Towards Responsible and Personalized Social Protection: Leveraging Machine Learning for Policy Innovation in Developing Countries – The Case of Morocco

Chaymae Sahraoui, ZARI Tarek
Faculty of Legal, Economic and Social Sciences – Ain Sebaa
University of Hassan II, Casablanca, Morocco
Email: sahraouic90@gmail.com/chaymae.sahraoui-etu@etu.univh2c.ma, profzari@gmail.com

***Abstract*—This article explores the potential of Machine Learning (ML) to enhance the effectiveness, fairness, and personalization of social protection systems in developing countries, with a particular focus on Morocco. In light of the demographic, institutional, and technological transformations underway, ML emerges as a promising tool to support eligibility assessment, fraud detection, beneficiary segmentation, and long-term forecasting.**

The paper proposes a structured five-step methodological framework for the ethical and context-sensitive integration of ML in welfare systems. Drawing from international experiences (Estonia, India, Rwanda, Chile), it identifies key success factors, including data interoperability, human oversight, and transparency. The article then outlines five strategic use cases tailored to Morocco's current reform of universal social protection, ranging from predictive targeting to territorial outreach.

To assess the social acceptability of these innovations, an exploratory survey was conducted among Moroccan citizens and stakeholders. Results indicate a cautiously optimistic attitude toward ML adoption in public services, tempered by concerns around algorithmic bias and institutional trust. Based on these findings, the paper offers concrete policy recommendations to foster inclusive, transparent, and human-centered AI governance in the social domain.

By combining theoretical insight, comparative evidence, and citizen perceptions, this study contributes to the global "AI for Social Good" agenda and informs the digital transformation of social policy in the Global South.

***Index Terms*—Machine Learning, Social Protection, Algorithmic Fairness, Public Policy Innovation, AI Governance, Morocco, Welfare Targeting, Developing Countries**

I. Introduction

Designing inclusive and adaptive social protection systems has become a central policy challenge for low- and middle-income countries (LMICs) in the 21st century. These systems face mounting structural pressures, including demographic shifts, persistent labor market informality, and recurrent socio-economic shocks, all of which undermine the efficacy of models historically built around formal employment and urban welfare institutions [1].

Traditional policy instruments often struggle to capture heterogeneous life paths and to anticipate emerging vulnerabilities, particularly in fragmented and rapidly evolving contexts. As a result, large segments of the population—informal workers, rural families, and socially excluded groups—remain underserved by conventional protection mechanisms [2], [3].

Amid these constraints, digital transformation presents a unique opportunity to rethink social policy delivery. Advances in Artificial Intelligence (AI), and particularly in Machine Learning (ML), offer powerful capabilities to process complex data, forecast risks, and inform more granular and adaptive interventions. Yet, the integration of ML into public systems raises fundamental questions regarding data availability, institutional capacity, algorithmic fairness, and citizen trust [4].

The Moroccan case provides a timely and relevant lens through which to explore these dynamics. Since 2021, the country has embarked on an ambitious reform aimed at achieving universal social protection, extending coverage to more than 22 million citizens, including informal workers, self-employed individuals, and low-income households [5]. While administrative modernization and institutional reform have made significant progress, the use of advanced data-driven tools—particularly ML—remains at an early stage.

This paper contributes to the emerging field of "AI for Social Good" by examining how ML can support a more efficient, transparent, and personalized social protection system in Morocco. The analysis proceeds in five steps. Section II presents a critical review of recent academic and institutional literature on AI in public policy and welfare delivery. Section III outlines a five-step methodological framework for the responsible deployment of ML in social protection. Section IV reviews international case studies to distill lessons applicable to the Moroccan context. Section V proposes five concrete ML use cases aligned with Morocco's ongoing reform. Sections VI and VII discuss governance, ethical, and implementation challenges, complemented by an exploratory survey on public perceptions of AI. Finally, Section VIII offers policy recommendations and outlines directions for future research.

II. Literature Review

The integration of Artificial Intelligence (AI), and more specifically Machine Learning (ML), into social protection

systems has gained increasing attention from scholars, policymakers, and international organizations. This literature review synthesizes recent developments along four major dimensions: public sector innovation, algorithmic targeting, personalization of welfare, and ethical challenges in low- and middle-income contexts.

A. AI and Public Sector Innovation

AI is increasingly being used to enhance public administration through data driven decision making, automation of services, and predictive modeling. In the domain of social protection, ML has shown promise in optimizing eligibility assessments, detecting fraud, and allocating resources more efficiently [6]. These tools support the transition from reactive to proactive welfare models, enabling governments to anticipate needs rather than merely respond to crises.

