



Article

# Embedding Ethical Principles into AI Predictive Tools for Migration Management in Humanitarian Action

Andrea Guillén \* and Emma Teodoro

Institute of Law and Technology, Faculty of Law, Autonomous University of Barcelona, 08193 Bellaterra, Spain

\* Correspondence: andrea.guillen@uab.cat

**Abstract:** AI predictive tools for migration management in the humanitarian field can significantly aid humanitarian actors in augmenting their decision-making capabilities and improving the lives and well-being of migrants. However, the use of AI predictive tools for migration management also poses several risks. Making humanitarian responses more effective using AI predictive tools cannot come at the expense of jeopardizing migrants' rights, needs, and interests. Against this backdrop, embedding AI ethical principles into AI predictive tools for migration management becomes paramount. AI ethical principles must be imbued in the design, development, and deployment stages of these AI predictive tools to mitigate risks. Current guidelines to apply AI ethical frameworks contain high-level ethical principles which are not sufficiently specified for achievement. For AI ethical principles to have real impact, they must be translated into low-level technical and organizational measures to be adopted by those designing and developing AI tools. The context-specificity of AI tools implies that different contexts raise different ethical challenges to be considered. Therefore, the problem of how to operationalize AI ethical principles in AI predictive tools for migration management in the humanitarian field remains unresolved. To this end, eight ethical requirements are presented, with their corresponding safeguards to be implemented at the design and development stages of AI predictive tools for humanitarian action, with the aim of operationalizing AI ethical principles and mitigating the inherent risks.

**Keywords:** AI predictive tools; migration management; humanitarian action; AI ethics



**Citation:** Guillén, Andrea, and Emma Teodoro. 2023. Embedding Ethical Principles into AI Predictive Tools for Migration Management in Humanitarian Action. *Social Sciences* 12: 53. <https://doi.org/10.3390/socsci12020053>

Academic Editor: Leah C. Schmalzbauer

Received: 30 April 2022

Revised: 9 October 2022

Accepted: 12 January 2023

Published: 18 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

According to the Global Humanitarian Overview 2022, 274 million people will need humanitarian assistance in 2022 (UN Office for the Coordination of Humanitarian Affairs 2021a). Managing humanitarian assistance in international migration demands the adoption of anticipatory and coordinated long-term actions, the improvement of data analysis, and the prioritization of people-centered responses by humanitarian actors (UN Office for the Coordination of Humanitarian Affairs 2021a). This must be conducted in full adherence to the humanitarian principles of humanity, neutrality, impartiality, and operational independence, which provide the foundations for humanitarian action (UN Office for the Coordination of Humanitarian Affairs 2012) and the humanitarian imperative of “do no harm.”

The development of AI predictive tools for humanitarian action can substantially help humanitarian actors to increase their decision-making capabilities based on anticipation, rather than reaction, to ensure effective migration governance (Beduschi 2021). However, the opportunity to make humanitarian responses more effective using AI predictive tools necessarily requires imbuing AI ethical principles in the design, development, and deployment stages of such tools. In the humanitarian context, embedding AI ethical principles into AI predictive tools for migration management preserves migrants' fundamental rights, while also seeking to improve their lives and well-being.

Studies on the current landscape of AI tools for migration management illustrate the ethical challenges raised by the use of AI tools in humanitarian contexts, such as accuracy, non-discrimination, transparency, and accountability (Berditchevskaia et al. 2021; Blasi Casagran et al. 2021; Pizzi et al. 2021), which reveal the need to put AI ethical principles into practice. This requires scrutinizing how the AI ethical principles of human autonomy, prevention of harm, fairness, and explicability, formulated by the High-Level Expert Group on Artificial Intelligence of the European Commission, must be embedded into AI predictive tools in the specific context of humanitarian action for migration management purposes.

However, rather than striving for broad ethical values, a pragmatic exercise must be conducted to translate these principles into actionable measures that can be implemented for their operationalization, consequently minimizing the potential risks at stake.

This paper is a contribution to the operationalization of AI ethical principles in AI predictive tools for migration purposes in the humanitarian field. It builds upon AI ethical principles and their corresponding requirements and refines them to the specific context of migration management for humanitarian purposes. As such, it accounts for the specificities stemming from the deployment of AI predictive tools in the humanitarian sector and the ethical risks that an AI tool may pose to migrants' fundamental rights. Therefore, AI predictive tools used for migration management purposes other than for humanitarian aid, such as border control, fall outside of the scope of this paper. The paper first identifies the AI ethical principles that must be embedded into any AI predictive tool deployed for migration management purposes in humanitarian action. AI ethical principles are then translated into practical requirements, whose content is tailored to the nature and specificities of using AI predictive tools in the humanitarian context and narrowly articulated into technical and organizational measures. This can be used as a practical guide for technical teams on how to design and develop ethical AI predictive tools for migration management in humanitarian action (Morley et al. 2020), and it can also serve those conducting impact assessments on these technologies to identify and evaluate risks and to implement appropriate safeguards to mitigate them.

This paper is based on the work conducted as part of the EU H2020-funded project ITFLOWS. It reflects the methodology followed for identifying, assessing, and mitigating risks posed by the design, development, and deployment of the ITFLOWS AI predictive tool, the EUMigraTool. This AI predictive tool is designed for first-line practitioners, second-level reception organizations, and municipalities and is meant to improve humanitarian assistance when managing EU migration flows in the phases of reception, relocation, settlement, and integration.

