ISDM
INDIAN SCHOOL OF DEVELOPMENT MANAGEMENT

AI Systems for Social Good

A Practitioner's Guide for Leaders of Social Purpose Organisations

March 2026

Version: 1.0 | March 2026

Created by: Centre for Data Science and Social Impact (CDSSI), ISDM

Purpose:
This guide serves as your step-by-step guide for designing responsible AI and data analytics solutions in the social sector. It guides both technical and non-technical teams through real-world projects, from understanding your context to monitoring your model post-deployment. Rooted in field-tested work, it aims to make AI more accessible, ethical, and impactful.

Author: Roopa Sharma

Contributors: Sreya Menon, Vaishnavi J

Under the Guidance of Dr Anand Srinivasa Rao, Swetha Prakash

Reviewers: Dr Pratibha Narayanan, Uthara Narayanan

Proofreader: Sowmya Rajaram

Design support: Bhavna Panda

Attribution: Please cite this work as follows:

ISDM. 2026.

AI Systems for Social Good. A Practitioner's Guide for Leaders of Social Purpose Organisations

DOI: http://dx.doi.org/10.58178/263.1078

Disclaimer: *In preparing this manuscript, we used an AI tool (ChatGPT) to summarise and reword the content for SPOs, and we take full responsibility for the content.*

Table of Contents

Introduction

What Is This Guide About?

Let's think of this guide as a travel companion. Our organisations already carry knowledge in registers, surveys, WhatsApp chats, or community memory. AI simply helps us listen better to that knowledge, spot patterns earlier, and act with foresight.

AI is shaping everyday social-sector work, spotting patterns earlier, testing "what if" ideas, and helping all of us make more confident decisions. When done well, it amplifies our wisdom and stretches scarce resources. When done poorly, it risks eroding trust and deepening inequalities. The responsibility is ours, together.

AI is never only technical. Every system reflects our human choices: what we collect, what we ignore, who we hear, and who we miss.

For organisations, this means scoping and governance.
For Data Professionals, it means working with messier problems, small datasets, and ethical trade-offs where fairness matters as much as accuracy.
For communities, it means AI designed with consent, transparency, and accountability, strengthening dignity, inclusion, and voice.

Who Is This Guide For?

This guide is best used collaboratively, as many reflections benefit from perspectives across programme, data, and leadership roles. However, it is not necessary to have a fully formed team to begin. A Program Manager, Monitoring, Evaluation & Learning Lead, or Project Owner can initiate and facilitate the process, bringing in relevant stakeholders at different stages as needed. The structure is flexible, allowing organisations to start small and build participation over time.

What Will You Learn?

- How to **scope value**: clarify the decision bottlenecks that AI can meaningfully improve.
- How to **assess data readiness**: know what you have, what you need, and what permissions matter.
- How to **choose an approach**: prediction, generation, or sometimes no AI at all.
- How to **design for constraints**: low connectivity, low compute, minimal maintenance.
- How to **govern responsibly**: simple but strong checks for fairness, transparency, and privacy.
- How to **deploy and steward**: handovers, adoption, monitoring, evaluation, and graceful exits.

How to Use This Guide

You do **not** need to read this book cover to cover.

If you are **exploring AI for the first time,** focus on

Phase 1 | Phase 2

If you are **already working with data or AI,** focus on

Phase 3 | Phase 4

- Use the **Reflection Corners** to guide internal discussions
- Treat technical details as **optional reference**, not mandatory reading.

Certain examples have been cited throughout the book from various works on the ground to make the context relatable.

Ethics Preview

Using data and AI in social impact work requires careful attention to fairness, privacy, and transparency. In this guide, ethics runs as a thread through every phase. Later chapters show how these principles become part of daily practice.

Data Analytics and AI Phases Overview

To drive meaningful and measurable outcomes through data and AI, we'll move together through a structured **9-step Value Realisation Framework.** This journey is grouped into four core phases: from initial scoping to long-term stewardship, ensuring that our solutions are grounded in organisational needs and sustained through continuous improvement.

Value Scoping

1 **Business & Data Understanding**
Understand the business challenges; identify and source data, including actual and synthetic

2 **Solution Design**
Design the solution, select the analytic & AI methods suited for the application & requirements

Value Discovery

3 **Data Extraction**
Data preparation, including data selection, cleansing, extraction & imputation

4 **Pre-Processing**
Iterative feature selection and engineering to create final ML ready dataset

5 **Model Building**
Build and validate the solution with continuous testing

Value Delivery

6 **Model Deployment**
Publication of a trained model for broader use

7 **Transition & Execution**
Implementation into business process and workflows; evangelization

Value Stewardship

8 **Ongoing Monitoring**
Ongoing monitoring of outcomes for continuous observation & auditing

9 **Evaluation & Check-in**
Evaluation of insights & actions against business objectives

AI Guide Ready Reckoner

(A quick guide for social sector organisations)

This guide is designed to help social sector organisations think clearly and responsibly about using data and AI. You do not need technical expertise to use it.

If you are short on time, start here.

What Is This Guide?

✅ **It will help you:**

- Decide whether AI is appropriate for your work
- Ask the right questions before engaging data or AI partners
- Move from ideas to real-world use responsibly
- Avoid common risks and misuses

❌ **It will NOT:**

- Teach you how to code or build models
- Recommend specific tools or vendors
- Promise quick or magical solutions

The AI Journey at a Glance

This guide follows a four-phase journey, from early thinking to long-term stewardship.

Phase 1: Value Scoping

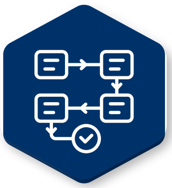
What problem are we trying to solve, and why does it matter? Are we solving the right problem?

- Clarify the real organisational challenge
- Understand what data exists and what does not
- Decide whether AI is even needed

Phase 2: Value Discovery

What kinds of questions can our data responsibly answer?

- Examine the quality and limits of available data
- Decide whether predictive or generative approaches are appropriate
- Identify risks, biases, and constraints early



Phase 3: Value Delivery

How will this actually be used by people, not just systems?

- Make solutions usable within real workflows
- Support adoption, training, and change
- Clarify ownership and accountability

Phase 4: Value Stewardship

How do we stay responsible as contexts and needs change? How do we sustain impact and trust over time?

- Monitor performance and outcomes
- Review ethical and social implications
- Adapt, scale, or stop solutions when needed

A Simple Way to Decide Where to Start

Use this quick guide:

- **Not sure whether AI is needed?** → Start with Phase 1
- **Have data but unclear on next steps?** → Go to Phase 2
- **Already building or piloting something?** → Focus on Phase 3
- **Running an AI-enabled system today?** → Pay attention to Phase 4

How to Engage with AI Partners or Vendors

You can use this guide even if you do not plan to build anything in-house. Before engaging external partners, you should be able to state clearly:

- the problem you are addressing,
- the role AI is expected to play,
- how success will be judged,
- and what uses are out of scope.

Using AI in the social sector is about making careful, values-driven choices that serve real people. This guide is meant to support that judgment, not replace it.

Phase 1
Value Scoping

Every meaningful AI journey begins with a pause. Value Scoping is where we slow down before grounding ourselves in the real problem, the people it affects, the opportunities ahead, and the risks we must avoid.

1

Think of this phase as sketching the map before a journey: Where are we? Who’s with us? Where might we go? And which paths should we avoid?