B. Algorithmic Targeting and Its Trade-Offs

Several studies have demonstrated that machine learning–based targeting approaches can outperform traditional mechanisms such as proxy means testing or categorical eligibility rules. However, these algorithmic models raise significant concerns regarding transparency, fairness, and social justice. As noted by Mehrabi et al., biases embedded in training data or model architecture can result in discriminatory outcomes, particularly for historically marginalized groups [7]. Furthermore, Eubanks highlights that the deployment of automated systems in public welfare services has, in many cases, reinforced structural inequalities and intensified the surveillance and control of vulnerable populations [8]. These risks emphasize the need for explainability, ethical safeguards, and strong human oversight in the design and deployment of ML-driven social protection systems.

C. Personalized Welfare and Adaptive Interventions

Machine learning facilitates highly personalized welfare interventions. By combining non-traditional big data sources—such as mobile phone usage—with advanced ML models, practitioners can more accurately target vulnerable populations than traditional eligibility systems allow, as demonstrated in recent humanitarian studies [9]. This innovative targeting enhances program effectiveness and reduces exclusion errors. However, these methods also raise significant concerns regarding surveillance, data privacy, and the erosion of consent-based service delivery models.

D. Challenges in Low- and Middle-Income Countries

The application of ML in LMICs is constrained by fragmented data systems, limited digital infrastructure, and weak institutional interoperability [10], [11]. These challenges hinder the development of robust predictive models and raise questions about the transferability of ML tools designed in high-income settings. Scholars increasingly call for locally trained models that are adapted to national socio-economic conditions [12].

E. Towards a Responsible AI Agenda

To mitigate these risks, a growing body of work advocates for a responsible AI agenda in social protection [13], [14]. This includes principles such as transparency, non-discrimination, and participatory design. Human oversight remains a cornerstone of this agenda, with scholars emphasizing the importance of embedding "human-in-the-loop" mechanisms to maintain legitimacy and prevent harm [15], [16].

In summary, the literature reveals both the transformative potential and the complex risks of applying ML in welfare systems. Effective deployment requires not only technical capability but also ethical safeguards, institutional readiness, and contextual adaptation. These insights guide the framework and use cases developed in the following sections.

III. Methodological Framework

The integration of Machine Learning (ML) into social protection systems requires more than algorithmic sophistication. It necessitates a comprehensive, ethically grounded, and context-aware approach to ensure relevance, fairness, and sustainability. This section introduces a five-step framework designed to guide the responsible adoption of ML in welfare programs, balancing technical feasibility with institutional readiness and normative safeguards.

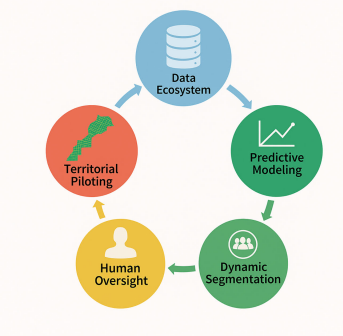

Fig. 1. Five-step framework for responsible ML integration in social protection

A. Step 1: Data Ecosystem and Interoperability

High-quality, interoperable data are the foundation of any ML application in social policy. This includes demographic registries, household surveys, tax and employment records, and, when ethically justified, geospatial or behavioral data (e.g., utility consumption or mobile usage) [17]. In fragmented systems, the creation of unified or federated social registries is a necessary precursor.

Data integration must follow strict governance standards: data minimization, explicit consent, transparency, and traceability. Aligning with international frameworks like the General Data Protection Regulation (GDPR) ensures legal robustness and social acceptability [18].

TABLE I
ILLUSTRATIVE DATA SOURCES FOR ML IN SOCIAL PROTECTION

Data Type	Potential Sources
Demographic	Civil registry, national census
Socioeconomic	Unified Social Registry (USR), household surveys
Administrative	Tax, labor, health, education, and pension databases
Geospatial	Satellite imagery, GIS platforms
Behavioral (optional)	Mobile records, utility bills, app metadata

B. Step 2: Predictive Modeling and Risk Scoring

Once data readiness is ensured, supervised learning algorithms can be employed to predict welfare-relevant outcomes. Techniques such as Random Forests, Gradient Boosting (e.g., XGBoost) [19], or logistic regression models can estimate the probability of poverty entry, dropout from health coverage, or benefit ineligibility.

These outputs must remain decision-support indicators—not automatic verdicts. Explainability tools such as SHAP (SHapley Additive exPlanations) [20] and LIME (Local Interpretable Model-agnostic Explanations) [21] are crucial for understanding model behavior and ensuring transparency.

C. Step 3: Dynamic Segmentation and Service Personalization

ML facilitates dynamic segmentation of beneficiaries into meaningful clusters—such as elderly rural dwellers, informal workers, or newly vulnerable households—based on both static and behavioral variables. This enables personalized services, differentiated contributions, or targeted outreach campaigns.

Clustering techniques such as K-Means, DBSCAN [22], or hierarchical methods are often used in this step. Segments should be regularly updated using temporal data to remain responsive to socio-economic changes.

