Draft – Open for Feedback

This document is a draft version and is temporarily available for review.

Following the presentation during the clinic session: *AI Readiness: Are you ready to use AI in Social Protection?,* we invite you to share any feedback on the content, structure, or usability of this tool. If you would like to contribute to shaping it further, please send your input by **28th of April 2026** to honga.ross@giz.de

AI Readiness for Social Protection

A Tool Outline for Consultation and Discussion

Part 1: Context and Purpose

This document is prepared for an internal discussion with AI Hub partners. It aims to inform the development of an AI readiness for social protection tool that will be shared with participants during **Webinar 1 – AI Readiness: Are you ready to use AI in Social Protection?**

The tool will respond to the lack of self-assessment resources designed specifically for the staff and leadership of social protection agencies (the programme managers, digital leads, policy officers, data teams, and senior officials) who make day-to-day and strategic decisions about whether, when, and how to introduce AI. Existing frameworks either operate at the level of national government readiness (too broad), or focus on technical AI governance in private sector or high-income country settings (too narrow or not contextually applicable).

The proposed aim for this tool is to help an agency answer a practical question: are we institutionally ready to introduce AI responsibly into our work - and if not, what do we need to do first?

This is deliberately an institutional readiness question, not a technical one. The evidence reviewed in preparing this outline points consistently to a single finding: when AI has caused harm in social protection, the failure has rarely been primarily technical. It has typically been a failure of governance, legal basis, data quality, human oversight, or accountability. The technology worked as designed, the institution was not ready for what the technology would do.

1.1. Potential audience for the tool

The proposed users are the staff and leaders of social protection agencies in all country contexts. Not AI specialists, data scientists, or lawyers, but the people responsible for running programmes and making decisions about digital transformation. The tool should therefore be usable without technical expertise, while still being substantive enough to surface real institutional gaps.

Part 2: What the Evidence Base Tells Us

This outline draws on an AI assisted literature review of 28 resources[1], including AI readiness frameworks, social protection AI guidance, and documented case studies of AI deployment in welfare and social protection systems. A brief summary of the most relevant findings is set out below.

2.1 The core frameworks informing this outline

Several frameworks have shaped the broad architecture of the areas proposed in this outline.

Frameworks shaping this outline
OECD: Harnessing Artificial Intelligence in Social Security (2025) | OECD The most comprehensive published analysis of AI in social security systems. Identifies governance, data quality, workforce capacity, and legal compliance as the core readiness dimensions and draws on real deployments across EU countries. Link
WHO / IDB: Artificial Intelligence in Public health Readiness assessment toolkit Health focused toolkit. Link
DCI: A Data Governance Framework for Digital Social Protection Systems (2025) | Digital Convergence Initiative The first data governance framework designed specifically for social protection, structured around management, quality, access, and security. Link
GIZ / ADB: AI in Social Protection — Exploring Opportunities and Mitigating Risks (2020) | GIZ and Asian Development Bank Foundational practitioner guide covering the main AI applications in social protection and the risks associated with premature adoption in LMIC contexts. Link
UNESCO: Readiness Assessment Methodology (RAM) | UNESCO Structured country-level diagnostic implemented in 58+ countries, covering legal/regulatory, socio-cultural, economic, scientific/educational, and technological dimensions. Link
UNDP: Artificial Intelligence Landscape Assessment (AILA) (2025) | UNDP Modular AI readiness diagnostic covering ecosystem, government use, and regulation. Conducted in 15+ countries. Useful for contextualising agency-level readiness within national conditions. Link
NIST: AI Risk Management Framework 1.0 (2023) | US National Institute of Standards and Technology The most widely adopted voluntary cross-sector AI risk framework. Four functions — Govern, Map, Measure, Manage — provide a lifecycle logic that informs the sequencing of readiness questions. Link
Ada Lovelace Institute, AI Now & OGP: Algorithmic Accountability for the Public Sector (2021) | Ada Lovelace Institute / AI Now Institute / Open Government Partnership Analysis of 40+ accountability policies from 20+ governments. Establishes that transparency without enforcement is ineffective, and that meaningful public participation is the least-developed dimension globally. Link

2.2 What case studies tell us about failure

A substantial body of documented case evidence now exists on what happens when AI is introduced into social protection without adequate institutional readiness. The cases share

[1] With human review of all resources

remarkably consistent patterns of failure, regardless of country income level or technical sophistication.

