October 24th, 2023

Recommendations

Access to justice through AI

By and for marginalized and/or under-represented communities

Participation of Marginalized and/or Under-Represented Groups:

Guide to supporting the inclusion of their needs and values

Karine Gentelet, Associate professor, UQO

Marie Zumstein (PhD Candidate, UdeM) and Lily-Cannelle Mathieu (PhD candidate, McGill)

INDEX

1. __Introduction:__

Research carried out as part of the Project for the Empowerment of Marginalized Peoples ("Chantier sur l'autonomisation des groupes et communautés marginalisées") examined the reasons why some tensions may arise when deploying AI tools dedicated to access to justice for marginalized populations. A summary of the initiatives that have already been put in place on themes surrounding social justice has thus enabled us to identify several elements to consider regarding the conditions for involving populations in the design, deployment or use of AI tools:

- Justice by design
- Witness through data and activism through evidence
- Popular education and access to information
- Citizen mobilization
- Counter-surveillance
- Counter-mapping
- Representativeness, inclusiveness and data sovereignty

Indeed, these tools developed by and for marginalized or/and poorly represented groups and communities enable us, based on their representations, to determine the most important conditions for the effective and efficient inclusion of these groups and communities within AI-based access to justice projects.

2. __Research project, methodology and results__

The purpose of this research was to study social justice initiatives already in place in different regions of the world, targeting marginalized and/or under-represented communities. The overall aim is to mitigate the potentially negative impacts of AI on bias and discrimination (see Annex 1 - Impact 1), a recurring concern in AI social ethics. As marginalized and under-represented populations already live on the margins of society, this situation must not be exacerbated by the creation and deployment of new AI tools for justice that could run counter to their rights to dignity and non-discrimination.

To this end, the question for us was: "How do marginalized groups and communities approach justice issues through digital technologies?" By starting with their own conceptions, we were able to learn about their priorities, their needs, and their methods of addressing them.

Initially, qualitative research was conducted on the internet between 2019 and 2021, using keywords in English and French. The intention was to identify social justice initiatives that used and developed digital and/or AI technologies _by_ and _for_ underrepresented, marginalized communities and groups, as well as by civil society organizations. This was followed by a dozen semi-structured interviews for the period 2021-2022. Stakeholders came from a variety of genders and backgrounds, both socio-cultural and geographical, to ensure the broadest possible representation.

3. **Guiding principles for effective representation and inclusion in access to justice projects**

Before turning to the recommendations themselves, we feel it is essential to present the results of the research. Based on analyses of the corpus of the pre-collected data, we have identified several principles put forward by research participants in their representations of tools that promote justice. In this sense, these principles constitute the foundations upon which any inclusive action must rest:

- Consider social justice issues well before dealing with issues of access to the courts. Marginalized groups and communities represent citizens who are under-represented in public spaces but over-represented in the justice system. This societal imbalance cannot be ignored in AI-based access to justice projects.
- Treat data representation issues as a factor of social exclusion and structural inequity. The challenges of data representations are decisive issues that structurally precede AI developments. These representation issues are societal well before being technical.
- Build a feeling of trust in these initiatives, their developers, and their operators by including people from the community in the early stages of the project as well as within the governance structures.

- Develop a solid understanding of the needs, values, objectives and internal dynamics of the groups and communities served, prior to their involvement in the project/initiative/tools. This can be achieved through regular meetings with the group in question, or through the development of collaborative tools, so that each party can express its needs and limitations. This knowledge and understanding will foster a collaborative, preventative approach and a long-term vision of potential societal problems affecting the community or population group concerned (as opposed to the reactive approach of the public and private sectors).
- Offer a local response to the needs expressed by the populations concerned (regardless of economic, political, or technological imperatives: the group's social needs and human rights must take precedence). It should be kept in mind that the needs of each group do not necessarily go hand in hand with the needs of society in general, or even of other communities and minorities.
- Objectively support the principles presented as fundamental by all the groups we interviewed when designing the tools: respect for people's self-determination, autonomy, and consent in the decision-making process. This must also include a right of withdrawal.
- Set up horizontal governance structures to leverage local interactions.
- Continuously include these groups and communities in the governance of their data and the tools deployed for them, so that they can participate in all decision-making.
- Jointly determine the choice of certain partners for the project and respect the right to exclude some of them due to pre-existing socio-historical contexts of discrimination or over-criminalization. Otherwise, the active participation of certain specifically targeted groups could be compromised.

We will look below at how digital tools can be deployed through human rights education. Indeed, it is important that the various parties and partners are made well aware of the best ways to include and involve marginalized and/or under-represented groups and communities in a collaborative and horizontal way when AI tools are developed and deployed for them. This is how their needs and rights will be respected in this new technological era.

4. **Recommendations/operating principles for the effective participation of marginalized groups and communities in AI access to justice projects**

The research results have enabled us to identify several recommendations that will help operationalize the principles presented above. These practices should make it possible to develop AI tools for access to justice that are more relevant, inclusive, and respectful of marginalized and/or under-represented groups and communities:

1. Clearly *identify* the group or the community potentially impacted by the AI tool to be deployed, taking into account the historical and systemic context of discrimination and/or human rights violations in the population concerned.

2. *Develop* a model specifically designed for this group or community. It is not conceivable to apply the same model in the same way to all communities.