D. Step 4: Human Oversight and Ethical Controls

Human judgment must remain central in any ML-supported decision system. Algorithms should inform, not replace, the professional expertise of social workers, case managers, and policy officers. This requires embedding "human-in-the-loop" mechanisms at all critical stages.

Governance structures should include audit trails, redress protocols, and independent ethics committees [13], [15]. These safeguards uphold principles of accountability, non-discrimination, and human dignity—especially when algorithms affect access to essential rights.

E. Step 5: Territorial Piloting and Institutional Integration

Before national deployment, ML solutions should undergo iterative piloting in select territories, chosen based on vulnerability prevalence and institutional maturity. Pilots serve both technical (model testing, error measurement) and social (participatory validation, trust-building) purposes.

Successful integration requires adapting workflows, training civil servants, and clarifying lines of responsibility. Legal harmonization, policy alignment, and multi-stakeholder engagement are essential to embed ML in existing administrative structures without undermining public legitimacy [23].

In sum, this five-step framework articulates a balanced approach to ML adoption in social protection: data-driven yet human-centered, technically sound yet ethically robust, and scalable while remaining locally grounded.

IV. INTERNATIONAL EXPERIENCES AND LESSONS LEARNED

Although the use of Machine Learning (ML) in social protection remains in its infancy across low- and middle-income countries (LMICs), several pioneering initiatives—both in the Global South and in more advanced economies—offer concrete examples of operationalization. This section presents a comparative overview of four relevant country cases and extracts key lessons for countries such as Morocco.

A. Applications of AI in Social Protection

Across public sectors, ML has been applied to diverse tasks such as eligibility verification, fraud detection, beneficiary profiling, and real-time monitoring of service delivery [23]. In the social protection domain, these innovations are increasingly framed as tools for improving targeting accuracy, operational efficiency, and responsiveness to shifting vulnerabilities [24].

However, early deployments also reveal substantial governance challenges—especially when algorithmic systems are implemented without adequate legal, ethical, or institutional safeguards [7].

B. Country Case Studies

Estonia. Estonia stands as a benchmark in digital government. Through a robust digital identity infrastructure and interoperable data platforms, its agencies deliver real-time eligibility assessments and personalized social services. ML has been used in early warning systems for child welfare and employment support. The success of e-Estonia is attributed to legal clarity, institutional coordination, and public trust [23].

India. The Aadhaar biometric identification system underpins many welfare schemes. ML algorithms are used to validate claims, detect duplicates, and segment beneficiaries. However, exclusion errors—especially for undocumented or marginalized populations—have raised serious ethical and legal concerns [25]. Litigation has led to restrictions on automatic enforcement of Aadhaar-based eligibility.

Rwanda. Rwanda has upgraded its Ubudehe community classification system using ML to improve targeting accuracy. Socioeconomic and geographic variables feed into classification models that inform benefit allocation. While promising, the initiative faces challenges in terms of model accuracy and transparency, due in part to fragmented data sources [11].

Chile. Chile's Registro Social de Hogares (RSH) integrates administrative data to calculate household vulnerability scores. ML supports the assignment of social benefits, with built-in appeal mechanisms and citizen feedback loops. The Chilean experience highlights the importance of accountability and iterative model refinement [26].

To synthesize these diverse country experiences, the table below presents a comparative overview of the main ML applications in social protection across four countries—Estonia, India,

Rwanda, and Chile—highlighting their respective strengths and implementation challenges.

Country	ML Application	Strengths	Challenges
Estonia	Real-time eligibility checks	Strong digital infrastructure	Limited replicability
India	Fraud detection (Aadhaar)	Scale and biometric verification	Exclusion risks, legal debates
Rwanda	Vulnerability classification	Local adaptation of data	Accuracy, data fragmentation
Chile	Household scoring	Data integration and appeal right	Transparency, algorithm opacity

Fig. 2. Comparative overview of ML applications in selected countries

C. Lessons for LMICs

A cross-case synthesis reveals five key lessons for developing countries aiming to adopt ML in social protection:

1) **Data readiness is critical.** Countries with unified digital IDs and interoperable registries are better positioned to implement ML tools effectively [27].
2) **Human oversight ensures legitimacy.** Fully automated decisions risk opacity and contestability. Hybrid models combining algorithmic scoring with social worker validation are more sustainable [13].
3) **Contextualization is essential.** Imported models often fail to capture local labor dynamics, informal structures, and cultural norms. Locally trained algorithms offer greater accuracy and legitimacy [12].
4) **Transparency fosters public trust.** Tools such as model cards, explainability dashboards, and grievance redress systems can enhance social acceptability [28].
5) **Pilots enable safe experimentation.** Territorial testing allows gradual scaling, contextual adaptation, and early detection of inclusion/exclusion risks [11].

These international experiences provide a conceptual and operational foundation to inform Morocco's ongoing social protection reform. The next section builds upon these insights to propose specific ML applications aligned with national priorities.

