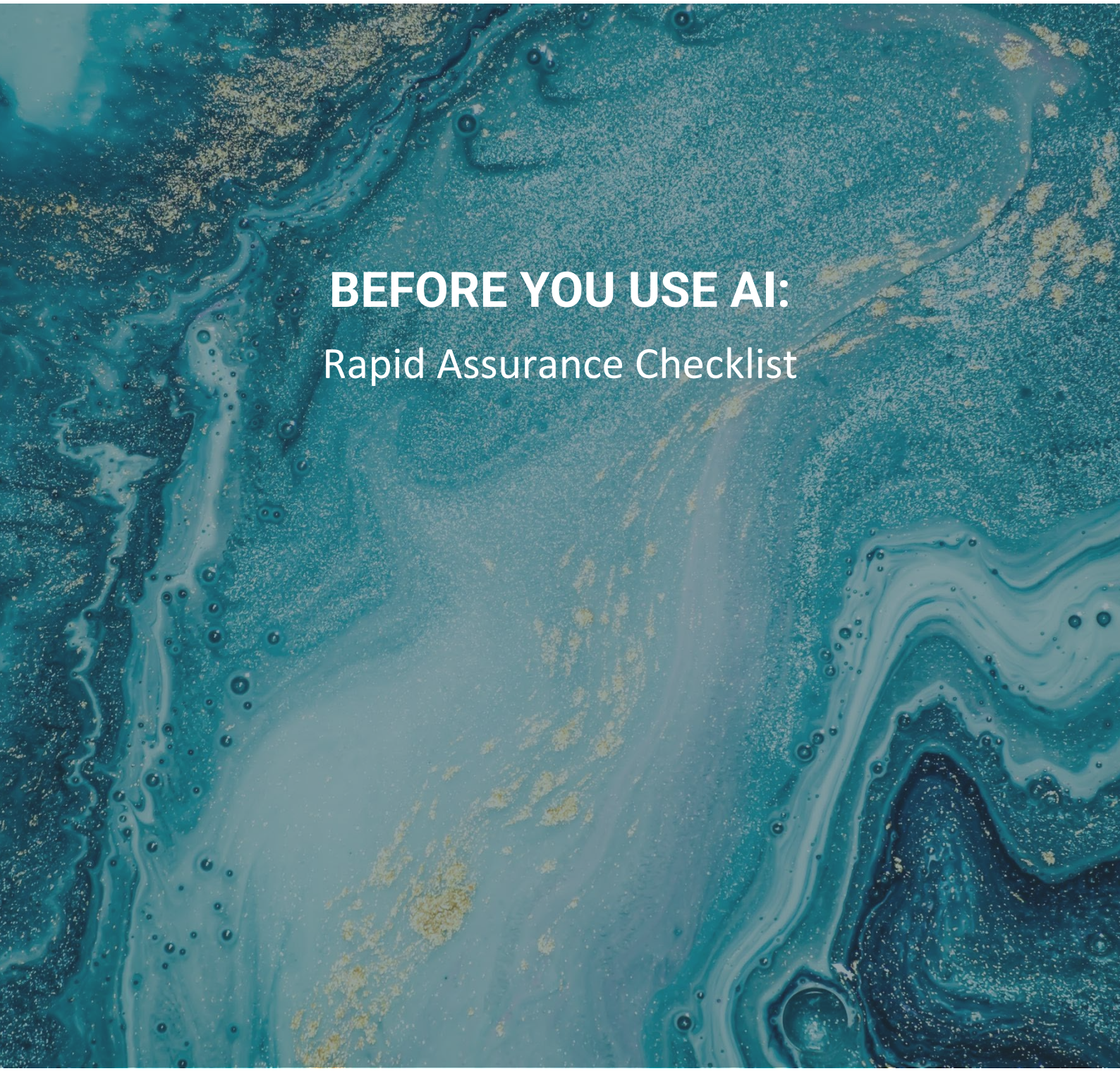
DATA CONSCIOUS



BEFORE YOU USE AI:

Rapid Assurance Checklist

AI ASSURANCE

AI assurance is the set of practices that let an organisation use AI in a way that is safe, lawful, and defensible, and be able to show evidence of that to stakeholders (management, donors, regulators, auditors, affected communities). In practice, AI assurance usually covers:

- **Risk & use-case classification**: what the AI is used for, who is affected, and how risky it is.
- **Data governance**: what data goes into the system, privacy/security controls, retention, lawful basis, DPIA where needed.[1]
- **Model/system evaluation**: testing for accuracy, hallucinations, bias/fairness.
- **Human oversight & accountability**: who is responsible, what must be verified by humans, sign-off points, escalation routes.
- **Transparency & documentation**: an audit trail of tools/models/versions, prompts, key decisions, and limitations (so results are reproducible and contestable).
- **Operational controls**: access control, monitoring, incident response, vendor due diligence.
- **Ethics & do-no-harm**: especially for vulnerable groups or fragile contexts.