The structure of the paper is as follows: Section 2 outlines the risks highlighted by migration scholars and civil society actors regarding the use of data-driven technologies in the field of migration. Section 3 explores the ethical frameworks to be observed when designing and developing AI tools and discusses why current ethical frameworks are not specific enough to have a real impact, thus requiring operationalizing AI ethical principles. Section 4 presents the proposed procedure for the operationalization of AI ethical principles, and Section 5 examines the eight AI ethical requirements, along with actionable technical and organizational measures that enable the practical implementation of AI ethical principles into AI predictive tools for migration management in the humanitarian sector. Finally, in Section 6 conclusions are provided.

## 2. The Use of AI Predictive Tools in Migration Management: Risks and Opportunities

Regarding the use of AI tools to control migratory movements and manage border spaces, scholars and civil society actors have voiced concerns about the increasing reliance on AI driven technologies to solve the complex problem of migration governance (Molnar 2019b; Bircan and Korkmaz 2021; Nalbandian 2022).

Concerning the deployment and use of AI tools for humanitarian purposes, Beduschi (2022) stresses the risks that may arise from (i) "the dangers of surveillance humanitarianism" such as data collection practices by humanitarian actors without implementing

appropriate safeguards; (ii) “the excesses of techno-solutionism,” which refers to the use of digital technologies as a solution for complex societal problems; and, (iii) “the problems related to techno-colonialism,” meaning the practices related to digital innovation that can perpetuate colonial relationships of dependency and inequality among different populations. These risks can have a severe impact on the rights of such vulnerable populations and expose them to additional harms.

From a human rights perspective, scholars and human rights advocates have highlighted the risks that the deployment and use of AI data-driven technologies by states, international organizations, and humanitarian actors involved in migration management may entail in terms of jeopardizing human rights (Molnar 2019b; Bircan and Korkmaz 2021; Beduschi 2021). Human rights of migrants, refugees, and asylum seekers, such as the right to life, liberty, equality and non-discrimination, and privacy and data protection, can be seriously impacted by the deployment and use of these tools in the domain of migration management (Molnar 2019b).

This article focuses on AI predictive tools for migration management in humanitarian contexts. The rationale behind this choice is that the debate is currently centered at the security domain, in particular, border control. At the same time, there is a growing deployment of data-driven technologies in the humanitarian sector, which also poses risks to be addressed. These may include: (i) a disconnect between the design and deployment stages that leads to a lack of contextual knowledge; (ii) a lack of end-user expertise and last-mile implementation challenges; (iii) the loss of privacy and control over the use of data; (iv) a lack of high-quality, representative, machine-readable data; (v) inequalities, discrimination, and bias; (vi) the undermining of trust in the outcomes generated by the AI tool due to proprietary models; (vii) a lack of transparency and explainability; and (viii) unclear accountability mechanisms, among others (Berditchevskaia et al. 2021; Pizzi et al. 2021).

The use of AI predictive tools in the humanitarian domain carries risks which must not be underestimated, and which should be properly addressed and mitigated, but it also brings significant opportunities for humanitarian actors by supporting them in providing effective people-centered and context-specific responses. In this regard, AI predictive tools for migration management in humanitarian actions can significantly help humanitarian actors in preparedness, as well as response and recovery situations, to augment their decision-making capabilities and improve the lives and well-being of migrants (Beduschi 2022).

Reaping the benefits of these technologies requires the implementation of safeguards that minimize the risks they may pose. To this end, the design, development, and deployment of AI predictive tools for migration management must be governed by the humanitarian imperative of “do no harm” and the human rights international framework. This can be achieved through the operationalization of current human-rights based AI ethical frameworks, as discussed in subsequent sections.

### 3. AI Ethical Frameworks: The Challenges of Highly Abstract Ethical Principles

In the past decade, several AI ethical frameworks have been developed by different actors, such as governments, companies, professional associations, and non-profit organizations (Future of Life Institute 2017), or multi-stakeholder initiatives to guide the ethical, right-respecting, and socially desirable development and use of AI technologies (Fjeld et al. 2020). All of these aim at mapping AI ethical principles from a rights-based approach. However, they differ in their purposes due to the differing nature of the actors behind them. For instance, the ethical frameworks of government and inter-governmental organizations are designed to support their governance strategies regarding AI; civil society’s and multi-stakeholders’ guidelines try to lay AI foundations for the design, development, and use of AI technologies, while the private sector seeks to govern its internal AI developments and uses.

From a technical perspective, the most significant AI ethical guidelines are the Ethics Guidelines for Trustworthy Artificial Intelligence of the High-Level Expert Group on

Artificial Intelligence of the European Commission (AI HLEG) ([High-Level Expert Group on Artificial Intelligence 2019](#)), and the report of the IEEE on Ethically Aligned Design ([The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems 2019](#)).

These technical AI ethical frameworks developed by the AI HLEG and by the IEEE present AI ethical principles to be embedded into any AI tool. As highlighted by the AI HLEG, these principles are grounded in the protection of individuals' fundamental rights as enshrined in EU Treaties, the Charter of Fundamental Rights, and International Human Rights Law ([High-Level Expert Group on Artificial Intelligence 2019](#)). Therefore, these ethical frameworks serve as general guidelines for the design, development, and deployment of AI tools. However, as acknowledged by the AI HLEG, due to the context-specificity of AI systems, different contexts present different ethical challenges ([High-Level Expert Group on Artificial Intelligence 2019](#)).