V. Potential Applications of ML in Morocco

Building on the international insights presented in the previous section, this part of the paper proposes concrete applications of Machine Learning (ML) in the context of Morocco's ongoing social protection reform.

A. Strategic Context

In 2021, Morocco launched an ambitious national initiative to achieve universal social protection. This multi-phased reform includes the extension of compulsory health coverage (AMO), the generalization of family allowances, the expansion of pension schemes, and the establishment of targeted social assistance mechanisms. These efforts primarily aim to include informal workers, low-income households, and other vulnerable groups previously excluded from contributory programs [5].

To deliver on these objectives, institutions must not only expand coverage but also optimize targeting, anticipate exclusion risks, and personalize service delivery. ML can provide scalable and adaptive tools aligned with these evolving priorities.

B. Use Case 1: Targeting for AMO Tadamoun

The AMO Tadamoun scheme offers subsidized health coverage for poor and vulnerable households not covered by formal employment. ML models trained on data from the Unified Social Registry (RSU) could identify households eligible but not yet affiliated.

Features such as household composition, income proxies, geographic isolation, and informal employment patterns could feed into supervised classification algorithms. Time-series models may further anticipate eligibility loss due to income fluctuations or demographic changes.

C. Use Case 2: Monitoring Contributions among Self-Employed (TNS)

Self-employed workers (TNS) are required to contribute to social insurance schemes, but dropout risks are high. ML could flag contributors at risk of disengagement by analyzing variables such as contribution history, income declarations, seasonality, and economic sector.

Segmentation algorithms can enable differentiated outreach strategies—e.g., tailored reminders, temporary exemptions, or administrative facilitation measures.

D. Use Case 3: Detecting Latent Eligibility in AMO Achamil

AMO Achamil is a contributory scheme targeting individuals without fixed employment contracts, such as caregivers, freelancers, or informal earners. Many potential beneficiaries remain invisible in administrative databases.

Unsupervised clustering techniques could detect latent eligibility by analyzing cross-sectoral data—such as utility records, health visits, and mobile usage patterns—while ensuring compliance with data privacy regulations. This would support proactive and voluntary affiliation strategies.

E. Use Case 4: Pension Coverage Forecasting

ML can help simulate long-term pension outcomes for workers with fragmented careers or low contribution densities. Regression models and survival analysis could estimate future entitlements and detect individuals at risk of receiving no or minimal pensions.

These insights could inform early interventions, such as enrollment in simplified schemes, contribution matching, or digital nudging toward formalization.

F. Use Case 5: Territorial Prioritization and Outreach

Geospatial ML techniques can support the deployment of mobile enrollment units and awareness campaigns in under-covered regions. Algorithms can identify clusters of vulnerable households not integrated into existing schemes based on location, infrastructure access, or public service usage.

Anomaly detection could also help spot unusual patterns—e.g., entire villages lacking affiliation—guiding institutional attention to service deserts or data blind spots.

G. Ethical and Operational Considerations

Each proposed use case involves specific ethical and technical challenges. It is essential to ensure fairness in model design, safeguard personal data, and avoid unintended exclusion effects.

As emphasized by the OECD and UN Digital Principles [14], [23], ML deployment should be guided by principles of explainability, participatory design, and human oversight. Algorithms must remain transparent and accountable, with mechanisms for appeal and redress embedded in every application.

Together, these use cases demonstrate how ML can enhance the precision, responsiveness, and inclusiveness of Morocco's social protection system—if deployed responsibly and aligned with institutional priorities.

VI. Discussion: Governance, Ethics, and Implementation Challenges

While the technical potential of Machine Learning (ML) in social protection is increasingly recognized, its operationalization in real-world policy environments—especially in developing countries—raises several critical challenges. These go beyond algorithmic performance and touch upon institutional capacity, ethical risks, legal ambiguity, and citizen trust.

A. Institutional Readiness and Capacity Gaps

Effective ML deployment requires more than data availability. Many public institutions in low- and middle-income countries face systemic limitations: fragmented databases, weak interoperability, and understaffed digital units [11]. Morocco, for example, has made notable progress through the Unified Social Registry (RSU) and CNSS reform, but full interoperability and real-time data exchange remain ongoing challenges.

Moreover, civil servants and local agents often lack the training needed to interpret algorithmic outputs or detect anomalies in automated recommendations. Without substantial investment in human capital and institutional infrastructure, ML tools risk being underutilized, misapplied, or abandoned.

B. Ethical Risks and Algorithmic Fairness

ML systems in welfare settings are particularly sensitive to bias and discrimination. If training data reflect historical exclusion—such as the underrepresentation of rural women or informal workers—algorithms may reproduce or even amplify these patterns [7]. In contexts where transparency is lacking, affected individuals may have no means to challenge or even understand automated decisions.