By the end of this phase, we will hold:

- ***A clearly scoped problem statement.***
- ***A stakeholder map.***
- ***A first-cut opportunity list.***
- ***A feasibility and ethics check.***
- ***Two canvases (Social Impact Canvas and AI Use Case & Ethics Canvas) to carry into Phase 2.***

Quick Reflection: *What decision in your work would become easier if you had better evidence for it?*

1.1: Understanding the Context & Problem Space

Without clarity on the real problem, we can easily end up building beautiful solutions to the wrong challenge. Together, let's explore the "why" before the "how." This anchors our intent in social value, so what we build truly matters to the people who matter most.

Example

In Tamil Nadu, livestock keepers often waited for government veterinarians, resulting in delays, animal deaths, and income losses. At the same time, SEWA had already documented 150+ herbal practices across languages, but this knowledge remained scattered and inaccessible due to language barriers and low digital literacy. The problem was not a lack of knowledge, but a lack of structured, timely access to it.

Reflection Corner

Use the questions below to reflect with your team or community partners:

1. What is the core challenge or need we want to address?

   ____

   ____

2. Why does this matter now? What will happen if nothing changes?

   ____

   ____

3. What systemic or historical reasons led to this issue?

   ____

   ____

4. Can the problem be solved with people, processes, or a simpler technological solution?

   ____

   ____

5. Why do we believe AI or data could help? What can it add?

   ____

   ____

Draft Your Problem Statement

- Write your problem[1] in **one sentence** (not longer than 15 words).
- Add **why it matters now** (1–2 lines).

1.2: Stakeholder & User Mapping

No model lives in isolation. AI touches all of us, communities, staff, funders, and regulators. Mapping helps us see who is central, who holds power, and who might be left out.

Reflection Corner

Reflect on the pointers below and create a simple map or list using this canvas

Primary Users
Who will use the solution directly?

Primary stakeholders
Who will be most positively or negatively affected?

Decision-Makers
Who approves, funds, or owns the solution?

Field Enablers
Who supports delivery (e.g. NGOs, local staff, volunteers)?

Data Actors
Who creates, owns, or is represented in the data?

1. **Formula to frame problem statement:** We need to decide **[decision]**, but currently **[what is missing or unclear]**. Using data could help us **[how it would improve the decision]**, so that **[who]** can **[act differently]**.

Activity

Build Your Stakeholder Map

Stakeholder Type	Name/ Role	Relevance to Project	Power/ Influence	Concerns or Needs
Primary Users				
Primary stakeholders				
Field Partner				
Data Owner				
Decision Maker				

Our map helps surface whose voice is missing, who holds power, and who we need to involve at each stage.

1.3: Opportunity Identification

AI can be tempting, but it is not always necessary. Here, we filter together where AI could meaningfully add value through speed, scale, or fairness.

Example

SEWA[2] had already trained 50,000 individuals and built a network of 1,200 livestock volunteers. The opportunity was to convert this structured, time-tested knowledge into a searchable AI system — allowing a farmer to ask a question in Tamil by voice and receive dosage guidance instantly, instead of waiting for a field worker.

Activity

Spot Opportunities

List 3 - 5 opportunities in your work where AI could help. Use the following lens:

- **Speed:** What tasks could be done faster with AI?
- **Scale:** What knowledge could reach more people?
- **Fairness and Effectiveness:** Where could AI reduce bias or exclusion?

2. Detailed report of SEWA can be found here

Reflection Corner

Discuss within your team and connect pain points with feasible AI support:

1. Which part of the problem could AI meaningfully address?
2. What would AI improve – speed, scale, targeting, prediction, insight?
3. What can't AI do in this context? What are the risks or limitations of AI?
4. Are there simpler (non-AI) interventions that should be tried first?
5. Can AI support or 'augment' existing human systems?

This helps teams focus on clear, narrow, high-leverage opportunities and avoid over-designing AI where it isn't needed

1.4: Feasibility Framing with Data & System Thinking

A great idea isn't enough; it also has to be doable. Together, let's run a quick reality check: do we have the data, permissions, people, and basic tech to make this idea work where it's needed?
Here we combine data readiness with system realities, considering factors such as infrastructure, organisational capacity, timing, and change readiness. This helps us set realistic boundaries before jumping into design.

Example

In Vivek's case, although the idea was promising, feasibility required checking: Were the 150 herbal practices already structured enough for AI ingestion? Could the LLM accurately retrieve context using RAG? Would voice recognition work reliably across Tamil dialects? Could farmers with basic Android phones access the web interface?

Reflection Corner: Are We Ready to Proceed?

Reflect on the guiding questions below and fill in the feasibility reflection table associated with the following:

Do we already have data[3] for this problem? If not, how will we collect it?	Is the data structured, reliable, clean, and regularly updated?
Is there a need for unstructured data, such as text, audio, video, or other formats? If so, do we have access to such unstructured data?	Do we have the digital systems or tools to host, process or act on AI?
Do we have (or can we access) people who can use or manage this solution?	Is now the right time? Do we have enough time, people, and money?

Activity

Feasibility Heat-Check

Score each domain from 1 (low) to 5 (high).

Domain	Our Readiness (1-5)	Notes or Gaps to Fix
Data		
Technology		
People & Skills		
Timeframe		
Organisational Support		

Use this tool to prioritise what needs strengthening before committing to building.

3. By data, we mean any information that an organisation already collects or observes in its work. This could include attendance records, primary stakeholder details, field notes, surveys, financial records, monitoring reports, or even informal tracking done by staff.

How to decide whether to move ahead

After completing the Feasibility Heat-Check, pause and look at your scores together.

- If **2 or more domains** score 1–2, do **not** move to model building yet.
  → Redesign the use case, simplify the ambition, or invest in readiness first.
- If **most domains** score 3, you may proceed **only as a small pilot**, with close monitoring.
- If **most domains** score 4–5, you are reasonably ready to explore modelling with a data partner.

This is not a pass/fail test. It is a **judgment call** designed to protect your organisation and communities from premature or fragile AI projects.

Non-negotiable Design rule
If your proposed solution cannot realistically survive low connectivity, low technical capacity, or limited maintenance, do not move forward yet.
Redesign the use case, simplify the ambition, or invest in readiness before proceeding.

1.5: Ethical Lens on Opportunity Framing

Trust is our licence to operate. Before we commit, let's look together at fairness, consent, privacy, and how people can question or override the system. Detailed governance structures will be defined later in Phase 4. At this stage, the goal is to surface risks early.

Decide on one red line[4] you won't cross and one plain sentence you'd say to a community member about why and how this will be used. When these are clear early, everything else is easier.

Example

Project Nclude was designed to support persons with disabilities in exploring careers and accessing government schemes. This raised critical ethical questions:

- *Would the AI give misleading career advice?*
- *Could it unintentionally reinforce bias about disability and employability?*
- *Would users over-trust the system and treat it as a final authority?*

To address this, the team grounded responses in 400+ verified lived career stories and scheme documents. They also monitored response accuracy (achieving ~95% groundedness) and introduced expectation-setting prompts so users understood the tool's limitations.

4. When we talk about problem framing or data and AI work, a red line usually means:

- A condition under which **we stop or do not proceed**
- A signal that says **"this is not appropriate"**, **"not yet"**, or **"not needed"**
- A boundary that protects **people, ethics, resources, or trust**

Reflection Corner

Think about each element in the table and list all risks or concerns for each. Also, discuss how to overcome those risks.

Ethical Theme	Risk or Concern	What Will You Do About It?
Inclusion: Who might be left out or harmed by this solution?		
Consent & Privacy: Are users aware of how their data is used? Do they have a real choice?		
Bias & Fairness: Could the data or model unintentionally favour or exclude groups?		
Transparency: Will the system's decision be understandable and explainable to its users? Will the models be interpretable?		
Responsibility: If something goes wrong, who is responsible?		

