

# Responsible AI and digital governance guidance for public sector capacity building (seed)

## Scope

This seed guide supports retrieval for projects training civil servants on AI use in policy and service delivery contexts. Replace with organization-approved or official guidance in production.

## Core principles for public sector AI capacity building

- Human oversight: AI-assisted outputs must be reviewed by responsible officials before operational use.
- Purpose limitation: approved use cases should have a defined administrative or service objective.
- Data minimization: use only necessary data for training and pilots.
- Privacy and security: align with agency data protection and information security requirements.
- Transparency: document when and how AI tools were used in workflows.
- Accountability: assign owners for validation, approval, and monitoring.
- Inclusion and accessibility: ensure training and deployment do not worsen service inequities.

## Training curriculum components

- AI literacy for public administration
- Safe prompt and workflow design
- Quality assurance and verification of AI-assisted outputs
- Bias, fairness, and inclusion considerations
- Data governance and records management
- Risk identification and escalation

## Operational safeguards for pilot use cases

- Approved use cases list and exclusion list
- SOP for testing in non-production or supervised environments
- Review checklist for policy notes, citizen communications, and service content
- Incident logging and corrective action protocol

## Suggested indicators for responsible AI adoption

- Percentage of trained staff passing post-training competency assessments
- Number of pilot units adopting approved AI-use SOPs
- Share of AI-assisted outputs reviewed and approved by supervisors
- Number of documented incidents or policy deviations and response time
- Improvement in processing time or consistency for selected administrative tasks

## Risk flags in project design

- Training without adoption support
- Tool access restrictions not addressed
- No documentation standards for AI-assisted outputs
- No focal points or governance owners