To mitigate these risks, ML systems must be designed with fairness constraints, audited regularly, and coupled with explainability tools (e.g., SHAP, LIME). Ethical oversight bodies—such as AI ethics committees—can also monitor impacts and advise on model updates [13], [15].

C. Public Trust and Social Acceptability

Public acceptance of ML in welfare depends on whether citizens perceive decisions as legitimate, fair, and contestable. Studies have shown that opaque systems reduce trust, particularly when benefits or entitlements are involved [29]. In Morocco, preliminary survey results (see Section VII) indicate cautious optimism but also concern about exclusion risks and institutional accountability.

Ensuring meaningful "human-in-the-loop" mechanisms—where social workers validate, override, or contextualize algorithmic outputs—is essential to preserve trust and respect human agency. Accessible grievance procedures and feedback loops further enhance acceptability.

D. Legal and Governance Frameworks

Most developing countries lack comprehensive legal frameworks to regulate the use of AI in the public sector. In Morocco, although progress has been made on data protection (notably with the CNDP), no dedicated AI governance law exists as of 2025.

Key legal questions remain open: Who is responsible for algorithmic decisions? How is consent obtained for data reuse across sectors? What redress mechanisms exist for those harmed by automated decisions? Tools such as algorithmic impact assessments, model cards [28], and participatory audits can help fill these gaps—provided they are adapted to the national legal and administrative context.

E. From Pilots to Policy: Scaling and Sustainability

Even when ML pilots succeed technically, scaling them across national systems involves coordination and long-term commitment. Risks include fragmentation between ministries, lack of budget continuity, and resistance from frontline staff. Sustainability requires clear mandates, cross-sectoral leadership, and iterative evaluation.

In the Moroccan context, integrating ML into social protection will require aligning innovations with broader institutional reforms, digitalization strategies, and governance structures. Multi-level collaboration—including with regional authorities and civil society—can enhance learning, accountability, and inclusiveness.

In summary, ML in social protection should not be viewed as a purely technical innovation. Its success depends on institutional maturity, ethical vigilance, transparent governance, and sustained public engagement. Addressing these challenges is critical for ensuring that data-driven systems enhance, rather than erode, social rights and democratic legitimacy.

VII. Survey on Public Perceptions of AI in Moroccan Social Policy

Following the discussion on institutional and ethical challenges, this section explores how citizens perceive the integration of AI and ML in Morocco's social protection system.

A. Objectives and Rationale

To complement the conceptual and policy discussion, this exploratory survey aims to capture how citizens perceive the integration of Artificial Intelligence (AI) and Machine Learning (ML) into Morocco's evolving social protection system. Understanding public sentiment is essential to ensuring the social acceptability and ethical legitimacy of algorithmic reforms [30], [31].

B. Methodology

This study relies on an online self-administered questionnaire designed to explore public perceptions of Artificial Intelligence (AI) and its potential role in social protection policy in Morocco. The survey was disseminated through academic mailing lists, social media platforms, and professional networks. Participation was entirely voluntary and anonymous, and the form was designed to be completed in less than ten minutes.

Sample Characteristics: A total of 50 responses were collected. While the sample is not representative in a statistical sense, it includes a diverse range of respondents from different socio-professional backgrounds, such as students, private sector employees, public sector workers, unemployed individuals, and retirees. This diversity provides useful exploratory insights into societal attitudes toward AI in welfare contexts.

Survey Structure: The questionnaire was composed of six thematic sections covering both demographic and thematic variables relevant to the research objectives:

1) **Gender:** The respondent was asked to indicate their gender (male or female).
2) **Socio-professional profile:** This included the current status of the respondent (public sector employee, private sector employee, student, retired, unemployed, or other).
3) **Familiarity with the CNSS and social protection system:** Respondents were asked whether they had previously used CNSS services (e.g., health coverage, pensions, benefits), and to self-assess their familiarity with Moroccan social protection policies (e.g., CNSS, AMO, social assistance).
4) **Awareness of Artificial Intelligence:** Questions focused on whether the respondent had heard about Machine Learning and their opinion on the potential of AI to improve the management of social assistance.
5) **Perceived opportunities and risks:** This section addressed the potential benefits (e.g., efficiency, transparency) and risks (e.g., bias, surveillance) of using AI in the welfare sector. Respondents were also asked about their comfort level with algorithmic decision-making and the conditions they deemed necessary for its acceptance.
6) **Trust in institutions and acceptability of AI use in CNSS:** Finally, the questionnaire assessed the respondent's trust in Moroccan institutions to use AI responsibly, and their opinion on the potential role of Machine Learning in improving CNSS services. A specific question also invited them to identify which CNSS services could most benefit from AI.

Data Collection and Limitations: The survey was conducted using a non-probability convenience sampling strategy. As such, the findings should be interpreted as exploratory rather than generalizable. In addition, the online nature of the questionnaire may have introduced a selection bias, particularly toward respondents with access to digital tools and some familiarity with technology [32].