This step helps us all walk forward with clarity and conscience so our AI solutions uplift the communities we aim to serve.

1.6: Closing the Loop: Synthesising with the Canvases

By now, we've scoped the problem, mapped stakeholders, identified opportunities, tested feasibility, and applied an ethics lens. To tie this together, we use the two canvases:

1. **The AI Use Case & Ethics Canvas –** to articulate the purpose, data, risks, and safeguards. Download the canvas here and adjust accordingly
2. **The Social Impact Canvas –** to summarise the problem, outcomes, stakeholders, and resources. Download the canvas here and adjust accordingly

These canvases aren't homework, they're **conversation tools**. Let's bring our teams together, fill them out, and carry them forward into Phase 2.

Summary

We now hold the basics:

- A clear problem statement
- A feasibility heat-check
- A stakeholder map
- An ethics test
- A shortlist of opportunities
- Two canvases that capture your scope

AI journey starts with people, not just data.

Takeaway:

*Strong scoping prevents wasted effort, protects trust, and ensures our **AI journey starts with people, not just data**.*

*This grounding sets a strong foundation for success. You are now ready to move from **'Why?' to 'How'**, diving into the actual discovery of data, models, and potential solutions.*

Next up

Phase 2 - Value Discovery, where you begin shaping data into predictive and generative models.

Phase 2
Value Discovery

Our ideas become real when they meet data. In this phase, we move from scoping into discovery — shaping information into working prototypes that give us foresight and help us communicate better.

Think of this as moving from sketch to prototype. Our sketch in Phase 1 told us where to go. Now we begin building a working version of the tool.

By the end of this phase, we will:

- ***Frame a prediction question clearly.***
- ***Test whether predictive modelling is appropriate.***
- ***Identify a generative use case and validate outputs.***
- ***Understand trade-offs between accuracy, fairness, and trust.***

2

Role clarity
Phase 2 is about discovery and decision-making, not technical ownership.
SPOs are not expected to build models themselves.
The goal of this phase is to **frame the right questions, test assumptions with partners, and decide whether and how to proceed.**

2.1: Predictive Modelling: Framing the Right Questions Before Building Anything

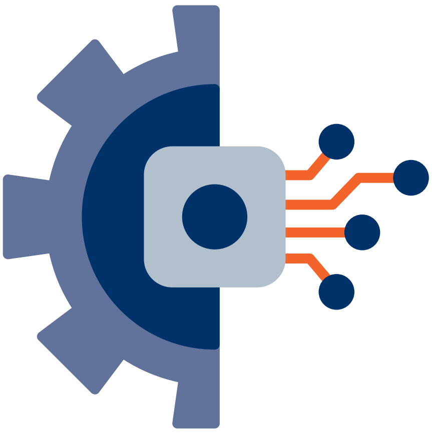
Predictive modelling[5] is often misunderstood as a technical exercise. In practice, it is a **decision exercise.**

Before any model is built, organisations must be clear about:

- what they want to anticipate,
- why they want to anticipate it,
- and how the prediction will be used in real decisions.

It helps organisations decide **whether predictive modelling is appropriate at all**, and if so, how to frame it responsibly.

What kind of organisational questions suit predictive modelling?



Predictive modelling is useful when an organisation wants to anticipate an outcome before it occurs, enabling early action.

Typical examples in the social sector include:

- Identifying Primary stakeholders at risk of dropping out, relapsing, or disengaging
- Anticipating service demand or resource shortages
- Flagging cases that may require early intervention

What must be true about your data for prediction to make sense?

Before considering predictive modelling, organisations should pause and assess their data against three conditions.

5. Readers who are interested in understanding how predictive models are technically built can refer to the Technical Appendix [Discovering value through predictive modelling](). This is not required to engage with the rest of the guide.

Do you have reliable past outcomes?
Predictive systems learn from history. If past outcomes are inconsistently recorded, politically influenced, or incomplete, predictions will reflect those distortions.

Are outcomes defined clearly and consistently?
If ‘success’, ‘risk’, or ‘failure’ mean different things across teams or time periods, predictions will be unstable and misleading.

Does the data reflect reality?
Many social-sector datasets capture access to services, not need or impact. Predicting from such data can reinforce existing inequities.

If any of these conditions are weak, predictive modelling should be approached with caution. In many cases, the right decision at this stage is to pause, simplify the use case, or choose a non-predictive approach.

What decisions will the prediction actually influence?

This is the most important question — and the one most often skipped.
Organisations should be able to state clearly:

- Who will see the prediction
- What decision will it inform
- What action will follow

For example:

- Will a prediction trigger outreach, additional support, or prioritisation?
- Will it be advisory, or will it override human judgment?
- What happens if the prediction is wrong?

If no concrete decision follows, predictive modelling risks becoming an academic exercise.

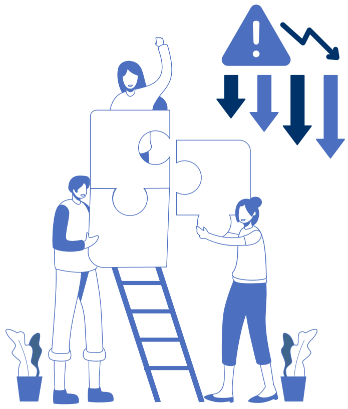
Understanding and accepting trade-offs

All predictions involve trade-offs. Organisations must consciously decide which kinds of errors are acceptable and which are not.
For instance:

- Is it worse to miss someone who needed help, or to flag someone who didn't?
- Are false alarms acceptable if they ensure no high-risk cases are missed?
- Should predictions be used differently for different groups?

These are value choices, not technical ones and must be made by programme leaders, not left to data teams alone.

Common pitfalls organisations should watch for



Predictive modelling often fails in the social sector due to:

- Treating predictions as neutral truths rather than probabilistic signals
- Ignoring structural biases embedded in historical data
- Using predictions to justify exclusion instead of support
- Expanding use beyond the original intent without review

Recognising these risks early allows organisations to design safeguards before harm occurs.

Reflection Corner

Before moving from discovery to model building, organisations should be able to answer the following clearly:

1. What outcome are we trying to predict, and why?
2. How is this outcome currently recorded?
3. What decision will this prediction support?
4. What are we not going to use this prediction for?
5. How will human judgment remain part of the process?
6. What risks might this prediction introduce for vulnerable groups?

If these questions cannot be answered confidently, the organisation is not yet ready for predictive modelling — and that is a valid outcome of Phase 2

Ethical check: Could any predictions cause harm if wrong? Who is impacted if the model makes a mistake?

Predictive models give us foresight. But foresight alone isn't enough. Together, we also need ways to communicate, explain, and act on those insights with staff and communities.

This is where Generative Models come in. If predictive models are your early-warning bells, generative models are your translators and scribes, helping you draft, explain, and share knowledge in ways people can understand and trust. Together, they form two complementary tools: one looks ahead, the other gives voice.

2.2: Working with Generative Models

Generative models don't predict outcomes; they create and communicate. They can help us draft reports, translate documents, summarise long surveys, or answer FAQs in local languages. Done carefully, they free our staff from repetitive tasks and expand access for our communities.

Why it matters

For us in SPOs, this means less time writing and more time engaging. For Data Professionals, it's about validating outputs for grounding, tone, and fairness. Generative models are powerful but prone to "hallucinations"; they need guardrails.

Example

The Project Nclude[6] did not build a new AI model. Instead, they connected an existing AI system to 400+ verified career stories from people with disabilities and to updated government scheme data. This meant when a blind student asked about engineering careers, the chatbot referenced real success stories — not generic internet advice. Early versions gave occasional incorrect answers, so the team refined how the system retrieved information to improve reliability.