To fill this gap, specific ethical frameworks for the design, development, and deployment of AI tools for humanitarian purposes have also been developed. For instance, the UN Office for the Coordination of Humanitarian Affairs ([UN Office for the Coordination of Humanitarian Affairs 2021b](#)), Nesta ([Berditchevskaia et al. 2021](#)), and the Humanitarian Data Science and Ethics Group ([Humanitarian Data Science and Ethics Group 2020](#)) have developed AI ethical guidelines for humanitarian action.

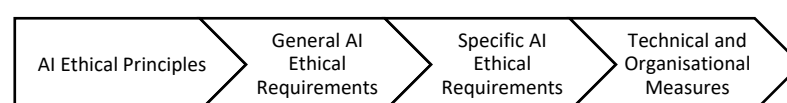
While these efforts aimed at contextualizing AI ethical principles in the humanitarian sector are welcome, such specific ethical frameworks will have little real impact unless AI ethical principles are effectively operationalized in the design, development, and deployment of AI tools ([Pizzi et al. 2021](#)). Putting AI ethical principles into practice requires their translation into low-level requirements with specific technical and organizational measures that consider the context of the application of the AI tool and its purpose, as well as the features of the technology ([Mittelstadt 2019](#)). Therefore, rather than aiming for high-level ethical principles, a pragmatic exercise must be conducted to translate AI ethical principles into actionable measures that can be implemented to operationalize them, and consequently, to minimize the potential legal and ethical risks at stake. However, the problem of how to operationalize AI ethical principles in AI predictive tools for migration management in the humanitarian sector remains unresolved.

Several methodologies for the operationalization of ethical principles have been proposed. For instance, the VCIO (values, criteria, indicators, and observables) model aims at making AI ethical principles practicable, comparable, and measurable ([Fetic et al. 2020](#)). The VCIO approach is comprised of four levels: (i) identifying the ethical values at stake given the context in which the AI system will operate; (ii) specifying the criteria that defines the fulfilment or violation of the corresponding value; (iii) establishing indicators to monitor whether a criterion is met; and (iv) defining observables that quantify or qualify if and to what extent indicators are met. In a similar vein, [Noriega et al. \(2022\)](#) propose eleven heuristics as part of the process of making ethical values operational for online institutions.

Building on the AI ethical frameworks and the methodologies for the operationalization of ethical principles, the next section presents the steps to be taken for the operationalization of AI ethical principles in the specific context in which the AI tool operates.

#### 4. From AI Ethical Principles to Actionable Measures

The steps proposed in this paper to translate AI ethical principles into practical guidance on how to operationalize them is depicted in [Figure 1](#).



**Figure 1.** Steps for operationalizing AI ethical principles.

The first step is the identification of the AI ethical principles to be embedded into AI predictive tools for migration management. The proposed AI ethical framework mainly

relies on the work conducted by the High-Level Expert Group on Artificial Intelligence of the European Commission (AI HLEG), which is grounded in the protection of individuals' fundamental rights, and also on the guidelines provided by the IEEE on ethically aligned design.

The four AI ethical principles to be imbued are: (i) human autonomy; (ii) prevention of harms; (iii) fairness; and (iv) transparency/explicability. According to the definitions provided by the AI HLEG, the AI ethical principles can be succinctly described as follows. Human autonomy entails that AI systems should not be designed to subordinate, coerce, deceive, manipulate, condition, or herd humans, but to augment, complement, and empower humans. The principle of prevention of harms means that AI systems should neither cause harm to individuals or groups, nor exacerbate it, nor have detrimental effects for human beings. Fairness means that the design, development, and deployment of AI systems must be fair in the sense of avoiding unfair bias, discrimination, and stigmatization, while also granting the opportunity to contest and seek effective redress against decisions made by AI systems. Lastly, transparency/explicability requires the transparency of processes, communicating the AI system's capabilities and purpose, and making the outcomes of the AI system explainable.

Secondly, the AI HLEG turns these principles into seven requirements for assessing risks. These requirements are: (i) human agency and oversight; (ii) technical robustness and safety; (iii) privacy and data governance; (iv) transparency; (v) diversity, non-discrimination, and fairness; (vi) environmental and societal well-being; and (vii) accountability. The AI ethical principles and their respective ethical requirements are listed in Table 1.

**Table 1.** AI ethical principles and the respective general AI ethical requirements.

AI Ethical Principles	General AI Ethical Requirements
Human autonomy	Requirement 1: Human agency and oversight
Prevention of harms	Requirement 2: Technical robustness and safety Requirement 3: Privacy and data governance Requirement 6: Societal and environmental well-being
Fairness	Requirement 5: Diversity, non-discrimination and fairness Requirement 6: Societal and environmental well-being Requirement 7: Accountability
Transparency/explicability	Requirement 4: Transparency

Similarly, the IEEE identifies the following eight principles for ethically aligned design: (i) human rights; (ii) well-being; (iii) data agency; (iv) effectiveness; (v) transparency; (vi) accountability; (vii) awareness of misuse; and (viii) competence.

Based on these approaches, the general AI ethical requirements presented by the AI HLEG and the requirement of awareness of misuse identified by the IEEE are assessed in this paper (Section 5). The awareness of misuse requirement must be included in the operationalization of AI ethical principles due to the potential ethical concerns that the misuse of a predictive tool for migration management may cause, e.g., the use of the AI tool for law enforcement purposes. As such, we propose to include this requirement as part of the AI ethical principle of prevention of harms.