Selected resources and case studies
Australia's Robodebt scheme (2016–2020) Automated debt recovery that issued 500,000+ illegitimate debts. A Royal Commission found it was unlawful from the outset. Internal legal warnings were suppressed. Human review was removed entirely. A class action was settled for AU$1.8 billion.
Netherlands: SyRI system and Childcare Benefits algorithm SyRI, a mass fraud profiling system, was banned by the Hague District Court in 2020 as a violation of the European Convention on Human Rights. The Childcare Benefits algorithm falsely accused 26,000+ families (disproportionately migrants) of fraud. The government resigned.
Michigan MiDAS unemployment system (2013–2015) Achieved a 93% error rate, wrongly accusing ~40,000 people of fraud. Human review was eliminated. The agency treated rising penalty revenue as a success indicator. Litigation was not resolved until 2024.
Serbia Social Card Registry (2022–present) A World Bank-funded system drawing data from multiple databases, but data for Roma and disabled people is structurally outdated. Caseworkers report being unable to override the system, even when they know it is wrong.
Denmark: Udbetaling Danmark welfare AI (2024) Amnesty International documented up to 60 AI models used to flag welfare fraud. Risk of discrimination against people with disabilities, migrants, and marginalised groups. Authorities rejected collaborative audit requests.
Human Rights Watch: Automated Neglect — Jordan Takaful programme (2023) Investigation of a World Bank-financed algorithmic cash transfer system. Documents how algorithm-driven poverty targeting excluded vulnerable families through flawed models and lack of transparency.
UN Special Rapporteur: Digital Welfare States and Human Rights (2019) Landmark report cataloguing AI-driven welfare harms across 30+ countries and establishing international human rights standards for digital welfare systems.
Virginia Eubanks: Automating Inequality (2018) Three US case studies documenting how automated systems consistently deepen disadvantage for low-income populations, regardless of the system's stated intentions.

2.3 What the evidence suggests about readiness

Across both the frameworks and the failure cases, a consistent picture emerges of the conditions that must be in place before AI is introduced into social protection. These are not primarily technical conditions. They are institutional ones.

- Every major documented failure involved a governance gap - no one with authority to stop the system when it began causing harm.
- Legal authorisation for automated or AI-supported decisions is frequently assumed rather than verified - and assumptions have proved wrong.
- Data quality problems for marginalised populations are known before deployment and treated as problems to be fixed later. They rarely are.
- The populations most affected by AI-supported decisions are the least involved in designing or reviewing those systems.

- Efficiency and cost metrics dominate performance frameworks; rights and equity metrics are often absent or unmeasured.

The implication for this tool is that readiness questions must probe institutional conditions - not just technical capabilities. An agency may have excellent data infrastructure but no governance structure capable of exercising oversight. It may have a legal team but no process for reviewing the legal basis of AI-supported decisions. The tool needs to surface these gaps.

Part 3: Draft Outline of the Tool

Based on the literature review and the case evidence, we propose that the tool should cover six broad areas. These are presented below as a proposed structure for discussion. For each area, we describe what it is about, why it matters for social protection specifically, which existing resources offer relevant material, and what kinds of questions a finished tool in this area might explore.

A key design question is whether these areas should be treated as equally weighted dimensions, or whether some should function as threshold conditions - areas where a significant gap means it is too early to proceed with AI, regardless of how well the agency scores elsewhere. The evidence suggests that governance and legal readiness are candidates for this threshold role.

Area A. Strategic and Political Readiness

What this area is about

This area examines whether the institution has a clear, committed, and accountable vision for AI - and whether that vision is grounded in an honest understanding of what AI can and cannot do in social protection contexts. It considers if AI use within the institution has clear ownership, a documented rationale, a risk appetite, and the authority to say no as well as yes.