Key Concepts Made Simple

1. Model Selection

Most SPOs will not build generative models themselves. In practice, they begin with prompt design or document-based approaches (such as retrieving answers from trusted internal or public documents), and work with Data Professionals only if deeper model adaptation is required.

Ask yourself:

Who is the audience? Do you need low-cost, multilingual ccess or privacy control?

6. Details of Nclude project can be found heranswerancwere

2. Model Adaptation

Although a model is pre-trained, it can be adapted in different ways:

- **Prompt Engineering**: guiding a pre-built model using clear instructions.
- **Document-based generation (RAG model)**: restricting answers to trusted documents so outputs stay grounded

Most SPOs use prompt design or document-based approaches, depending on their needs and constraints.

3. Model Validation

With generative AI, accuracy alone isn't enough. We need to test together: relevance, grounding, tone, and hallucination.

- **Relevance:** Does it answer the user's actual question?
- **Grounding:** Is the answer based on trusted information?
- **Tone:** Is it respectful and context-aware?
- **Hallucination:** Is it making things up?

Teams can use simple checklists or pilot conversations to test these. Always review responses before deploying at scale.

Quick win

Take one paragraph from a recent report. Run it through a generative model and ask for a plain-language summary. Then read both versions aloud to your team. In 10 minutes, you'll surface what "clarity" means for your community.

Activity

Reviewing AI Outputs with Your Team

Use this checklist after testing a generative model:

Validation Question	Pass / Fail	Notes
Is the answer relevant to the user's question?		
Is it grounded in your trusted sources?		
Is the tone respectful and context-aware?		
Any made-up or false content (hallucination)?		
Organisational Support		

Reflect & Try It Yourself

1. What's one repetitive task you'd hand over to AI tomorrow?

2. What trusted documents or knowledge sources would you want the AI to use?

3. Who will use the model, and what risks should you plan for?

4. How will you ensure answers are not only correct but respectful?

Summary

In this phase, we:

- Framed a prediction question
- Clarified requirements for a predictive use case
- Identified a generative use case and reviewed outputs

Trust and fairness guide useful AI

Takeaway:

*Predictive models help us **look ahead**; generative models help us **communicate and create**. Both are valuable, but only when grounded in trust, fairness, and clear purpose.*

Next up

Phase 3 - Value Delivery, where your model moves from lab to field and becomes part of real workflows.

Phase 3
Value Delivery

We've built a prototype. Now comes the question: can it live outside the lab?

Value Delivery is about embedding a model into real systems and workflows — ensuring it is not only technically sound but also practically usable and trusted.

Think of this as moving from **prototype to practice**. In Phase 2, we experimented; here, we make the model useful in day-to-day work.

3

By the end of this phase, we will:

- ***Plan for deployment.***
- ***Integrate the model into existing workflows.***
- ***Support trust, adoption, and responsible use.***
- ***Set up transition and ownership for maintenance.***

Role clarity

SPOs are not expected to manage technical deployment themselves.
The role of SPOs in this phase is to make informed choices, support adoption, and ensure accountability, working closely with Data Professionals.

Model Deployment[7]

We've all seen models stay stuck in notebooks. The real challenge is deployment: getting the model into places where people can actually use it inside an app, on a dashboard, or through a chatbot.

As you plan deployment, run a dry-run discussion with your team. Walk through where the model should "live" and who will actually touch it day to day. You'll find that people closest to the workflow often surface the most practical insights about what will or won't work in reality.

Why it matters

For us in SPOs, deployment means our staff and communities see value in real time, not just reports. For Data Professionals, this is the bridge between coding and actual service delivery. A model that isn't deployed is a story half-told. In practice, this means:

- The model is accessible and embedded in daily work.
- Users understand outputs and can get simple explanations.
- The system can flag failures or performance drift.
- Teams know when to review, retrain, or retire the model.

This is the phase where your AI tool becomes a reality, a valuable asset, and an integral part of the team.

Example

In the E-QLT project[8], a model was built to estimate household social security scores. But that was just the beginning. The team then had to:

- *Package the model so it can run reliably on the NGO's systems, with support from Data Professionals.*
- *Decide on batch prediction and run it once a day to update the dashboards.*
- *Integrate those outputs into the NGO's portal so field workers can actually use the scores during visits.*
- *Set up a monthly review where users sample predictions and flag anything unusual.*
- *Develop a retraining plan based on feedback and updates to ground data.*

This ensured the model wasn't just built but also used, maintained, and improved over time.

7. In this guide, deployment refers to making a solution usable in real organisational contexts, not technical infrastructure alone.
8. Detail report of EQLT can be found here

Example

In SEWA's AI Vetbot project, deployment meant making the chatbot accessible through a web app and WhatsApp integration so livestock keepers could use it easily

The team also incorporated speech-to-text tools to support low-literacy users and hosted the system on cloud infrastructure. Early user testing helped refine language accuracy and usability.

Choosing the Right Interface

Users do not need to be technical experts; we just need to choose a "home" for the model that fits their daily rhythm and local constraints, such as connectivity. The data solution can be delivered through different interfaces depending on user needs.

Mobile Apps: Best suited for field workers in remote areas as they support offline data entry and frequent use.

Web Dashboards: Ideal for program managers and NGO leaders who require complex data visualisation and back-office analysis.

WhatsApp or IVR: Best suited for communities with low digital literacy due to support for voice-based interactions and large-scale reach.

Field Tip:

- Don't build a new app if your team already spends 80% of their time on WhatsApp or an existing Management Information System (MIS). Embed the AI where the work already happens.
- If a solution cannot function reliably under real field constraints (connectivity, devices, staff capacity), it must be simplified or redesigned.

Key Concepts Made Simple

Question You Might Ask	What This Means	How It Helps You
Where should the model live?	Together, we decide whether it will run on the cloud, a local server, or directly on staff devices.	Makes the model accessible to your users
How will it run?	We choose batch mode (e.g., nightly updates) or real-time responses. Balance speed, cost, and simplicity.	Balances speed, cost, and simplicity
How do we know it's working?	We need to set up simple checks and reviews; Track accuracy, adoption, and fairness in practice.	It helps catch errors before they grow
How will users trust it?	We provide short guides, role-based training, and ways to explain predictions simply; Design for privacy and safety from the start	Builds confidence and reduces misuse
Who will maintain it?	We need to nominate a “model steward”, someone accountable for monitoring, retraining, and responding to issues	Ensures it stays alive after the project ends

Transition & Execution

Deployment is just the start. Transition means embedding our tool into existing workflows, training staff, and ensuring a smooth handover from the pilot team to long-term owners.

Many of our projects risk failing here: the model works in pilots but struggles in practice when staff don’t use it, trust it, or see value.

Together, let's ask:

- Who needs training, and in what format?
- How will you reassure users that AI is a support, not a threat?
- Who owns the tool once the pilot team steps back?

Example

***Nclude:** Before scaling, over 500 users tested early versions and reported issues with speed and clarity. The team introduced guided questions and integrated the chatbot into WhatsApp groups already used by disability organisations. This helped grow to 2,000+ users without replacing existing counsellors*

Activity

Transition Checklist

Transition Element	Completed (Y/N)	Notes
Training sessions for staff		
User manuals / quick guides		
Feedback channels created		
Governance owner assigned		
Maintenance team onboarded		

Summary

We've now moved from prototype to practice.

Deployment → choosing the right platform and interface.

Transition → we built workflows, training, and ownership so it can last.