C. Public Awareness of Machine Learning and Familiarity with Social Protection Policies

A key dimension of the survey focused on gauging baseline knowledge of both AI-related concepts and the functioning of Morocco's social protection system. These elements are critical for evaluating the feasibility and legitimacy of introducing algorithmic tools into welfare governance.

As shown in Figure 3, a large majority of respondents (78%) indicated that they were familiar with the term "Machine Learning," suggesting a strong presence of AI-related topics in public discourse. In contrast, only 26% reported being "very familiar" with the national social protection architecture—including institutions like CNSS or programs such as AMO—while 22% acknowledged having no familiarity at all.

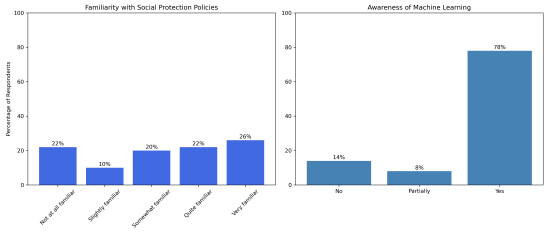

Fig. 3. Self-reported awareness of Machine Learning and familiarity with Moroccan social protection policies (n=50)

This gap reveals a disconnect between digital awareness and policy literacy. While the penetration of AI-related terminology may reflect increased exposure through media or education, understanding of social protection rights and structures remains fragmented. Bridging this gap is essential for fostering inclusive and informed debates on digital transformation in public services [33].

D. Perceived Opportunities and Concerns about AI

Beyond general awareness, the survey explored respondents' perceptions of the potential impacts of integrating AI into Morocco's social protection programs. Participants were asked to identify what they viewed as the main benefits and risks associated with such a reform.

Figure 4 summarizes the results. Among the most frequently cited benefits were time savings for administrative staff (86%), enhanced fraud detection (58%), and fairer targeting of beneficiaries (54%). These responses indicate a belief that AI could streamline operations and improve efficiency within institutions such as the CNSS.

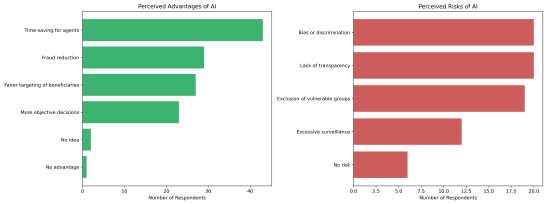

Fig. 4. Perceived benefits and risks of AI in social protection – based on survey responses (n = 50)

On the other hand, respondents expressed concerns regarding algorithmic bias and lack of transparency (each mentioned by 40%), as well as the possible exclusion of vulnerable groups (38%). These concerns echo broader discussions in the literature about the ethical risks of automated decision-making [7], [13].

E. Institutional Trust and Perceived Usefulness of ML for CNSS

Trust in institutions emerged as a pivotal determinant for AI acceptability. Respondents were asked both how much confidence they placed in Moroccan institutions to manage AI ethically and under what conditions they would accept its use in CNSS services.

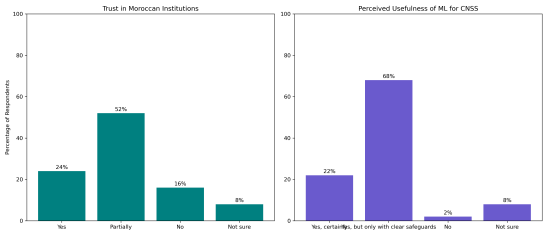

Fig. 5. Perceptions of Institutional Trust and ML Utility

Only 24% of respondents reported full confidence in Moroccan institutions to use AI responsibly, while 52% indicated partial trust. Importantly, 68% supported the use of ML for CNSS services if legal safeguards were in place. This reinforces the idea that the perceived benefits of AI are conditional on robust regulation, accountability, and transparency mechanisms [8], [34].

F. CNSS Services Most Likely to Be Enhanced by Artificial Intelligence

To identify perceived priorities, respondents were asked to select which CNSS services would most benefit from AI. As seen in Figure 6, fraud detection (40%) and targeting of beneficiaries (24%) were the top responses.

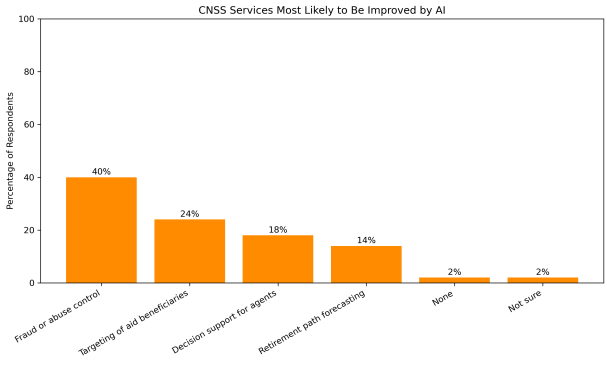

Fig. 6. Top CNSS services perceived as most improvable by AI (n=50)

Decision support for agents (18%) and forecasting retirement trajectories (14%) were also noted. These insights reflect citizen expectations for AI tools that enhance both the efficiency and fairness of public service delivery [35].