Why it matters for social protection agencies

The most consistent precursor to AI failure in social protection is the absence of meaningful leadership engagement. Specifically, leadership that understands the risks, not just the potential. In every documented failure case, there was political or managerial commitment to the outcome (efficiency, fraud reduction, speed) but no commitment to the process of responsible adoption. The Robodebt Royal Commission found that ministers were told the scheme was lawful when it was not, and that no one in the system had both the knowledge and the authority to stop it.

For agencies in lower-income contexts, a distinct risk is that AI adoption is driven primarily by donor enthusiasm, vendor demonstrations, or national digital strategy pressures, without internal ownership or genuine institutional readiness assessment. Readiness requires that someone inside the institution actually owns the question.

Types of questions this area might explore

The questions below are illustrative. They are intended to prompt discussion about what a finished tool should ask in this area, not as a proposed final question set.

Types of questions this area might explore
Does the institution have an explicit, documented rationale for why AI is being considered - and does that rationale go beyond efficiency or cost reduction to include rights and equity considerations?
Who is responsible for decisions about AI adoption - and do they have both the knowledge to understand the risks and the authority to pause or stop an AI deployment if problems emerge?
Has the institution defined which AI use cases it will not pursue, and why? (The absence of any prohibited uses list is itself informative.)
Is AI adoption in this institution being driven primarily by internal analysis and readiness, or by external pressures - donor requirements, vendor proposals, national strategy mandates?
Does leadership have a realistic understanding of what AI can and cannot do in this institutional context - including the quality of available data, the maturity of existing systems, and the capacity of staff?
Is there a process for reviewing and approving AI initiatives before they are started - not just monitoring them after they are deployed?

Questions for discussion

How should this area treat the difference between an agency that has a strong AI strategy but weak implementation, versus one that has no formal strategy but is making careful, supervised decisions in practice? Should this area focus on documented commitments, or on observable behaviours?

Area B. Legal and Regulatory Readiness

What this area is about

This area examines whether there is a sound legal basis for the ways AI is being, or is planned to be, used - and whether the legal rights of beneficiaries are protected in practice, not only on paper. It covers data protection law, the legal authorisation for automated or AI-supported decisions, the right to explanation and appeal, and anti-discrimination obligations.

Why it matters for social protection agencies

Legal readiness is the dimension most frequently underestimated and most catastrophically consequential when absent. Australia's Robodebt scheme was not a technical failure - it was an unlawful scheme. Internal legal advice that the automated income averaging on which it was based had no legal foundation was ignored for years. The Dutch SyRI fraud profiling system was operating for years before a court ruled that it violated the European Convention on Human Rights and ordered it shut down.

A common assumption among social protection agencies is that legal questions are for lawyers, not programme teams. But the evidence suggests that the most dangerous legal gaps are not obscure points of law - they are fundamental questions that programme teams need to be asking themselves. Does existing legislation actually permit us to use AI to make or support this decision? If a beneficiary is denied a payment partly on the basis of an AI assessment, can they find out why, and challenge it? Does the data we are feeding this system contain categories (e.g. nationality, disability, ethnicity) that require specific legal authorisation to process in this way?

The EU AI Act (2024) is now the most significant international regulatory development in this space. It classifies AI systems used to determine eligibility for or access to essential public services (which includes social protection) as high-risk, imposing mandatory requirements. While this regulation applies directly only in the EU and to systems deployed there, it is shaping standards globally and is likely to influence the regulatory environment in many countries over the coming years.

Types of questions this area might explore

Types of questions this area might explore
Is there a clear legal basis (in legislation, regulation, or ministerial order) for using AI or automated tools to support eligibility, payment, or targeting decisions in this programme?
Do national data protection laws cover the personal data being processed by AI systems in this agency, and is there an authority with the capacity to enforce them?
Are beneficiaries legally entitled to an explanation of AI-supported decisions that affect them - and is this right accessible in practice, not just on paper?
Has the agency conducted a legal review of any planned AI use case against data protection, anti-discrimination, and administrative law - and was that review independent and documented?

