

NIST AI Risk Management Framework

*A tool, not a syllabus: **four things you do**, in service of **seven things you're aiming for**. (NIST AI 100-1, 2023 — voluntary, rights-preserving, sector-agnostic.)*

The four functions — what you do

- **Govern** — who's accountable, and what's the culture? The standing function: roles, policies, and oversight that make the other three actually happen.
- **Map** — what is this system, in context, and where could it go wrong? Frame the purpose, the people affected, and the risks before you build or deploy.
- **Measure** — how bad, how likely, and are we watching it? Test, quantify, and monitor the risks Map surfaced.
- **Manage** — so what do we do about it? Prioritize, act, and keep responding as things change.

The seven trustworthy-AI characteristics — what you're aiming for

- **Valid & reliable** — it does what it claims, consistently, under real conditions.
- **Safe** — it doesn't endanger life, health, property, or the environment.
- **Secure & resilient** — it withstands attack and recovers from failure.
- **Accountable & transparent** — someone is answerable, and how it works is disclosed.
- **Explainable & interpretable** — you can say what it did and why, usefully.
- **Privacy-enhanced** — it safeguards identity, autonomy, and data.
- **Fair, with harmful bias managed** — benefits and harms aren't distributed by prejudice.

Use it: hold a case up to the four functions and ask, at each, *what would this have asked, and would it have caught the harm?* Its strength is a shared language; its limit is that it is voluntary — it binds no one, and a checked box is not a prevented harm.