The third step is to fine-tune these general ethical requirements to enable their operationalization by taking into account the purpose and context for which the AI predictive tool is deployed. As a result, specific AI ethical requirements are provided for each ethical requirement to facilitate the identification of technical and organizational measures required to make AI ethical principles actionable.

Lastly, technical and organizational measures are the actionable mechanisms to put AI principles into practice. Additionally, they serve as observables, as they can help in

determining whether an ethical requirement is being properly embedded into the design, development, and deployment stages of the AI tool.

The specific AI ethical requirements for AI predictive design for migration management in humanitarian action and the respective safeguards are examined in the following Section.

### 5. Specific Ethical Requirements and Safeguards to Embed into AI Predictive Tools for Migration Management

AI predictive tools for migration management must respect and promote migrants' human rights. Operationalizing AI ethical principles in AI predictive tools for migration management requires accounting for the specificities stemming from their application in the humanitarian field and the risks that AI tools may pose to migrants ([Humanitarian Data Science and Ethics Group 2020](#)).

This section delves into the eight ethical requirements that must be evaluated when designing and developing AI predictive tools for migration management in the humanitarian sector. Following the methodology provided in [Figure 1](#) for the operationalization of AI ethical principles, the ethical requirements have been fine-tuned to adapt them to the purpose and context for which the AI predictive tool is deployed. As a result, several specific AI ethical requirements are provided under each of the eight ethical requirements to facilitate the identification of the safeguards required to operationalize them.

What follows below can serve as a practical guide for technical teams on how to design and develop ethical AI predictive tools for migration management ([Morley et al. 2020](#)). These eight ethical requirements and their respective organizational and technical measures, if implemented at the design and development stages, can have a real impact ([Pizzi et al. 2021](#); [Morley et al. 2020](#)), significantly helping to mitigate risks. At the same time, it can aid those assessing the risks of the deployment of an AI predictive tool in migration management for humanitarian purposes. In this regard, the implementation (or lack thereof) of technical and organizational measures can help in determining whether an ethical principle is being properly imbued in the design, development, and deployment stages. For instance, if this evaluation is conducted via an impact assessment, the questions contained in the impact assessment questionnaire should aim at determining whether and to what extent such safeguards are implemented. This would enable the identification of risks posed by the AI predictive tool and the provision of mitigation measures to address them.

The eight general ethical requirements listed in [Section 3](#) are presented below, with their corresponding specific ethical requirements. Technical and organizational measures are also included as examples of how to achieve the specific ethical requirements, thereby operationalizing the respective AI ethical principle.

**Requirement 1: Human agency and oversight.** AI predictive tools should be designed and deployed to respect and promote human rights. This is paramount, given the array of human rights that can be at stake, as shown in [Section 2](#), and the common portrayal of migrants as security threats instead of human beings with rights ([Bircan and Korkmaz 2021](#)). Among other human rights, AI predictive tools must ensure human dignity and autonomy, which can be achieved through human agency and oversight. Human agency implies that AI predictive tools should aid humanitarian actors in making better and more informed choices when managing migration flows, while human oversight helps to prevent or minimize the potential risks of AI predictive tools ([European Commission, Directorate-General for Migration and Home Affairs 2021](#)). Human oversight and control mechanisms must be put in place in AI predictive tools deployed in humanitarian contexts to reduce inaccurate predictions of migration flows ([Humanitarian Data Science and Ethics Group 2020](#)).

Specific ethical requirements include embedding user-centric design principles that meet the needs of humanitarian actors to ensure human oversight and control and to prevent end-user overconfidence in, or overreliance on, the AI predictive tool ([Bircan and](#)

[Korkmaz 2021](#)). Likewise, end-users must be clearly informed about the AI predictive tool's functionalities, capabilities, and limitations.

Human agency and oversight can only be achieved if end-users have the expertise, competence, and authority to exercise human control effectively ([Article 29 Data Protection Working Party 2018](#)). End-user expertise and competence can be improved by adopting a collaborative approach to involve end-users in the design and development of the AI predictive tool. This can ensure that user-centric design principles are embedded into the tool, thus fulfilling end-users' needs, and allowing end-users to learn about the tool. Close collaboration between technical teams, NGOs, migration scholars, legal and ethical experts, end-users, and other stakeholders is also key to engage in discussions that enable the identification of legal and ethical risks, as well as mitigation measures to address them ([Bircan and Korkmaz 2021](#)). In this regard, the involvement of migrants is highly encouraged ([Molnar 2019a](#)). This participatory approach can provide developers with the adequate contextual knowledge and political, societal, and cultural sensitivities they may lack ([Pizzi et al. 2021](#)). Raising awareness of some unique features of a migration route or a country of origin can also improve fairness, e.g., this knowledge can help developers to spot that some populations are underrepresented in a given dataset, thus ultimately leading to potential biases ([Pizzi et al. 2021](#)). Another organizational measure that enhances human agency and oversight is delivering training sessions for end-users that enable their understanding of the AI predictive tool's functionalities, capabilities, and limitations. This can also prevent end-user overconfidence in or overreliance on the tool, as well as automation bias ([Skitka et al. 2000](#)), among other issues. From a technical perspective, human agency is preserved by designing AI predictive tools as decision-support systems that aid humanitarian actors in their decision-making processes, so that end-users can make informed and improved decisions based on the outcomes provided by the tool jointly with any other information or evidence they may have. Lastly, embedding reporting mechanisms into the tool is highly encouraged to allow end-users to flag errors, potential biases, and system malfunctions, thereby enhancing the AI predictive tool and its outcomes.