Are there categories of sensitive data being processed (nationality, disability, ethnicity, health) for which specific legal authorisation is required, and has that authorisation been confirmed?
Is the agency's AI procurement process checking whether vendors' systems comply with applicable legal requirements, including data residency, security, and accountability obligations?

Questions for discussion

Many agencies in lower-income countries operate in contexts where the legal framework for AI in public services is underdeveloped or absent. How should this area treat contexts where the gap is not a failure of the agency but a gap in the national legal environment? Should the tool flag this as an external constraint to be escalated, rather than an internal readiness gap?

Area C. Data Foundations

What this area is about

This area examines the quality, governance, and completeness of the data that AI systems in social protection will use. It is concerned not only with whether data exists and is digitised, but with whether it is accurate, current, standardised, and whether it is equally reliable for the populations the agency is trying to serve, including those who are hardest to reach and most in need.

Why it matters for social protection agencies

AI systems in social protection are only as good as the data they consume. This is not a new observation, but the evidence base now shows with particular clarity that data quality problems in social protection are not random, they are systematically worse for the people social protection systems most need to include.

The Serbia Social Card case is instructive. The system was technically sophisticated. But the data it drew on for Roma populations and people with disabilities was structurally outdated and incomplete - a legacy of decades of administrative exclusion. The AI did not create this inequality; it amplified and automated it. The same pattern appears in the India Aadhaar cases, in Jordan's Takaful targeting algorithm, and in multiple other documented cases. Algorithmic systems inherit and scale whatever biases exist in their training and input data.

The DCI Data Governance Framework (2025) identifies data quality for marginalised populations as the single most important and most neglected readiness dimension globally in social protection digitalisation. An agency's readiness to use AI responsibly depends heavily on whether it has genuinely confronted this question - not whether it has good data on average, but whether it has good data for the people most at risk of exclusion.

Types of questions this area might explore

Types of questions this area might explore
For the populations this AI system would affect, how complete and current is the underlying data - and are there known gaps for specific groups (women, people with disabilities, rural households, migrants, ethnic minorities)?
Has the agency conducted a data quality assessment specifically for the data it plans to use in an AI system - not just an overall data audit, but one focused on the populations the AI will make decisions about?
Are data-sharing arrangements between agencies and systems formalised in legal agreements, with clear ownership and accountability for data quality?
Is there a designated person or team responsible for data quality - someone who is accountable when the data is wrong, not just when the system produces a wrong answer?
Does the agency use standardised data objects (unique identifiers, consistent address formats, verified disability status codes) across the systems that would feed an AI?

Are there feedback mechanisms that allow errors in AI-processed data to be identified, reported, and corrected systematically - not just on a case-by-case basis when an individual complains?

Questions for discussion

This area overlaps significantly with the DCI's existing Country Readiness Questionnaire (Section 3). How should the AI readiness tool relate to that questionnaire in practice - should this area essentially cross-reference and build on it, or should it ask different questions with a specific AI focus? And how should the tool handle the common situation where an agency knows it has data quality gaps but is proceeding anyway - is this a readiness failure, or a managed risk?

Area D. Technology and Infrastructure

What this area is about

This area examines whether the underlying digital infrastructure is capable of supporting AI deployment in a way that is safe, reliable, and inclusive. It covers the maturity of core information systems, connectivity, digital identity infrastructure, cybersecurity, and the agency's understanding of different AI deployment options and their implications.

Why it matters for social protection agencies

Infrastructure readiness is the technical floor beneath all other readiness dimensions. An agency cannot deploy AI responsibly on top of fragmented legacy systems, unreliable connectivity, or digital identity infrastructure that excludes the people it is meant to serve.

The India Aadhaar-linked welfare cases illustrate what happens when infrastructure readiness is assumed rather than verified. Biometric authentication fails systematically for manual labourers with worn fingerprints, for elderly people, and for rural populations with poor connectivity. In some documented cases, these failures meant that people entitled to food rations could not access them - with fatal consequences. The infrastructure problem was not hidden: it was known, and the system was deployed anyway.