Good model: used, trusted and sustained

Takeaway:

A good model is not just one that works technically, but one that is used, trusted, and sustained in real life

Next up

Phase 4 - Value Stewardship, where you ensure the model stays ethical, fair, and effective over time.

Phase 4
Value Stewardship

Building and deploying a model is only half the journey. The real test comes after launch: does it stay fair, accurate, and trusted as the world around it changes?

Value Stewardship is about **responsibility**, setting up systems to monitor, evaluate, and adapt. In social impact, AI must not only work on day one, but remain reliable, ethical, and transparent as contexts change.

Think of this as moving from use to care: just like a clinic doesn't stop after opening day, our model needs ongoing care to keep serving communities well.

By the end of this phase, we will learn to:

- ***Monitor and evaluate your model regularly.***
- ***Detect and respond to model drift over time.***
- ***Engage communities in oversight.***
- ***Build systems of accountability and trust.***

4

Role clarity
In this phase, SPOs are responsible for stewardship decisions and accountability.
Data Professionals may support monitoring and retraining, but responsibility for impact, fairness, and use always remains with the organisation.

Why it matters

Without stewardship, even good models can silently cause harm: a prediction that worked last year may fail today because patterns shifted; a chatbot that served one group well may unintentionally exclude another. Stewardship is how we ensure our AI continues to create value, stays aligned with purpose, and earns trust over time.

This phase is how SPOs turn one-time projects into long-term value.

Activity

Draft a Governance Charter

This charter should be no longer than one page and written in plain language. A one-page “promise to users,” covering:

- **Purpose**: what the AI is (and isn’t) for.
- **Consent**: how users give or withdraw consent.
- **Transparency**: how outputs are explained in plain language.
- **Oversight**: who handles complaints and monitors risks.
- **Accountability**: Who is responsible if something goes wrong?

The Stewardship Cycle

Stewardship is not a one-time act but our shared rhythm:

Monitor
Track performance, fairness, and adoption in real time.

Review
Compare outputs with ground reality; is the model still solving the right problem?

Retrain
We use fresh data and insights to refresh our model, staying mindful of new risks, biases, or shifts in the system.

Feedback
Create loops with staff, communities, and partners to flag gaps or harms.

Use these four steps as a quarterly check-in exercise with your team: ask together what's drifting, what still feels relevant, and what needs retraining. You'll discover that the act of reviewing as a group often strengthens trust more than the technical fixes themselves.

Quick win

Pick one model you've used. In 10 minutes, answer together: "What has changed since we first built this?" This quick check often reveals drift before the numbers do.

Example

In Vivek's example, after deployment, stewardship meant:

- *Tracking how many farmers were actively using the chatbot each quarter.*
- *Monitoring whether AI responses were accurate and culturally appropriate.*
- *Expanding the database from 150 to 500 practices across more languages.*
- *Periodically retraining the model as new herbal practices were added.*
- *Reviewing whether antibiotic use actually reduced in target communities.*

Activity

Model Health Check

Use this table every quarter with your team:

Area	Question	Status	Notes / Action
Accuracy	Are predictions still correct compared to reality?		
Relevance	Is the model still solving the original problem?		
Fairness	Is any group being excluded or disadvantaged?		
Adoption	Are staff and users actively engaging with it?		
Governance	Do users know their rights (consent, grievance)?		

Activity

Annual Evaluation Canvas

Once a year, step back with your team and ask:

Question	Notes	Decision
Does the model still align with our mission?		Continue / Redesign / Retire
Has the context or user group changed?		Continue / Redesign / Retire
Is the model delivering real outcomes for people?		Continue / Redesign / Retire
Are risks (bias, misuse) under control?		Continue / Redesign / Retire
Is the cost of upkeep worth the impact?		Continue / Redesign / Retire

Reflection Corner

1. What 2-3 signals will tell you if your model is drifting?

____

____

2. Who in your organisation is responsible for stewardship?

____

____

3. When would you retire a model rather than maintain it?

____

____

4. How will you explain your AI tool in plain language to users?

____

____

Summary

Value stewardship does not happen in isolation

To monitor models, respond to drift, and act responsibly, organisations need internal capabilities — clear ownership, trusted governance, and people who know how to work together across programme, data, and field realities.

The next section focuses on these ***organisational capabilities****: how SPOs build trust, roles, and ways of working that enable AI systems to remain ethical, effective, and useful over time.*

Organisational Capabilities for Responsible AI

These capabilities cut across all four phases and determine whether AI systems are trusted, sustained, and responsibly governed over time.

Trust Management

This section focuses on governance, accountability, and safeguards that allow AI systems to be used with confidence and care.

Trust is an organisational capability, not a feature of the model. In the social sector, trust determines whether an AI system is used, questioned, or quietly ignored. Trust Management therefore focuses on the structures, checks, and habits that help organisations earn and maintain the confidence of staff, partners, and communities throughout the AI lifecycle.

Why it matters

Without trust, adoption falters. SPOs need confidence that their systems are fair and transparent, and communities need reassurance that they remain safe, included, and respected.

Example

In Nclude, many users initially distrusted AI-based guidance, especially when it related to sensitive decisions like career pathways. Others expected the chatbot to behave like a human counsellor.
Instead of positioning the system as a replacement for career facilitators, it was introduced as a support tool within disability organisations and WhatsApp groups already trusted by users. Early pilots with 500+ users helped refine clarity, tone, and reliability before scaling. The system clarified what it could and could not do, reducing overconfidence and building gradual trust.

SEWA's chatbot provided herbal treatment guidance, but the team was careful not to position it as a replacement for veterinarians. It was framed as a support tool that complemented existing livestock volunteers and field workers. By integrating it into existing WhatsApp groups and training networks, trust among farmers gradually grew.

Key Concepts Made Simple

Concept	What It Means	How to Apply It in SPO Work
Three Lines of Defence	A way to clarify who builds, who reviews, and who is accountable.	**First line:** The team or partners building and using the AI system **Second line:** Internal reviewers or leaders who question outputs, risks, and fairness **Third line:** Periodic external or independent review (advisors, funders, auditors)
End-to-End Governance	Trust is not only managed during deployment but across the full lifecycle from strategy to retirement.	Build in feedback loops, sunset reviews, and clear ownership
AI Risk Domains	Common areas where trust may be broken (Aligned with frameworks such as the NIST's AI Risk Management Framework)	E.g., Bias, privacy, misinformation, over-reliance, or lack of transparency
Causal Sources of Risk	Helps us understand where risks come from	Human error? Model misuse? Deployment conditions? Map them clearly
Transparency vs Interpretability	Transparency = How open the system is. Interpretability = How understandable it is to humans	Strive for both, especially for frontline users and affected communities

In AI systems, trust is not a feature added at the end. It is built through transparency, responsibility, monitoring, and consistent human oversight.

Trust Health Checklist

Trust factors	What it means	How to apply it in SPOs
Fairness & Bias	Is any group left out or disadvantaged?	Check whether your data accurately reflects the community's diversity. Talk to people affected by the AI system to understand the unintended impact.
Diversity	Are frontline workers, Primary stakeholders, and partners included in design and review?	Include voices from the field, frontline workers, and affected groups while designing or reviewing the model.
Transparency & Interpretability	Can staff and users understand how the model works, in plain language?	Keep a simple record of how the model was built, what data it uses, and who is responsible. Share this with partners and users.
Consent & Privacy	Do people know and agree to how their data is used?	Collect only what's necessary. Avoid using data that could expose individuals without their consent.
Safety & Security	Protecting the system from harm, both technical (hacking) and social (misuse)	Keep data access limited, store it securely, and consider how the system could be misused or misunderstood.
Reliability & Robustness	Does the model hold up across conditions, or does it break down when the data change?	Test your model in different field conditions. Ask: Does it still work if data quality changes? What happens if some inputs are missing?
Resilience	The system can recover from failure and adapt over time	Build in simple feedback loops. Can the model be updated if it starts behaving poorly or the context changes?
Human Fallback	Is there always a route back to a human decision?	
Right to Contest & Appeal	Users can question or challenge AI outputs	Create a simple way for staff or communities to flag "this feels wrong" and have it reviewed by a human
Accountability	Is it clear who is responsible when something goes wrong?	Name an internal owner for decisions, redress, and escalation.