**Requirement 2: Technical robustness and safety.** AI predictive tools must be robust, resilient, secure, safe, accurate, reliable, and reproducible. Technical robustness and resilience should be ensured to prevent intentional harm, such as malicious attacks, as well as unintentional harm. Therefore, the existence of potential security risks must be evaluated at the design, development, and deployment phases, and mitigation measures must be implemented in accordance with the severity and likelihood of the risks. AI predictive tools must also provide accurate results, and the consequences of inaccurate outcomes must be evaluated. A thorough evaluation is required when the results of the AI predictive tool can affect human lives ([High-Level Expert Group on Artificial Intelligence 2019](#)).

Given that the AI predictive tools in question are deployed in the humanitarian sector for migration management purposes, accuracy rates should be high because their outcomes can have a direct impact on vulnerable individuals and groups. Among other consequences, inaccurate predictions could lead to miscalculations in the provision of critical aid, inefficient allocation of resources, or poor decisions regarding migrant placement. From a security perspective, the integrity and resilience of the AI predictive tool against malicious attacks must be ensured, and a fallback plan must be designed and tested. Lastly, humanitarian actors need to trust the tool to effectively use it. Therefore, reliability is a key aspect to ensure that end-users employ the technology and that those affected by the outcomes of the AI predictive tool are accepting of it ([Fournier-Tombs 2021](#)).

Accuracy thresholds or benchmarks must be determined and technically implemented. Below such a threshold, predictions should not be displayed to end-users. Instead, end-users must be warned that a prediction could not be made due to a low accuracy rate. Additionally, accuracy rates should be properly communicated to end-users of the tool. In case of low or medium accuracy rates for a prediction, a warning must be issued to alert humanitarian actors to the poor accuracy of the prediction. Technical measures should also be implemented to measure the frequency of the tool's inaccurate predictions. Accuracy

must be monitored on an ongoing basis, and procedures to improve accuracy rates must be implemented and reported. Security measures to be put in place include: SSL certificates; authentication and authorization of user access; two-factor authentication; secured servers, firewalls, and regular updates of the software, among other measures. In addition, the AI predictive tool should also be regularly backed up off-site to ensure that in case of an attack or any other event, the system can be back online quickly, with minimal to no data loss.

**Requirement 3: Privacy and data governance.** The rights to privacy and to the protection of personal data must be preserved and promoted, given the potential risks that AI predictive tools pose to fundamental rights through the processing of personal data. In the humanitarian sector, this also entails risks related to consent in order to ensure the voluntary participation of vulnerable people, the adoption of robust anonymization procedures, and the effective implementation of mechanisms that allow individuals to exercise their data protection rights ([Humanitarian Data Science and Ethics Group 2020](#)).

The lack of high-quality data has been highlighted as a challenge in the deployment of AI tools in the humanitarian context ([Beduschi 2022](#); [Berditchevskaia et al. 2021](#); [Pizzi et al. 2021](#)). In this field, datasets containing information to forecast migration flows are heterogeneous ([European Commission, Directorate-General for Migration and Home Affairs 2021](#)) and the need for more reliable, timely, and comparable statistical data on migration flows has been pointed out by several authors ([Beduschi 2021](#); [Singleton 2016](#)). The lack of high-quality data increases the risks that AI predictive tools can produce biased and incorrect results, which can ultimately lead to rights-infringing outcomes ([Molnar 2019a](#)).

Therefore, overreliance on existing data sources must be prevented ([Beduschi 2021](#)). Technical measures to ensure the quality and integrity of the data used for the predictive tool must be implemented. Oversight mechanisms must be put in place to mitigate the risks of using biased, inaccurate, or compromised datasets to ensure data quality and integrity. This can be achieved through regular quality checks of the external data sources fed into the tool. Regular checks should also be aimed at assessing the type and scope of the data in the datasets—in particular, to determine whether they contain personal data. This is particularly important because even when these AI predictive tools are fed with publicly available datasets that contain aggregated data, privacy-preserving mechanisms must be established, since personal data can be inferred from aggregated data ([Wachter and Mittelstadt 2019](#)). Lastly, non-authorized access to datasets can be prevented through authorization and authentication mechanisms.

When processing personal data, compliance with the provisions laid down by the General Data Protection Regulation (GDPR) must be ensured. This includes relying on a lawful basis for personal data processing (Article 6), abiding by the principles of lawfulness, fairness, and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; and accountability (Article 5), and embedding the principles of data protection using design and data protection by default (Article 25), among other obligations.

The GDPR also requires conducting a data protection impact assessment (DPIA) before processing data when it “is likely to result in a high risk to the rights and freedoms of natural persons” (Article 35(1)). As [Beduschi \(2022\)](#) notes, even if there is no legal obligation to perform a DPIA, DPIAs are highly encouraged because they serve as a robust protection tool. Although DPIAs are conceived as *ex ante* mechanisms to detect potential data protection harms and to identify appropriate measures to address and mitigate them, their scope is not limited to data protection risks. Given that Article 35(1) GDPR refers to “the rights and freedoms,” the DPIA is not circumscribed to data protection, as it may also involve the assessment of other human rights, such as non-discrimination ([Article 29 Data Protection Working Party 2017](#)). As such, DPIA questionnaires could include questions that help determining whether and to what extent technical and organizational measures are implemented, thereby also enabling the identification of legal and ethical risks posed by the AI predictive tool and the provision of additional safeguards to mitigate such risks.