A distinct infrastructure question for social protection agencies is the choice of how AI is deployed. Bozdag and Bennati (2026) distinguish three deployment models - using AI as a service through a vendor application (SaaS), integrating a third-party AI model via API, or hosting AI on infrastructure the agency controls. Each model has very different implications for data sovereignty, accountability, and what the agency is actually responsible for. Many agencies are making these choices without fully understanding the implications, particularly in low and middle-income country contexts where cloud-based SaaS may be attractive for cost reasons but raises significant data residency and security concerns.

Types of questions this area might explore

Types of questions this area might explore
How mature are the core information systems that an AI system would operate within or alongside - and are they capable of supporting the data flows, logging, and audit trails that responsible AI use requires?
Is connectivity sufficiently reliable across all the service delivery points where AI would be used - including rural and remote areas - to ensure the system functions consistently and does not create new access inequalities?
Does the digital identity infrastructure exclude any groups from accessing services that would be partly administered by AI - and are there alternative pathways for people who cannot use digital authentication?
Does the agency understand the difference between using AI as a cloud-based service, integrating AI via an API, and hosting AI on its own infrastructure - and has it considered the data sovereignty and accountability implications of each?

Are there documented cybersecurity standards and active security monitoring for systems that process beneficiary data - including any vendor systems the agency is using or considering?
Does the agency have the technical capacity to audit AI systems it is using - or is it entirely dependent on vendors' self-reporting?

Questions for discussion

There is a risk that infrastructure questions dominate readiness discussions at the expense of governance and legal questions, because they feel more tangible and solvable. How should the tool be designed to ensure that infrastructure gaps do not mask - or be used to distract from - more fundamental governance and accountability gaps? And should this area include questions about the agency's approach to AI procurement, given that many infrastructure and data sovereignty risks are created at the point of vendor selection?

Area E. Institutional Capacity and Governance

What this area is about

This area examines whether the institution has the structures, skills, and human oversight mechanisms needed to deploy and manage AI responsibly. It covers internal governance arrangements, the role of human judgment in AI-supported decisions, staff capacity and literacy, accountability mechanisms, and the institution's ability to respond when things go wrong.

Why it matters for social protection agencies

This is the dimension most consistently implicated in failures - and most consistently underestimated. Michigan's MiDAS system achieved a 93% error rate not because of poor engineering, but because human review was eliminated entirely, a third of the workforce was made redundant, and the agency treated rising penalty revenue as evidence the system was working. The system produced harm at scale because there was no institutional capacity to exercise judgment over it.

The problem is not simply about having technically skilled staff. It is about whether any person inside the institution has both the authority and the responsibility to question an AI system's outputs, override them when they are wrong, and escalate concerns without career risk. The Serbia case documents caseworkers who told beneficiaries: 'There is nothing I can do, it is the system from Belgrade that decided.' This is the institutional failure that human oversight requirements are designed to prevent - not just staff who could question the system, but staff who understood they were permitted and expected to do so.

Bozdag and Bennati (2026) identify the Responsible AI function (a cross-functional team with legal, technical, and ethics expertise, and real authority) as the governance linchpin without which all other governance measures are nominal. In a social protection agency, this function need not be a large dedicated unit; but it must have a genuine mandate, not just a name.

Types of questions this area might explore

Types of questions this area might explore
Is there a person or team with both the expertise and the authority to review AI systems before deployment, monitor them in operation, and require changes if they are causing harm?
Are caseworkers and frontline staff who interact with AI-supported decisions expected to exercise their own judgment - and are they given the tools, training, and institutional permission to do so?
Does the agency have an AI incident or error reporting process - and is there evidence it is actually used, rather than just documented?
What happens when an AI system produces an output that a member of staff believes is wrong? Is there a clear, accessible process for raising this - and evidence that such concerns are taken seriously?
Does the institution have enough internal technical capacity to understand what an AI system is doing - or is it entirely dependent on vendor explanations?