G. Interpretation and Implications

While exploratory, the survey reveals a cautiously optimistic public stance: citizens are generally supportive of AI in social protection, especially for fraud control and targeted distribution—on the condition that institutions adopt transparent and ethical governance practices.

These findings align with prior sections highlighting the importance of combining technical innovation with institutional legitimacy. For policymakers, they emphasize the need to pair AI initiatives with public communication strategies, legal safeguards, and participatory approaches [31], [36].

VIII. Conclusion and Future Research Directions

This article has explored the transformative potential of Machine Learning (ML) in modernizing social protection systems in developing countries, with a particular focus on Morocco. By combining a conceptual analysis, a review of international experiences, and an exploratory public perception survey, the study provides a multidimensional understanding of the opportunities and challenges that accompany the integration of AI into welfare governance.

Several key insights emerge:

- ML can enhance social protection through improved targeting, fraud detection, forecasting, and operational efficiency — provided it is implemented with transparency and oversight.
- International case studies show that success requires robust legal frameworks, institutional capacity-building, and participatory design to avoid reinforcing inequalities or biases.
- In the Moroccan context, while public attitudes toward AI are generally positive, trust in institutions remains moderate. Citizens express support for digital reforms when accompanied by legal safeguards and inclusive governance.

Based on these findings, several policy recommendations can be outlined:

1) **Strengthen legal and ethical frameworks** to govern algorithmic decision-making, including provisions for transparency, explainability, and accountability.
2) **Invest in civic and digital literacy**, ensuring citizens understand both social protection rights and the implications of AI-based tools.
3) **Promote human-in-the-loop systems** to retain institutional oversight and mitigate automation-related risks.
4) **Encourage co-design and participatory implementation** of AI systems to align technical design with societal values and lived experiences.

From a research perspective, several concrete avenues remain open and will guide future work:

- **Empirical testing of ML use cases with administrative data.** A first line of inquiry will consist of developing and validating ML models—such as logistic regression, Random Forests, XGBoost, and clustering algorithms—using real administrative data from the Unified Social Registry (RSU), the CNSS, and AMO schemes. This will make it possible to rigorously assess the contribution of ML to eligibility assessment, fraud detection, contribution monitoring, and proactive outreach, compared with existing rule-based approaches.
- **Operationalization of governance and ethical safeguards.** A second avenue will focus on designing and testing practical governance tools for responsible ML deployment in social protection. Future work will explore algorithmic impact assessments, transparency dashboards, explainability mechanisms (e.g., SHAP, LIME), and human-in-the-loop validation protocols, in order to ensure that predictive systems remain fair, contestable, and socially acceptable.
- **Territorial pilots and institutional readiness.** A third research direction will involve implementing territorial pilots in partnership with Moroccan institutions to examine how ML tools can be integrated into existing workflows. These pilots will help identify organizational constraints, capacity gaps, and training needs for civil servants, while generating empirical evidence on the operational and social impacts of ML-supported decision-making.
- **Deepening the analysis of public perceptions and participation.** Finally, future work will extend the exploratory survey presented in this article through larger and more representative samples, complemented by qualitative methods such as interviews and focus groups. This will make it possible to better understand citizens' expectations and concerns and to inform participatory co-design processes for AI in social policy.

In sum, ML offers promising tools to enhance social protection systems in developing countries. However, its success will depend not only on technical innovation but also — and critically — on institutional trust, ethical safeguards, and inclusive public engagement. Policymakers must therefore adopt a dual strategy: building algorithmic capacity while reinforcing democratic legitimacy, accountability, and citizen participation in digital welfare reform.

REFERENCES

[1] International Labour Organization (ILO), "World social protection report 2020–22: Social protection at the crossroads – in pursuit of a better future," Geneva: ILO, 2021. [Online]. Available: https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed_protect/@soc_sec/documents/publication/wcms_817572.pdf

[2] Organisation for Economic Co-operation and Development (OECD), "Breaking the vicious circles of informal employment and low-paying work," Paris: OECD Publishing, 2024. [Online]. Available: https://doi.org/10.1787/f95c5a74-en

[3] Overseas Development Institute (ODI), "Getting to grips with heterogeneity in the informal economy," ODI, 2024. [Online]. Available: https://odi.org/documents/9238/ODI_Literature_Review_-_Getting_to_grips_with_heterogeneity_in_the_informal_economy.pdf

[4] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark, and F. Fuso Nerini, "The role of artificial intelligence in achieving the sustainable development goals," *Nature Communications*, vol. 11, no. 1, pp. 1–10, 2020, [Indexed in Web of Science, Scopus].

