bern, 2011, pp. 403–424.

[4] KOSA AI, "The importance of AI governance and 5 key principles for its guidance", Accessed on April 2022, https://kosa-ai.medium.com/the-importance-of-ai-governance-and-5-key-principles-for-its-guidance-219798c8f407.

[5] HEDY project, *Life in the AI era*, KA220-HED 0C8D3623 - Cooperation partnerships in higher education, Accessed on August 2022, https://lifeintheaiera.eu.

[6] A. Zuiderwijk, Y.-C. Chen, F. Salem, "Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda", *Government Information Quarterly*, vol. 38, no. 3, July 2021.

[7] Google, *Perspectives on Issues in AI Governance*, Accessed on April 2022, https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf.

[8] J. Butcher, I. Beridze, "What is the state of artificial intelligence governance globally?" *The RUSI Journal*, vol. 164, n. 5-6, pp. 88–96, Nov. 2019.

[9] J. Schneider, R. Abraham, C. Meske, J. Vom Brocke, "AI governance for businesses", arXiv:2011.10672 [cs.AI], Nov. 2020, Accessed on August 2022, https://doi.org/10.48550/arXiv.2011.10672.

[10] A.F.T. Winfield, M. Jirotka, "Ethical governance is essential to building trust in robotics and artificial intelligence systems", *Philosophical Transactions of the Royal Society A*, vol. 376, no. 2133, 20180085, Nov. 2018.

[11] M. Mäntymäki, M. Minkkinen, T. Birkstedt, M. Viljanen, "Defining organizational AI governance", *AI and Ethics*, Feb. 2022.

[12] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, M. Srikumar. "Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI" *Berkman Klein Center for Internet & Society*, Feb. 2020.

[13] M. Sokalski, "The shape of AI governance to come", KPMG, December 2020, Accessed on April 2022, https://assets.kpmg/content/dam/kpmg/xx/pdf/2021/01/the-shape-of-ai-governance-to-come.pdf.

[14] S. Wachter, B. Mittelstadt, L. Floridi, “Why a right to explanation of automated decision-making does not exist in the general data protection regulation”, *International Data Privacy Law*, vol. 7, no. 2, pp. 76–99, May 2017

[15] European Commission, *Regulation laying down harmonised rules on artificial intelligence*, April 2021, Accessed on August 2022, https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

[16] US Congress, *Algorithmic Accountability Act of 2019*, H.R.2231, 116th Congress, April 2019, Accessed on August 2022, https://www.congress.gov/bill/116th-congress/house-bill/2231.

[17] R.A. Krueger, M.A. Casey, “Focus groups: a practical guide for applied research”, Newbury Park, Sage Publications, Aug. 2014.

[18] V. Braun, V. Clarke, “Using thematic analysis in psychology”, *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77-101, 2006.