Has the agency conducted any training or awareness-raising on AI — its potential, its risks, and how staff should interact with AI-supported decisions — and does this reach frontline staff, not just senior managers?

Questions for discussion

This area is likely to be the most challenging for agencies to self-assess honestly, because it requires acknowledging governance gaps that may feel embarrassing or imply criticism of leadership. How should the tool be designed to encourage candid responses in this area? And should this area function as a threshold condition - where a significant gap here means the agency is not ready to proceed, regardless of how it scores elsewhere?

Area F. Accountability, Ethics, and the Rights of Beneficiaries

What this area is about

This area examines whether AI-supported decisions can be explained and contested, whether the agency has assessed potential for bias or discrimination, whether affected communities are involved in shaping systems that affect their lives, and whether there are meaningful boundaries on how AI is and is not used. It is the area most directly concerned with the rights and interests of the people social protection systems serve.

Why it matters for social protection agencies

Explainability and contestability are not technical features - they are rights. Under international human rights law and in most domestic administrative law systems, people are entitled to understand the basis of decisions that affect them and to challenge those decisions. When an AI system makes or supports those decisions, the agency is responsible for ensuring these rights remain real in practice, not just nominal.

The evidence base shows that these rights are routinely undermined in practice, in ways that agencies often do not recognise as a failure until it becomes a crisis. The Amnesty International investigation into Denmark's welfare AI system found that the authority responsible for administering it was unable to explain how the models worked, declined to allow external audit, and had no systematic process for identifying whether specific groups were being disadvantaged. None of this was the result of bad intentions - it was the result of deploying AI without the accountability infrastructure that responsible use requires.

Community participation is the least-developed dimension in almost every accountability framework reviewed. Beneficiaries are the people with the most at stake in these systems and frequently the most relevant knowledge about how they work in practice - which categories are systematically miscoded, which verification requirements are impossible to meet, which appeal routes do not function. Their involvement in design and review is not a consultation exercise; it is a quality assurance requirement.

Types of questions this area might explore

Types of questions this area might explore
Can the agency explain, in plain language, how an AI-supported decision was reached - and can it do so in a way that is genuinely useful to the beneficiary who received it, not just to a technical reviewer?
Is the appeals mechanism for AI-supported decisions accessible in practice to the people most likely to be affected - including those who are not literate, do not have internet access, or live far from administrative offices?
Has the agency assessed whether the AI system it is using or planning to use could produce different outcomes for different groups - by gender, disability, ethnicity, location, or other characteristics - and is it monitoring for this?

Are the people whose lives are most affected by AI-supported decisions involved in reviewing and improving those systems - not just informed about them?
Has the agency defined any AI use cases it will not pursue - and if so, on what basis?
Is the agency open to external review of its AI systems (by regulators, civil society organisations, academic researchers, or beneficiary groups) and are there mechanisms to make such review possible?

Questions for discussion

This area raises the deepest questions about the purpose of social protection - who it is for, who has a right to shape it, and what accountability to beneficiaries actually means in practice. Some agencies will have sophisticated views on participation and co-design; many will not. How should the tool handle this area in a way that is challenging and aspirational without being unrealistic or preachy? And should there be explicit 'minimum standard' requirements in this area (things every agency should be doing before deploying AI) distinct from aspirational good practice?

Part 4: Design Considerations for Discussion

The six areas set out above are a proposed structure, not a settled one. Before the tool is developed further, there are a number of design questions that need to be resolved through consultation. These are set out below as prompts for discussion, not as recommendations.

4.1 Scope: agency-level or use-case level?

Should the tool assess the readiness of an institution as a whole to adopt AI, or the readiness to implement a specific AI use case? The two assessments would ask somewhat different questions. An agency-level assessment examines general governance, legal, and data conditions. A use-case assessment examines whether a specific application (eg using machine learning for fraud detection in a particular programme) is ready to proceed.