In addition, DPIAs should be understood as an ongoing process ([Article 29 Data Protection Working Party 2017](#)) which would allow for a regular assessment of the inputs, outputs and the logic of the system, as new data can be collected and processed, potentially leading to new or higher risks.

**Requirement 4: Transparency.** Transparency encompasses three elements: (i) traceability, (ii) explainability, and (iii) open communication about the limitations of the AI system. For systems to be transparent, traceability measures must be implemented. Traceability implies that the datasets and the technology underlying the AI predictive tool should be documented. Given that traceability allows for the identification of the reasons behind the tool's predictions, it enables explainability, which is the ability to explain the predictions made by the AI tool intelligibly. To this end, the rationale behind the tool's prediction should be understood and traced by end-users.

Transparency and explainability in decision making within the humanitarian sector become key to build trust, to improve coordinated efficient responses, and to provide evidence-based decision making for targeting and prioritizing humanitarian assistance ([Humanitarian Data Science and Ethics Group 2020](#)). Therefore, explainability features should be implemented to provide explanations on how the predictions are made to foster humanitarian actors' views that the AI predictive tool is reliable ([European Commission, Directorate-General for Migration and Home Affairs 2021](#)). Lastly, communication channels must be established to raise awareness of the capabilities and limitations of the AI predictive tool. Consequently, end-users must be informed in a clear and understandable manner about the tool's purpose, capabilities, and limitations ([European Commission, Directorate-General for Migration and Home Affairs 2021](#)).

Technical measures to ensure traceability should be implemented. This may include documenting all methods used for designing, developing, testing, and validating the AI predictive tool and its outcomes. Explainability and interpretability can be improved by performing analysis control over training and testing data used to develop the AI predictive tool. In this regard, end-users must have access to understandable information regarding the reasoning and criteria behind the AI predictive tool's outcomes. This information should be provided in clear and plain language, free from technical jargon, and should be visible and easily accessible. From an end-user's perspective, it is also crucial to implement a reporting mechanism that allows end-users to provide feedback on the performance of the AI predictive tool. Lastly, organizational measures include training sessions for end-users that enable their understanding of the functionalities, capabilities, and limitations of the AI predictive tool.

**Requirement 5: Diversity, non-discrimination, and fairness.** Diversity, fairness, and non-discrimination requirements must be put into practice by implementing monitoring and mitigation measures to tackle both the risks related to the validity and quality of the humanitarian data in terms of its representativeness, completeness, and inclusiveness; and algorithmic biases that influence the tool's predictions ([Humanitarian Data Science and Ethics Group 2020](#)). Systems developers should ensure that the tool uses accurate AI models and that the data is not biased against attributes such as nationality, race, gender, age, religion, and sexual orientation. Otherwise, the algorithm would perpetuate discriminatory trends ([Beduschi 2022](#); [European Commission, Directorate-General for Migration and Home Affairs 2021](#)). Fairness can also be at stake due to the common lack of transparency of AI tools ([Pasquale 2015](#)) and the risk of automation bias ([Skitka et al. 2000](#)). This produces significant power imbalances between decision makers and migrants, which must be addressed ([Beduschi 2021](#)). As highlighted by [Fournier-Tombs \(2021\)](#), building trust is one of the key challenges in ensuring the adoption of AI predictive tools by humanitarian actors and the acceptance of those that can be affected by their outcomes. This difficulty can only be overcome if transparency and fairness are enhanced.

Diversity, non-discrimination, and fairness can be achieved with oversight mechanisms that identify, examine, address, and test biases in the datasets and in the model at the design and development phases. Oversight mechanisms must also be implemented

to ensure that the datasets used are not incomplete, outdated, or otherwise inadequate. Engagement with stakeholders with diverse backgrounds in the design, development, and deployment phases must be sought to enhance diversity and fairness ([Access Now 2018](#)).

From a design perspective, the technology should be understandable and accessible to all end-users, regardless of their age, abilities, or characteristics. In this regard, the active collaboration of relevant stakeholders with diverse backgrounds and viewpoints at different stages of technological development is highly encouraged to avoid discrimination and the perpetuation of existing inequalities ([Molnar 2019b](#); [Bircan and Korkmaz 2021](#)). Such a participatory approach involving NGOs, other humanitarian actors, and migrant communities helps to embed fairness into the system. As mentioned above, close collaboration with end-users and other stakeholders can also equip developers with adequate contextual knowledge ([Pizzi et al. 2021](#)), which could sensitize them to potential inherent biases that could unconsciously be introduced into the AI tool. Lastly, a reporting mechanism would also enhance fairness, since end-users could flag errors and potentially biased results. End-users would be more likely to use the AI predictive tool if such a reporting mechanism were in place as a key element to build trust.

**Requirement 6: Societal and environmental well-being.** AI predictive tools for migration management should aim at benefitting migrants and society at large. Such predictive tools should be needs-based instead of technology-based ([Humanitarian Data Science and Ethics Group 2020](#)). They must be designed to strive for fairness and to prevent individual and societal harms, as well as with sustainability and environmental friendliness in mind.