If one of these begins to crack, trust erodes quietly. Catch it early.

Governance & Risk in Practice

This section translates trust principles into concrete governance actions that organisations can actually run.
SPOs face real risks when using AI, including bias and misinformation, privacy breaches, and over-reliance. But identifying the risk is only the first step.

The **NIST AI Risk Management Framework (AI RMF 1.0)** offers a simple, practical structure to guide your next steps. It encourages teams to:

- **Map the risks:** Who might be harmed? How?
- **Measure them:** How likely, how severe, how widespread?
- **Manage them:** What controls or actions reduce harm?
- **Govern them:** Who is responsible for oversight?

It helps SPOs move from concern to confidence by providing a clear path to steward AI responsibly throughout its lifecycle.

Watch For

These are signals that trust may be eroding quietly, even if no one says it aloud:

- AI recommendations that no one questions anymore
- Field teams saying, "I don't know why it flagged this"
- Tech teams assume the model is fine unless errors are reported
- Primary stakeholders showing hesitance or disengagement
- Processes that feel opaque or overly automated

Reflection Corner

1. Where in our system could trust be easily broken, and for whom?

____

____

2. Who plays our "second line of defence," and do they have real authority?

____

____

3. How would you explain this tool to your grandmother in two sentences?

____

____

You may use a simple risk-mapping table (see Appendix) to document and review these risks.

What Comes Next

Trust isn't a one-off activity; it's a discipline. By checking for fairness, transparency, and accountability across the lifecycle, SPOs can ensure their AI systems don't just work technically but are used with confidence, dignity, and care.

Now that we've thought deeply about trust, it's time to focus on the people who carry that trust forward. In the next step, we turn to **Talent Management**, building the right team capacities, structures, and values to sustain ethical, reliable AI systems. Because no matter how strong a model is, it is the people behind it who shape its impact.

Once trust structures are in place, the next question is whether the organisation has the people and ways of working to carry them forward.

Talent & Process Management

Great models fail without the right people and rhythms. Talent & Process Management focuses on the minimum viable team, clear ways of working, and crisp handovers so models survive beyond pilots.

Why it matters

SPOs often run lean. Data professionals can be hard to hire or retain. The answer isn't always "build a large in-house data team"; it's designing a sensible operating model: who does what, how decisions are made, what "done" means, and how you keep momentum without burning people out.

Example

In the MAAP project on malnutrition, the AI team included developers, public health experts, and government partners. Asha and Anganwadi workers played a significant role in data collection and feedback, even though they were not "AI experts." This made the tool more usable and ensured early-stage buy-in from frontline workers.

Minimum viable team design

- Trust A program or domain lead (problem ownership and decision-making)
- A data/technical partner (model design and implementation)
- A community or ethics advisor (fairness, consent, and ground reality)

These roles may sit with individuals, teams, or partner organisations, but the relevant party must explicitly own them. Titles matter less than clarity on who builds, who reviews, and who is accountable.

Reflection Corner

1. What roles are essential now, and later?
2. Which roles can be internal vs partnerships?
3. How will you promote collaboration between the field, tech, and leadership?
4. How will ethical concerns be surfaced and acted upon?

Ways of Working in Practice

Why it matters

SPOs are used to complexity, ambiguity, and change. Your AI process needs to work in that environment. That means working iteratively within structured phases toward **agile, iterative cycles** that can learn from the field and respond to ethical concerns early.

Example

IIn the EQLT simulation project, development was iterative. The team adapted the model based on feedback from civil society actors, not just technical metrics. The process wasn't linear; it circled between prototyping, simulation, feedback, and learning. This kept the tool grounded in real-world needs.

Nclude's development process relied heavily on user testing. Over 500 users tested early versions and provided structured feedback. The team iteratively improved speech models, multilingual support, and response grounding across quarters. Expert advisors also guided the prioritisation of key use cases over spreading efforts too thin

Key Concepts Made Simple

Concept	What It Means for SPOs
Agile + AI	Combine fast iterations (Agile) with field insights—plan for versioning, testing, and course correction.
Document everything	AI systems are complex. Keep lightweight records of decisions, changes, roles, and assumptions.
Process owner	Appoint someone to hold together the tech, ethics, and program goals, a “translator” across domains.
Start early	Don’t wait for deployment to think about governance or handover. Process discipline begins at the idea stage.
Handover planning	Decide early who owns the system after pilots, and how knowledge is transferred

Reflection Corner

1. Are your AI and program teams in sync on timelines and iteration cycles?

2. Do you have a “process bridge” person who understands both ethics and engineering?

3. Could you run a short simulation or dry run to test your AI process before scaling?

The Journey Ahead: You, Your Team, and AI for Social Good

We've made it through the guide. But this is not the end, it's the beginning of a deeper journey.

Building AI systems in the social sector is not about speed or scale alone. It's about stewardship. We are not just launching models, we are shaping decisions that touch lives, redistribute power, and influence how justice and dignity show up in everyday systems.

What It Takes

- **Clarity:** A grounded problem that matters to people on the margins
- **Care:** A commitment to fairness, inclusion, and non-harm
- **Collaboration:** Teams that bridge field, data, design, and leadership
- **Courage:** To start small, ask uncomfortable questions, and change course if needed

Even a pilot that maps patterns or makes simple predictions can help shift decision-making.

- Use this guide to track your assumptions
- Revisit reflection tables every few months
- Update your models as context shifts
- Talk to users. Talk to each other.

Closing note

This guide is a guide, not a guarantee. Its value lies in how honestly it is used, questioned, and adapted. If it helps your organisation pause before building, listen before deploying, and care after launch, it has done its job.

Please share your valuable feedback.

Appendices and Supplemental Sections

Appendix A: Walkthrough Case Study:

A1. Janhit Collective – AI for Women's Health Access

Janhit Collective is a grassroots organisation that supports low-income women in rural districts to access maternal healthcare. Their field teams noticed a worrying pattern: many women begin antenatal care, but nearly one in three drop out after just one or two visits. This puts both the mother and baby at serious risk.

They asked, 'Can we use our data to predict who might drop out early?' And if so, could that help ASHA[9] workers prioritise their follow-ups?

Value Scoping

They began by listening. ASHAs described their long lists of women to follow up with and how little time they had to reach everyone. Programme managers shared past attempts to identify high-risk cases. Everyone agreed that a tool that could quietly support better targeting would be helpful, provided it was done carefully.

The team gathered available data: paper registers, mobile app logs, and ASHA notes. It was patchy but promising. They set a clear, shared goal: reduce antenatal care dropout from 30% to under 15% in one year.

They mapped what this tool would look like, not a dashboard, not a prediction shown to the woman, but a quiet flag embedded in the ASHA's daily list. It would gently suggest: "Pay special attention here." If a woman is flagged, it might prompt the ASHA to check in more personally, offer a home visit, or emphasise the importance of antenatal care, especially for those facing barriers such as long travel distances or irregular support. The aim was not to automate decisions, but to help frontline workers act with more insight and care.
They also surfaced ethical questions from the start. What if we misclassify someone? Will ASHAs trust the model? Will women feel labelled? These concerns shaped the design from day one.