THE RAPID ASSURANCE CHECKLIST

This checklist is designed as a **rapid assurance check** for teams planning to use **LLMs in an evaluation workflow** (e.g., transcription support, translation, summarizing, qualitative analysis/coding support, drafting). It reflects the most common concerns raised by Data Conscious' AI assurance clients, especially **data sensitivity, output reliability, bias, and defensibility** in public-facing findings and recommendations. It is based on a synthesis of widely used governance and sector guidance, including:

Table 1: Common AI Governance Frameworks and Guidelines[2]

GUIDANCE	SCOPE & COVERAGE
GDPR	Lawful data processing, minimization, security, DPIA/accountability.
UN	Principles for the Ethical Use of Artificial Intelligence in the United Nations System.
EU AI ACT	Risk-based AI implementation approaches, transparency and human oversight expectations, and high-stakes "red line" use cases.
CDAC'S SAFE AI	Safe AI guidance: do-no-harm, participation, accountability in humanitarian contexts.
NTEN	Nonprofit responsible AI guidance: practical organizational governance and responsible adoption.
NETHOPE	Responsible AI / suitability-oriented resources used in the nonprofit/humanitarian sector.

This checklist is provided as **practical guidance only**. It is **not legal advice**, and it does not replace your organization's policies, donor requirements, or professional legal/compliance review where needed, particularly for **high-stakes uses** (e.g., safeguarding triage, eligibility/access decisions, sanctions/discipline, or profiling of individuals).

[1] A Data Protection Impact Assessment (DPIA) is a structured process designed to identify, assess, and mitigate data protection risks associated with a project or processing activity, particularly when it is likely to result in high risk to individuals' rights and freedoms under GDPR.

2 This checklist draws on the guidance referenced in Table 1, but is an independent synthesis. Data Conscious has no affiliation with the guideline authors and did not seek or receive their endorsement.

1) MAP YOUR USE CASES

Complete Table 3 (overleaf) by listing each intended use of an AI tool you have for your evaluation. for each intended use, record the following information in the appropriate column:

USERS: The roles or people who will actually operate the AI tool in practice (e.g., evaluators, research assistants, MEAL staff, analysts). Include whether they're internal staff, consultants, or partners, and who has permission to run prompts/uploads.

DATA IN: What information will be provided to the AI tool. Be specific about format and sensitivity (e.g., raw audio files, verbatim transcripts, de-identified excerpts, survey free text, policy documents, notes). This is where you note whether it includes personal data, special category data, safeguarding content, etc.

USE: The intended purpose and workflow step: what you are using the AI output for (e.g., speed up transcription; produce a first-pass translation; generate an evidence-linked summary; propose initial codes to be reviewed; draft report structure). Clarify whether it's for internal working notes, or directly feeding into report findings/recommendations.

OWNER: The single accountable person (named role) responsible for ensuring the use case is approved, controls are followed, outputs are verified, and issues are handled. Not necessarily the person running the tool day-to-day. It's more like the accountable lead.

2) ASSIGN A RISK RATING

Tag each intended AI use with a traffic light rating to identify the levels of risk associated with each one:

Table 2: Risk Rating

RISK RATING	EXAMPLES
GREEN	non-sensitive edits/formatting of documents; summaries of publicly available documentation; structural drafting only
AMBER	sensitive data or vulnerable groups; protection contexts; outputs inform public findings or recommendations
RED	safeguarding triage; eligibility/access decisions; sanctions/discipline; profiling/scoring

Table 3: AI Use-Case Mapping

USE CASES	USERS	DATA IN	USE	OWNER	RISK RATING
TRANSCRIPTION					
TRANSLATION					
SUMMARIZING TEXT					
QUALITATIVE ANALYSIS					
DRAFTING FINDINGS					
OTHER(S)					

3) ESTABLISH MINIMUM CONTROLS

GLOBAL CONTROLS:

☐ AGREE THE "NEVER UPLOAD" LIST: documents or data types you will not upload to any AI system throughout the evaluation.

☐ MINIMISATION/REDACTION PROCESS AGREED: how and what you will redact from the documents you intend to upload to an AI system.

☐ TOOL BEHAVIOUR CONFIRMED: defining rules for the AI tool regarding data retention, training on/off, data sovereignty and storage locations.

☐ VERIFICATION PLAN AGREED: defining how you verify the validity of the results (e.g. bilingual checks; quote verification; sampling approach).

☐ AUDIT TRAIL REQUIREMENTS AGREED: decide what records you will keep to prove responsible use (e.g. which tool was used, key settings, prompt templates, what data processed, what outputs were generated, how checked/approved them).

☐ HUMAN SIGN-OFF POINTS AGREED: which steps require a human to review and approve before anything is used further (e.g. verbatim quotes, translation accuracy, key findings and recommendations).

CASE-SPECIFIC CONTROLS:

USE CASES	MINIMUM CONTROLS
TRANSCRIPTION	☐ Benchmark on a short segment, record typical errors and update the prompt
TRANSLATION	☐ Fluent speaker spot-checks a defined sample (e.g., 5–10% of segments). ☐ Keep original and the translation linked for traceability
SUMMARIZE	☐ "Grounded summaries only": ensure the AI provides supporting references or citations. ☐ "No new facts" enforced and uncertainty or contradictions are flagged.
QUALITATIVE ANALYSIS	☐ Human-Defined Codebook Used: ensure the AI sticks to the analytic categories you provide. ☐ Each code/theme includes supporting excerpt(s).
DRAFT	☐ Major findings triangulated: decide minimum number of sources or appropriate range of data sources/types. ☐ Causal claims written by humans with limits/counter-evidence noted. ☐ Two-person check for key findings and recommendations in sensitive contexts.

DATA CONSCIOUS