Therefore, the social and ecological impact of the AI system should be regularly assessed. In this regard, risks must be identified and their likelihood or severity assessed to ultimately provide effective mitigation measures with the potential to mitigate or at least minimize such risks.

The societal impact of the AI predictive tool should be evaluated, both at the individual and societal level, through an integrated impact assessment that covers human rights, ethical, and societal aspects—the so-called human rights, ethical, and social impact assessment ([Mantelero 2018](#)). As part of this integrated impact assessment, the effective implementation of the AI ethical principles can be evaluated according to the level of compliance with the requirements and measures presented in this section. Crucially, impact assessments must be carried out before the AI tool is deployed to enable the identification of risks and their corresponding mitigation measures without affecting individuals' rights. After the deployment of the AI tool, regular impact assessments are encouraged to identify new risks, to keep track of the effectiveness of the mitigation measures already in place, and to provide additional measures, if needed.

Likewise, the ecological impact of the system should be evaluated throughout the system's lifecycle, and measures to reduce such impact should be encouraged. Additionally, AI tools should be user-centric and designed in a way that it is usable and accessible to all end-users, regardless of their personal characteristics.

**Requirement 7: Accountability.** Accountability requires the implementation of technical and organizational measures to report the tool's performance and to provide adequate and accessible remedies and redress. Such measures include the assessment of design processes, the AI predictive models, the datasets used, and the predictions, which allows for the auditability of the system. In this sense, auditability involves reporting the negative impacts of the system, identifying appropriate mitigation measures, and feeding them into the system. As mentioned above, these negative impacts can be identified and evaluated through comprehensive impact assessments that must be conducted regularly.

Technical measures must be embedded into the AI predictive tool to allow end-users to report potential vulnerabilities, risks, or biases. Other technical mechanisms that must be implemented include: (i) the authentication and authorization of components; (ii) the definition of users' roles and privileges; and, (iii) logging mechanisms to record when, where, how, by whom, and for what purpose the tool was used. Accountability also

involves providing explanations of the tool's predictions so that humanitarian subjects can have the opportunity to challenge the tool's outcomes and to seek redress through judicial or extra-judicial mechanisms.

To this end, legal responsibility and liability must be clearly established ([Humanitarian Data Science and Ethics Group 2020](#)). Lastly, the design and development of open-source AI predictive tools is highly encouraged, since the availability of the code source enables public scrutiny of the AI system, allowing any third-party to audit it.

**Requirement 8: Awareness of misuse.** The risk of misuse refers to potential uses of the tool for unintended purposes.

Particularly, in the humanitarian context, AI predictive tools can be misused for non-intended purposes, such as border control. Therefore, the potential misuse of the AI predictive tool must be considered at the design and development stages.

When designing and developing AI predictive tools for migration management, the risk of misuse must be anticipated, and technical measures must be put in place to minimize it. Technical measures include: (i) implementing authorization and authentication components; (ii) recording all logging activities; (iii) defining users' roles; and (iv) clearly stating its humanitarian purpose and its scope of application in the terms of use of the AI tool.

## 6. Conclusions

Humanitarian actors are increasingly adopting AI predictive tools for migration management. Migration scholars and civil society actors have voiced concerns about the multiple risks arising from the use of such tools. However, these tools also provide opportunities for humanitarian actors in different situations, including preparedness, response, and recovery ([Beduschi 2022](#)). AI predictive tools can improve the decision-making capabilities of human actors; however, to reap its benefits, the risks must be addressed and mitigated to the greatest extent possible. Thus, embedding AI ethical principles into AI predictive tools for migration management becomes paramount. Embedding AI ethical principles into such tools requires the effective implementation of safeguards that mitigate the risks.

Current AI ethical guidelines present high-level ethical principles that are not action-guided, thus curbing their practical implementation. This paper aims at addressing the question of how to imbue AI ethical principles into the design, development, and deployment of AI predictive tools for migration management in the humanitarian field, with the ultimate goal of minimizing the legal and ethical risks they pose. To this end, eight ethical requirements are presented, with their corresponding organizational and technical measures to be implemented at the design, development, and deployment stages of AI predictive tools for humanitarian action with the aim of operationalizing AI ethical principles and mitigating risks. These safeguards can be used as observables to determine whether and to what extent AI ethical principles are implemented. The implementation status of the principles enables, in turn, the identification of risks posed by the AI predictive tool. This can lead to the provision of additional safeguards that serve as mitigation measures to address these risks, operationalizing the respective AI ethical principle.

This paper is a first attempt to put AI ethical principles into practice in AI predictive tools for migration management in humanitarian action. It provides a starting point for the discussion on how the effective operationalization of AI ethical principles can be achieved. Embedding AI ethical principles in these tools is paramount, as the safeguards required to be implemented aim at reaping the benefits of this technology for humanitarian action while minimizing its risks.