Value Discovery

With help from a local data volunteer, they extracted three months of data across three systems. They built in consent checks and anonymisation at the start, making sure personal identifiers were kept separate from the modelling process.

9. ASHA workers (Accredited Social Health Activists) are community-based frontline health workers who support maternal and child healthcare in rural India.

The data was messy: inconsistent names, missing dates, reused phone numbers. Data cleansing became a project of its own, sometimes frustrating, often illuminating.

Two models were tested. The team chose the one that was easiest to explain, even though it was not the most complex. Programme staff could trace how decisions were made and question them. This clarity mattered more than marginal gains in accuracy.

Before moving forward, they held a joint review session. ASHAs, supervisors, and programme managers tested outputs with real examples. Some surprises emerged. They revised feature weightings and updated labels before deployment.

Value Delivery

The model was packaged into a lightweight file and integrated into the ASHA mobile app. It ran in the background each night, quietly refreshing risk scores.

This made adoption smoother and reduced resistance from frontline staff. The tool simply reordered the ASHA's task list, with women at higher risk appearing near the top. A small icon marked these cases, but it could be ignored.

Training sessions were held in groups of 6 to 8. The team used roleplay to simulate scenarios: What if the model flagged someone you know is doing well? What if it missed someone you're worried about?

Significantly, they added a simple feedback form inside the app: "Was this flag helpful?" ASHAs could tap once to agree, disagree or leave a voice note. This feedback loop became a core part of the delivery.

Value Stewardship

Monitoring began with a light but regular rhythm. The data team sampled 30 flagged and unflagged cases every quarter. Supervisors reviewed follow-up visits and compared them to predictions.

One quarter, accuracy dropped. A local maternity scheme had quietly ended. The dropout pattern changed, but the model hadn't caught up. They paused and retrained.

At the end of the year, the entire team, from ASHAs to state-level partners, reviewed the project. The dropout rate had reduced to 14%. However, more importantly, ASHAs reported greater clarity on where to focus their time and energy.

Together, they decided to retrain the model annually and add new features, such as distance from the health centre and family support indicators. They also recommended that future tools be developed with ASHAs involved from the outset.

What this case shows

- AI can support frontline workers without replacing their judgement
- Simpler, explainable models often work better in social contexts
- Trust grows when users can question and override AI outputs
- Feedback loops are as important as accuracy
- Stewardship is about adapting to real-world change, not fixing models once

Appendix B: Core Concept Reference

B1. What Is Data Analytics and AI?

Data analytics and AI are mathematical or computational representation that helps you understand patterns, make predictions, or generate new data from existing information. There are several types of data analytics and AI we can draw from, each serving different purposes.

- **Descriptive Models:** help you see what's happening now (patterns, trends), for example, grouping Primary stakeholders by demographics.
- **Diagnostic Models:** help you understand why something happened, such as analysing factors behind low attendance.
- **Predictive Models:** act as early-warning bells, signalling who or what may need support. For example, predicting which students are likely to drop out.
- **Prescriptive Models:** act as guides, recommending the next best actions, such as optimising resource allocation.
- **Generative Models:** are translators and scribes, creating text, images, or answers
- **Adaptive Models:** learners improve themselves as new data arrives.

While each of these models has its place, in this guide, we focus on predictive and generative models as powerful tools for forecasting future events and creating new content or data capabilities that are especially impactful in the social sector.

B2. Predictive Models

Machine Learning (ML) is used to identify patterns in past data and make predictions. Machine Learning is of three different types.

1. **Types of Machine Learning**
    a. **Supervised Learning:** You have input data and a set of labelled output data (e.g., "Did the household income recover? Yes or No - a binary classification of an output variable"). Based on this labelled training data, the model learns to predict the household income recovery variable.
    b. **Unsupervised Learning:** You only have input data and want to find patterns or groupings (e.g., clustering households into similar livelihood profiles).
    c. **Reinforcement Learning:** The system learns by interacting with the outside environment and obtaining feedback in the form of rewards and penalties.

*This section focuses on **supervised learning** using structured data.*

2. **Feature engineering (what Data Professionals mean by it)**

Features are the inputs a model uses to make predictions. In practice, this means deciding what information matters and what does not.

Domain knowledge is critical here. Frontline teams often know which variables reflect real-world behaviour and which are misleading.

SPOs play an important role by questioning features that feel unfair, irrelevant, or hard to explain.

3. **Types of Predictive Models**

Here are some model types commonly used in social data:

Model Type	What it does	When it's useful	Example
Linear Regression	Finds a pattern between numbers and draws a simple upward or downward trend.	You want to predict a continuous outcome like expected income or test score	Predicting household income recovery based on past income, occupation, and family size
Logistic Regression	Predicts a Yes/No answer based on input data	You're predicting binary outcomes, like loan default or school dropout	Predicting which students are likely to drop out of a skilling programme
Decision Tree	Creates a series of if-then rules based on data	You want transparency, easy to explain to programme teams or stakeholders	Mapping out decision rules for prioritising follow-up visits based on Primary stakeholders' risk levels
Random Forest	Combines many decision trees to improve accuracy	You want better accuracy and have enough data, even if you lose some transparency	Predicting loan repayment success using many small decision paths from multiple variables

SPOs do not choose models on their own. This overview is meant only to help you understand trade-offs when discussing options with data partners.

B3. Model Evaluation & Validation

When a model makes predictions, it will sometimes be wrong. What matters is how it is wrong. Missing a real problem can cause harm (e.g., a student dropping out unnoticed).

False alarms can waste time and resources (e.g., unnecessary follow-ups). Different projects tolerate different kinds of mistakes. Programme teams must decide which errors are acceptable. This is a value judgement, not a technical one.

How teams usually start

- A typical modelling cycle (simplified)
- Prepare and clean data
- Decide what to predict and why
- Train a simple model
- Test it on new examples
- Review errors and risks together

Keep in Mind:

1. Over time, our models may become less accurate due to data drift (when input values change) or concept drift (when the definition of "success" changes). That's why we need to review and refresh them regularly (refer to chapter Stewardship cycle)
2. We need to be cautious of overfitting when a model becomes overly accurate on current data but fails to generalise to new data. That's why testing and validation matter for all of us.

Appendix C: Risk, Ethics & Compliance (Reference)

C1. Risk Mapping Table

Risk	Who is affected	Severity of impact	When it may occur	Frequency of occurrences	Who is responsible	How to detect it early

C2. Overview of Data Privacy & Compliance

Key Regulations:
These regulations shape how SPOs must think about consent, data use, and accountability, even when working with external Data Professionals.

- **GDPR (EU):** Protects personal data, requires user consent, data minimisation, and breach notifications.
- **HIPAA (US):** Governs the privacy and security of health data.
- **India's Digital Personal Data Protection Act (DPDP Act):** Proposed data protection law emphasising user rights and data localisation.

Best Practices:
- Collect minimal personal data necessary.
- Obtain informed user consent explicitly.
- Anonymise or pseudonymise data wherever possible.
- Secure data using encryption and access controls.
- Maintain audit trails and ensure transparency in data use.

C3. AI Ethics Frameworks

Principles to Uphold:

- **Fairness:** Avoid bias and discrimination.
- **Transparency:** Explain AI decisions clearly.
- **Accountability:** Define who is responsible for the outcomes.
- **Privacy:** Protect user data vigilantly.
- **Inclusivity:** Ensure AI benefits diverse populations.
- **Sustainability:** Consider long-term social and environmental impact.