3. *Carry out* an in-depth audit of the data that will be used in the development and deployment of the AI algorithm to ensure that the target population is effectively and adequately represented, responding to issues of discrimination. Indeed, the issue of data use is not independent of that of representation and visibility of minorities and under-represented populations. It is also fundamentally important for groups that are still systematically poorly represented in the data.

4. *Consult*, prior to project design, with representatives of the targeted community for which the AI tool is intended, in order to include the group's needs, values, objectives, and internal dynamics. This step is all the more crucial as it enables us to provide a locally adapted response that respects the experiences and internal governance structures.

5. *Involve* representatives of the community or group affected by the project in its further implementation stages, for instance by setting up a citizen steering committee. Care should be taken to ensure that this committee is representative of the composition, structure and functioning of the community in question. This initiative not only ensures recognition of the value of organic structures and an inclusive process, but also the implementation of

actions and decisions rooted in experiential and concrete knowledge and the response to real and useful needs.

6. Then, *ensure* an ongoing consultation process and the effective participation of this citizen steering committee at all stages of AI tool development and deployment. The effective participation of this committee fosters horizontal governance models based on the inclusion of the community served in the decision-making process. This is crucial if the AI initiative deployed, along with its operators, are to be trusted by the population concerned.

7. *Take into account* the opinions and needs expressed by the citizen steering committee, even when they do not represent the objectives initially pursued. Indeed, the principles of self-determination, autonomy and effective power in the decision-making process are decisive in promoting the use of the AI system by members of the population concerned. This approach ensures not only that the AI deployed is appropriate to the socio-cultural context in question, but also that the horizontal governance structure model is respected.

8. *Monitor* the impact of the AI tools' deployment by the citizen steering committee and make any necessary readjustments at any stage of the implementation. This process should never be one-sided or one-dimensional. The implementation of an AI system within a community must remain a flexible and continuous process, where feedback is possible at every stage.

9. *Give* community representatives or structures the means to take ownership of the AI technology deployed in order to ensure the long-term sustainability of the project through local anchoring. This could be through open technology or technical training, for example. The issues of autonomy and emancipation of under-represented and/or marginalized communities will ensure their effective inclusion in the implementation of the AI tools intended for them.

10. *Re-evaluate* the participation tool and procedure if it is envisaged to deploy the same AI system to another group that would be marginalized or under-represented.

Annex 1: General incidences of AI on access to justice

The research also enabled us to identify 6 more general and potentially negative impacts of AI on access to justice. We include them in this paper as they are important to consider when developing AI for access to justice, and all the more crucial when seeking to deploy AI projects, initiatives or tools in marginalized and/or under-represented populations and communities.

- Impact 1: Discrimination due to structural bias

The very existence of AI technologies is based on the assimilation of data into the algorithms that compose them. However, this data is far from neutral and is based on an already biased social fabric, as well as on inequitable power relationships between the various actors in society. AI systems can therefore reproduce or even amplify these biases and reinforce existing discrimination. This could prove problematic or even dangerous if such systems were to be deployed in communities that are already under-represented or marginalized.

- Impact 2: Standardization of socio-cultural values

The deployment and use of AI systems composed of similar features is homogeneous, despite differences in context and culture. However, the application and use of these technologies in certain social-cultural contexts can prove problematic, as the values, ideas or priorities are not always the same as those in which the AI was conceived and created. It is therefore important to develop AI systems that are adapted to the situation of the populations concerned and avoid falling into what some authors describe as "colonization by AI".

- Impact 3: Data challenges

In some cases, citizens do not consider their personal data to be private, especially when they are not deemed sensitive. And yet, even if these data are not necessarily sensitive, they have an impact not only on the fundamental right to privacy, but more broadly on the way in which these data will be used to represent or even define an individual, and then indirectly, groups of individuals. Contrary to popular belief, these data are not always neutral, and the impact they can have on marginalized communities is not negligible.

- Impact 4: Automated decision-making

The great trust generally placed in new technologies, particularly AI, and the perception that these systems are objective and infallible, seem to mean that the results or outputs of these AI systems are too little monitored, verified, evaluated, or adjusted, with the result that certain problems can unfortunately go unnoticed. The negative repercussions on individuals (particularly under-represented populations or minorities) can be significant and can contravene human rights or promote discrimination and bias. All decision-making and the responsibilities that go with it must never be delegated to a single AI system, as a human being must be able to take ownership of it and maintain oversight of its operation.

- Impact 5: irrelevance of AI technology

The positive portrayal of AI is sometimes over-emphasized. However, it is essential to think critically about the appropriateness of using AI technology, its potential benefits, or the appropriateness of using it in a population or population group, in relation to the objective to be achieved. This is even more important when we consider the potentially negative impact that AI can have not only on the fundamental rights of individuals, but also on the environment. It is important to remember that the tool is a means to an end, not an end in itself: is there not a social, economic, or political response to be applied before a technological solution?

- Impact 6: Lack of clarity and direct legal recourse

It is regularly repeated that the lack of clarity about the data embedded in AI systems, their opacity, complexity and learning capabilities (particularly in the case of deep-learning AI) mean that the usual mechanisms for awarding liability can sometimes be difficult to apply in the event of harm caused by an AI technology. This has a potentially negative double impact: how to preserve and facilitate redress mechanisms in the event of damage caused by an AI system, and how to empower populations, particularly the most marginalized and under-represented, to obtain effective redress. This has wider implications for the maintenance of effective social justice.