**Author Contributions:** Conceptualization, A.G. and E.T.; methodology, A.G. and E.T.; investigation, A.G. and E.T.; writing—original draft preparation, A.G. and E.T.; writing—review and editing, A.G. and E.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the ITFLOWS project as part of the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 882986.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank the anonymous reviewers for the insightful comments and suggestions, which helped us to improve the quality of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Access Now. 2018. Human Rights in the Age of Artificial Intelligence. Available online: <https://www.accessnow.org/human-rights-matter-in-the-ai-debate-lets-make-sure-ai-does-us-more-good-than-harm/> (accessed on 3 March 2022).
- Article 29 Data Protection Working Party. 2017. Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is ‘Likely to Result in a High Risk’ for the Purposes of Regulation 2016/679. (wp248rev.01). Available online: <https://ec.europa.eu/newsroom/article29/items/611236/en> (accessed on 3 March 2022).
- Article 29 Data Protection Working Party. 2018. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. (WP251rev.01). Available online: <https://ec.europa.eu/newsroom/article29/items/612053> (accessed on 3 March 2022).
- Beduschi, Ana. 2021. International Migration Management in the Age of Artificial Intelligence. *Migration Studies* 9: 576–96. [CrossRef]
- Beduschi, Ana. 2022. Harnessing the Potential of Artificial Intelligence for Humanitarian Action: Opportunities and Risks. *International Review of the Red Cross* 104: 1149–69. [CrossRef]
- Berditchevskaia, Aleks, Eirini Malliaraki, and Kathy Peach. 2021. *Participatory AI for Humanitarian Innovation: A Briefing Paper*. London: Nesta.
- Bircan, Tuba, and Emre Eren Korkmaz. 2021. Big Data for Whose Sake? Governing Migration through Artificial Intelligence. *Humanities & Social Sciences Communications* 8: 1–5. [CrossRef]
- Blasi Casagran, Cristina, Colleen Boland, Elena Sánchez-Montijano, and Eva Vilà Sanchez. 2021. The Role of Emerging Predictive IT Tools in Effective Migration Governance. *Politics and Governance* 9: 133–45. [CrossRef]
- European Commission, Directorate-General for Migration and Home Affairs. 2021. *Feasibility Study on a Forecasting and Early Warning Tool for Migration Based on Artificial Intelligence Technology*. Brussels: Publications Office.
- Fetic, Lajla, Torsten Fleischer, Paul Grünke, Thilo Hagendorf, Sebastian Hallensleben, Marc Hauer, Michael Herrmann, Rafaela Hillerbrand, Carla Hustedt, Christoph Hubig, and et al. 2020. *From Principles to Practice: An Interdisciplinary Framework to Operationalise AI Ethics*. Gütersloh: Bertelsmann Stiftung. [CrossRef]
- Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Christopher Nagy, and Madhulika Srikumar. 2020. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Cambridge, MA: Berkman Klein Center for Internet & Society at Harvard University.
- Fournier-Tombs, Eleonore. 2021. Towards a United Nations Internal Regulation for Artificial Intelligence. *Big Data & Society* 8: 205395172111039493. [CrossRef]
- Future of Life Institute. 2017. Asilomar AI Principles. Available online: <https://futureoflife.org/open-letter/ai-principles/> (accessed on 1 March 2022).
- High-Level Expert Group on Artificial Intelligence. 2019. *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission.
- Humanitarian Data Science and Ethics Group. 2020. *A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector*. Available online: <https://www.hum-dseg.org/dseg-ethical-framework> (accessed on 1 March 2022).
- Mantelero, Alessandro. 2018. AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. *Computer Law & Security Review* 34: 754–72. [CrossRef]
- Mittelstadt, Brent. 2019. Principles Alone Cannot Guarantee Ethical AI. *Nature Machine Intelligence* 1: 501–7. [CrossRef]
- Molnar, Petra. 2019a. New Technologies in Migration: Human Rights Impacts. *Forced Migration Review* 61: 7–9.
- Molnar, Petra. 2019b. Technology on the Margins: AI and Global Migration Management from a Human Rights Perspective. *Cambridge International Law Journal* 8: 305–30. [CrossRef]
- Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics* 26: 2141–68. [CrossRef] [PubMed]
- Nalbandian, Lucia. 2022. An Eye for an ‘I’: A Critical Assessment of Artificial Intelligence Tools in Migration and Asylum Management. *Comparative Migration Studies* 10: 1–23. [CrossRef] [PubMed]
- Noriega, Pablo, Harko Verhagen, Julian Padget, and Mark d’Inverno. 2022. Design Heuristics for Ethical Online Institutions. In *Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XV. COINE 2022. Lecture Notes in Computer Science*. Edited by N. Ajmeri, A. Morris Martin and B. T. R. Savarimuthu. Cham: Springer, vol. 13549, pp. 213–30. [CrossRef]
- Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

- Pizzi, Michael, Mila Romanoff, and Tim Engelhardt. 2021. AI for Humanitarian Action: Human Rights and Ethics. *International Review of the Red Cross* 102: 145–80. [\[CrossRef\]](#)
- Singleton, Ann. 2016. *Migration and Asylum Data for Policy-Making in the European Union—The Problem with Numbers*. Brussels: CEPS Paper in Liberty and Security in Europe No. 89.
- Skitka, Linda, Kathleen Mosier, and Mark D. Burdick. 2000. Accountability and Automation Bias. *International Journal of Human-Computer Studies* 52: 701. [\[CrossRef\]](#)
- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2019. *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*. Piscataway: IEEE.
- UN Office for the Coordination of Humanitarian Affairs. 2012. *What Are Humanitarian Principles?* New York: OCHA.
- UN Office for the Coordination of Humanitarian Affairs. 2021a. *From Digital Promise to Frontline Practice: New and Emerging Technologies in Humanitarian Action*. New York: OCHA.
- UN Office for the Coordination of Humanitarian Affairs. 2021b. *Global Humanitarian Overview 2022*. New York: OCHA.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.