Well-known AI ethics frameworks (for reference)

- OECD AI Principles
- IEEE Ethically Aligned Design
- EU Ethics Guidelines for Trustworthy AI

SPOs are not expected to adopt these frameworks wholesale. They serve as reference points for shaping internal principles and discussions.

Appendix D: Canvases & Practical Tools

Appendix D1: Core Canvases (For SPO Teams)

Social Impact Canvas

Print: 1 per organisation
Page size: 20 x 11.25”
Download ppt (editable): Link

AI Use Case & Ethics Canvas

Print: 1 per organisation
Page size: 20 x 11.25”
Download ppt (editable): Link

Appendix D2: Advanced / Optional Canvases

Machine Learning Canvas (Only for Tech partners)

Print: 1 per organisation
Page size: 20 x 11.25”
Download ppt (editable): Link

AI ROI (for leadership and funder discussions)

Print: 1 per organisation
Page size: 20 x 11.25”
Download ppt (editable): Link

Appendix E: Learning & Reference Resources

E1. Recommended Resources

Books:

- Artificial Intelligence: A Guide for Thinking Humans – Melanie Mitchell
- Prediction Machines – Ajay Agrawal, Joshua Gans, Avi Goldfarb
- Weapons of Math Destruction – Cathy O'Neil

Online Courses:

- Coursera: AI For Everyone – Andrew Ng
- Fast.ai: Practical Deep Learning for Coders (for Data Professionals or fellows)
- Hugging Face Course: Transformer Models (advanced, technical reference)

Tools & Platforms:

- Python libraries: scikit-learn, TensorFlow, PyTorch
- No-code: Google AutoML, Microsoft Azure ML Studio
- Model Hosting: AWS SageMaker, Google Cloud AI Platform

Note: Specific tools and platforms change rapidly and are best chosen by Data Professionals. SPOs should focus on problem clarity, data readiness, and governance rather than tool selection.

Appendix F: Glossary & FAQs (Quick Support)

F1. Glossary of Key Terms

This glossary provides definitions of key terms used throughout the guide to help you quickly understand technical jargon and domain-specific language.

AI and Machine Learning Term

- **Artificial Intelligence (AI):** Technology that enables machines to simulate human intelligence, including learning and decision-making.
- **Bias:** Systematic error or unfairness in data or model predictions that can lead to discrimination.
- **Data Drift:** Changes over time in the data's statistical properties, which can reduce model performance.
- **Generative Model:** A model that can create new data similar to the input data (e.g., text generation).
- **Predictive Model:** A model designed to forecast outcomes based on input data.
- **Explainability:** Ability to understand how AI models make decisions.
- **RAG (Retrieval-Augmented Generation):** Combining AI generation with external data retrieval for accuracy.
- **Ethical AI:** AI designed and used with fairness, transparency and responsibility

Social Sector & Project Terms

- **Primary stakeholders:** The individuals or communities who benefit from a programme or intervention.
- **Ethical AI:** The practice of designing and deploying AI systems that are fair, transparent, and accountable.
- **Impact Evaluation:** Assessing the changes that can be attributed to a programme or intervention.
- **Stakeholders:** Individuals or groups with an interest or role in a project's success.
- **Use Case:** A specific problem or scenario where AI or machine learning can be applied to add value.
- **Data Privacy:** Protection of personal or sensitive information from unauthorised access or disclosure.
- **Frontline Workers:** Staff or volunteers who interact directly with communities.
- **Programme Theory / Theory of Change:** A framework explaining how activities lead to intended outcomes.
- **Governance:** Structures and processes that ensure accountability and responsible decision-making.
- **Model Drift:** When a model's performance degrades because conditions or data patterns change.

F2. Frequently Asked Questions (FAQs)

These FAQs address questions that often remain after completing the guide.

A. Understanding the Basics

Q: Do I need to be a tech expert to use this guide?
Not at all. This guide is designed for SPOs like you, grounded in social purpose, not technical degrees. What you need is intent, clarity of purpose, and a willingness to learn collaboratively.

Q: Can small SPOs with limited data still use AI?
Yes — but often in smaller, simpler ways. Many SPOs begin by clarifying decisions or piloting lightweight solutions with partners. The goal is not sophistication, but usefulness and trust.

B. Applying the Phases

Q: Do we need to follow all nine steps exactly?
Think of the guide as a map, not a rulebook. Each step adds value, but how deeply you engage depends on your goals, team, and context.

Q: What if our problem doesn't feel "AI-worthy"?
That's a valid signal. Many good outcomes of Phase 1 involve deciding that AI is not the right tool — for now. This guide values clarity and restraint as much as innovation.

C. Working with Data

Q: What if our data is messy or incomplete?
That's completely normal. Most SPO data is messy. Phase 1 helps you decide whether your data is good enough for the decision you want to make. If gaps are too large, the right choice may be to redesign or pause.

Q: How do we keep our data safe and private?
Data safety and consent are non-negotiable. Phase 1 helps you surface risks early, and Phase 4 helps you put simple governance in place. If trust cannot be protected, the project should not proceed.

D. Building Teams

Q: Do we need a data scientist on our team?
Not necessarily. Many SPOs partner with academic institutions, fellows, or technical volunteers. What matters is that someone understands your context deeply, and someone understands the data deeply.

Q: How can we train our team to be AI-ready?
Start small. Focus on shared understanding rather than technical skill. This guide is designed to help teams ask better questions, not become AI experts overnight.

E. Ethics, Risk & Governance

Q: How do we make sure our AI model is fair and unbiased?
Bias often hides in data and design choices. The guide helps you surface these risks early (Phase 1) and monitor them over time (Phase 4). Community feedback remains the strongest safeguard.

Q: What are the risks of using AI in social impact work?
Risks include exclusion, misuse, and over-reliance on automation. This is why the guide emphasises clear boundaries, human oversight, and the option to pause or stop.

Q: Do we need consent to use community data in our model?
Yes, not just legally, but ethically. Even if anonymised, data represents people's lives. They have a right to know how it's being used and to benefit from its insights.

F. Tools, Budget & Sustainability

Q: How much does it cost to build an AI system?
Costs vary widely. Some pilots can begin with minimal expense, while others require sustained funding. Phase 1 helps you assess whether the expected value justifies the investment — and whether doing nothing is the better choice.

Q: Once we build a model, who maintains it?
That's covered in Phase 4. AI systems are like gardens; they need ongoing care. You can build internal capacity or set up lightweight partnerships for maintenance, updates, and monitoring.

Still Wondering About Something?

If your question isn't here, that's okay. Use the Worksheets or the Appendix to jot it down. Every responsible AI journey starts with honest, human questions.

Annexure

Attribution & Licensing

Feedback & Suggestions
We welcome feedback to improve future editions of this guide.

Email: cdssi@isdm.org.in
Online feedback form

A Closing Note

AI in social impact is different. It operates in contexts of scarcity, complexity, and lived human consequence. Data is often incomplete. Resources are limited. And the cost of getting things wrong is borne by communities, not systems.

This guide was written to support thoughtful, grounded choices — not faster deployment, bigger models, or technological spectacle. It is a guide for organisations that want to pause before building, listen before automating, and care after deployment.
Whether you are a programme manager, a data partner, or a funder, the invitation is the same: walk together, ask better questions, and build systems that strengthen human judgement rather than replace it.

If this guide helps you decide not to use AI when it isn't right, it has succeeded.
If it helps you build something smaller, slower, and more just, it has done its job.

ISDM

INDIAN SCHOOL OF DEVELOPMENT MANAGEMENT