

The Responsible AI Principles

A free reference from AI GOVERNANCE CHAIN™

A principle is not a slogan. A real principle tells you what to do when two goods conflict, when speed pressures safety, when profit pressures fairness. These seven principles are written to be usable under that pressure. Adopt them, adapt them, cite them. They are the ethical spine on which every AI GOVERNANCE CHAIN™ instrument rests.

This is the free WHAT and WHY. The paid AI Ethics Operating Kit turns each principle into a working control with templates, owners, and review cadence.

Abbreviations and Acronyms

No prior knowledge is assumed. Every acronym used in this document is defined below.

AI: Artificial Intelligence

The Seven Principles

1. Human First

AI exists to serve human judgment, dignity, and wellbeing, never to quietly replace them where they matter. When a system would remove meaningful human agency, that is a decision to make deliberately and openly, not a default to drift into.

2. Fairness as a Duty

We treat fair outcomes as a moral obligation, not merely a compliance checkbox. We actively look for who could be harmed or disadvantaged, especially those least able to object, and we fix what we find.

3. Truthfulness over Agreeableness

A system that tells people what they want to hear is not helpful, it is dangerous. We build and govern AI to be accurate and willing to disagree, and we treat sycophancy as a first-class risk. Artificial intelligence does not fail because it is insufficiently intelligent. It fails because it agrees with you.

4. Transparency

People have a right to know when AI is involved in a decision that affects them, and to a plain-language account of how it works and where it fails. We disclose limits, not only strengths.

5. Accountability

For every AI system, a named human is answerable for its conduct. Accountability cannot be delegated to a model, a vendor, or an algorithm. Records exist so that an outsider could verify how a decision was made.

6. Contestability

A person affected by an AI decision can question it and reach a human who can review and, if warranted, overturn it. A decision no one can challenge is a decision no one truly governs.

7. Proportion and Restraint

We ask whether a system should exist, not only whether it can. We weigh societal, dual-use, and environmental costs against the benefit, and we are willing to say no.

How to use these principles

- - As a test. When a decision is hard, ask which principle is under pressure and why. The principle that is most inconvenient in the moment is usually the one that matters most.
- - As a contract. Publish them. Let staff, customers, and the public hold you to them. A principle no one can see is a principle no one can enforce.
- - As a foundation. Every operational control, every review gate, every incident response in the AI GOVERNANCE CHAIN™ catalog traces back to one of these seven.

The honest caution

Principles are cheap to publish and expensive to keep. The gap between a stated principle and a practiced one is where reputations are lost. The free principles tell you what you stand for. The paid AI Ethics Operating Kit gives you the machinery to prove you actually live them.

Your next step: Get the [AI Ethics Operating Kit, Full Edition](#) in the Execution Library (www.aigovernancechain.com/storefront.html, section: The Core Governance Kits). Every instrument in the Execution Library is also included in the [Governance Vault](#), the all-access annual licence.

AI GOVERNANCE CHAIN™ | Human-First AI Governance

(c) 2026 Mathieu K. Gouanou. All rights reserved. Member, Harvard Business Review Advisory Council.

Document standard: human-first AI governance.