[5] Caisse Nationale de Sécurité Sociale (CNSS), "Rapport annuel 2022," 2022. [Online]. Available: https://www.cnss.ma/fr/actualites/rapport-annuel-2022

[6] Organisation for Economic Co-operation and Development (OECD), "Ai and the future of social protection in oecd countries," https://doi.org/10.1787/7b245f7e-en, 2021.

[7] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021, [Indexed in Scopus and Web of Science].

[8] V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

[9] E. Aiken, S. Bellue, D. Karlan, C. Udry, J. E. Blumenstock *et al.*, "Machine learning and phone data can improve targeting of humanitarian aid," *Nature*, vol. 603, pp. 864–870, 2022, [Indexed in Scopus and Web of Science].

[10] M. C. Elish and d. boyd, "Situating methods in the magic of big data and ai," *Communication Monographs*, vol. 86, no. 2, pp. 213–229, 2019, [Indexed in Scopus].

[11] World Bank, "Reshaping social protection systems for resilience and inclusion," 2023. [Online]. Available: https://www.worldbank.org/en/news/feature/2023/04/18/reshaping-social-protection-for-inclusion-and-resilience

[12] S. Mohamed, M.-T. Png, and W. Isaac, "Decolonial ai: Decolonial theory as sociotechnical foresight in artificial intelligence," *Philosophy and Technology*, vol. 33, no. 4, pp. 659–684, 2020, [Indexed in Scopus].

[13] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, J.-C. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, and E. Vayena, "Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations," *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018, [Indexed in Scopus].

[14] U. Nations, "Roadmap for digital cooperation," 2021. [Online]. Available: https://www.un.org/en/content/digital-cooperation-roadmap

[15] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint*, 2017.

[16] R. Binns, "Fairness in machine learning: Lessons from political philosophy," *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency*, pp. 149–159, 2018, [Indexed in scopus].

[17] C. Lowe, J. Rigolini, L. S. Castro, and F. Bastagli, "Pathways toward digitalization in social protection and labor service delivery," World Bank, Social Protection & Jobs Discussion Paper 2307, 2023. [Online]. Available: https://documents1.worldbank.org/curated/en/099100523151038686/pdf/P173530036a7f30f60899c06a6172040bf3.pdf

[18] European Union, "General data protection regulation (gdpr)," 2016, regulation (EU) 2016/679. Available at: https://gdpr.eu/.

[19] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016, [Indexed in Scopus and WOS].

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017, [Indexed in Scopus and WOS].

[21] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016, [Indexed in Scopus and WOS].

[22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, 1996, [Indexed in Scopus and WOS].

[23] Organisation for Economic Co-operation and Development(OECD), "Digital government index: 2022 edition," 2022, [Online]. Available: https://www.oecd.org/gov/digital-government-index.htm.

[24] T. Sun and R. Medaglia, "Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare," *Government Information Quarterly*, vol. 36, no. 2, pp. 368–383, 2019, [Indexed in Scopus and Web of Science].

[25] K. Muralidharan, P. Niehaus, and S. Sukhtankar, "Building state capacity: Evidence from biometric smartcards in india," *American Economic Review*, vol. 110, no. 10, pp. 3368–3416, 2020, [Indexed in Scopus].

[26] Government of Chile, "Registro social de hogares," 2021. [Online]. Available: https://rsh.ministeriodesarrollosocial.gob.cl/

[27] A. Gelb and A. Metz, "Identification revolution: Can digital id be harnessed for development?" 2018. [Online]. Available: https://www.cgdev.org/publication/identification-revolution-can-digital-id-be-harnessed-development

[28] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru, "Model cards for model reporting," *Communications of the ACM*, vol. 64, no. 12, pp. 56–65, 2021, [Indexed in Scopus].

[29] V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

[30] M. C. Elish and d. boyd, "The moral crises of ai: A case for algorithmic accountability," *SSRN*, 2019. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3432050

[31] United Nations, "Un secretary-general's roadmap for digital cooperation: Principles for digital governance," 2021, [Online]. Available: https://www.un.org/en/content/digital-cooperation-roadmap.

[32] J. A. Maxwell, *Qualitative Research Design: An Interactive Approach*. SAGE Publications, 2012.

[33] P. M. Krafft *et al.*, "Machine learning for social good in africa: A landscape review," *Nature Machine Intelligence*, vol. 3, no. 2, pp. 89–93, 2021.

[34] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, pp. 389–399, 2019.

[35] R. Lentz and T. Tranaes, "Targeting and monitoring in social assistance: Evidence from a digital reform," *American Economic Journal: Economic Policy*, vol. 14, no. 3, pp. 1–27, 2022, [Indexed Scopus and Web of Science].

[36] M. Mitchell *et al.*, "Model cards for model reporting," *Communications of the ACM*, vol. 64, no. 12, pp. 56–65, 2021, [Indexed Scopus and Web of Science].