The case for a use-case focus is that readiness is highly context-dependent: an agency might be well-prepared to use AI for administrative efficiency in back-office processes but not for eligibility determination. The case for an institution-level focus is that it is more actionable for leadership, and that many of the foundational conditions (legal basis, governance structure, data quality) need to be in place regardless of the specific use case.

4.2 Threshold conditions versus dimensions

Should some areas function as threshold conditions (eg where a significant gap means it is not appropriate to proceed with AI regardless of how the agency performs in other areas?). The evidence base suggests that governance (Area E) and legal basis (Area B) are candidates for this role. A technically sophisticated agency with excellent data and infrastructure but no human oversight structure and no legal basis for AI decisions is not ready, regardless of its other scores.

This is a design choice with significant implications for how the tool is used and communicated. It would mean that the tool is not just a readiness profile but a gating mechanism - and that needs to be discussed with the agencies who will use it.

4.3 Format and facilitation

How should the tool be administered? Options include: a self-assessment questionnaire completed individually by a designated focal point; a facilitated group conversation across functions; a supported assessment with external facilitation from the AI Hub; or some combination. The DCI's experience with the Country Readiness Questionnaire and the World Bank's Playbook self-assessment offer useful reference points for what works in practice for social protection agencies.

The evidence from similar tools (UNESCO RAM, UNDP AILA) suggests that group facilitation consistently produces richer and more honest assessments than individual completion - particularly for the governance and accountability areas, where the most important gaps may only become visible when people from different parts of the institution compare their understanding of the same system.

4.4 Relationship to existing tools

The DCI Country Readiness Questionnaire for DCI Standards Adoption covers significant ground that is relevant to this tool, particularly in the data and infrastructure areas. The design challenge is to create a tool that complements rather than duplicates this questionnaire. Ideally one that can be used in sequence with it, and that explicitly cross-references relevant sections rather than requiring agencies to answer the same questions twice.

Similarly, national-level tools such as the UNESCO RAM and UNDP AILA assess country readiness conditions that form the context within which agency-level readiness sits. The tool should help agencies understand their position within that national context, not just their internal capacity.

4.5 Audience and language

The intended users are the staff and leadership of social protection agencies, not AI specialists. The tool must be usable by a programme manager who has not worked with AI before, alongside a digital transformation lead who has. This means the language must be accessible, the concepts must be grounded in social protection practice rather than AI theory, and the examples used must reflect the realities of agencies in diverse country contexts - not just high-income country examples.

At the same time, the tool must be substantive enough to surface real institutional gaps, not so simplified that experienced practitioners find it superficial. Getting this balance right will require iteration and testing with practitioners.

Part 5: Questions for the Consultation Group

On the scope and structure

- Do the six areas proposed cover the right ground? What is missing? What is over-emphasised?
- Are any of the areas wrongly framed for the context of social protection agencies in lower or middle-income countries?
- Should any area function as a threshold condition (a gate that must be passed before AI use is appropriate) rather than simply a dimension to be assessed?

On the questions

- Within each area, what are the two or three questions that matter most for social protection agencies in practice?
- Are there important readiness questions that cut across multiple areas (for example, questions about community participation, or about vendor accountability) that should be treated as their own strand?
- What does 'good enough' look like in each area for an agency in a fragile state, a lower-middle income context, and a high-income context? Should the tool have context-adjusted baselines?

On the design

- Should the tool assess institutional readiness in general, or readiness for a specific use case? Or both?
- What format would work best for the intended users - a structured questionnaire, a facilitated discussion guide, a scoring tool, or something else?
- How should the tool relate to the existing DCI Country Readiness Questionnaire? Complement, cross-reference, or integrate?
- What would make this tool genuinely useful to a senior official in a social protection ministry who has been asked to decide whether to proceed with an AI pilot? What would they need it to tell them?

On what the tool should not do

- Are there framings or approaches in this outline that would not work in practice - that would feel preachy, unrealistic, or disconnected from the realities of social protection administration?
- Are there areas where the risk of the tool being used to justify premature AI adoption (rather than to honestly assess readiness) is high - and how should the design guard against this?