Principles for Deployable Industrial AI

A public-facing baseline for trustworthy AI, robotics, and autonomy in critical industry

Why Minerva, Why Now

AI-enabled autonomy is moving from experimentation into industrial operations. In critical infrastructure, the opportunity is not simply automation - it is safer operations, faster recovery, higher productivity, and decisions at machine speed. MINERVA helps industry make that transition responsibly: adopting AI and robotics while preserving reliability, safety, security, workforce capability, and public trust.

The Need

Industrial AI cannot be governed by aspiration alone. Critical sectors need shared principles broad enough for energy, utilities, transportation, and manufacturing and specific enough to guide measurable action inside real organizations.

These principles provide Minerva's baseline for trustworthy, deployable AI in operational technology and cyber-physical environments.

How to Use the Action Guidance

Action Guidance translates each principle into operation. Use them to shape AI governance, procurement language, pilot gates, operator training, board reporting, and post-deployment assurance.

The forthcoming guide will add maturity levels, No-Regrets Controls, and implementation tools.

THE PRINCIPLES - PAGE 1 OF 2

1. Mission-Aligned Industrial Benefit

Principle: Adopt AI only where measurable value advances safety, resilience, productivity, competitiveness, and trust.

Action Guidance: Organizations should adopt AI only where there is a clear operational, safety, economic, resilience, environmental, or workforce benefit - and where that benefit outweighs the risk introduced by automation, model uncertainty, increased connectivity, or system complexity.

2. Use-Case Discipline Before Deployment

Principle: Start with a defined operational problem. Classify risk, autonomy, reversibility, data sensitivity, and impact before launch.

Action Guidance: Before deploying AI in industrial environments, organizations should inventory and classify each use case by risk level, operational criticality, autonomy level, data sensitivity, reversibility, and potential impact on people, assets, continuity, and public services.

3. Safety, Reliability, and Cyber-Physical Resilience

Principle: Engineer AI for safe behavior under normal, degraded, adversarial, and failure conditions.

Action Guidance: Organizations should treat AI-enabled industrial systems as cyber-physical systems. AI safety cases should address latency, drift, false positives, false negatives, degraded sensors, incorrect recommendations, unsafe automation, operator cognitive load, network interruptions, and manipulation.

4. Human Authority, Accountability, and Control

Principle: Keep humans accountable and able to supervise, challenge, pause, override, or disable AI-enabled operations.

Action Guidance: Organizations should define when AI can advise, when it can automate, when it requires human approval, and when it must be prevented from acting. Human operators, engineers, and accountable executives must understand how AI affects operational decisions and have authority, time, and protocols to authorize or override AI.

5. Secure-by-Design AI and Agentic Autonomy

Principle: Treat AI agents as security-relevant actors with identities, scoped permissions, boundaries, logs, and constraints.

Action Guidance: Organizations should apply least privilege, identity and access management, segmentation, defense-in-depth, continuous monitoring, adversarial testing, and incident response to AI systems, especially where agents can take actions, call tools, access sensitive data, or interact with OT/ICS environments.

6. Data Integrity, Provenance, and OT Boundary Protection

Principle: Govern industrial data quality, lineage, ownership, permitted use, and exposure across OT and enterprise boundaries.

Action Guidance: Organizations should protect industrial data as a strategic asset. Training data, operational data, sensor streams, maintenance histories, engineering diagrams, logs, model inputs, prompts, and retrieval sources must be governed for quality, provenance, security, privacy, and appropriate use.

The Principles, Continued | Commitments for safe, secure, and resilient AI adoption

THE PRINCIPLES - PAGE 2 OF 2

7. Transparency, Explainability, and Audit Evidence

Principle: Make AI understandable to operators and traceable enough for oversight, audit, incident review, and executive trust.

Action Guidance Organizations should maintain evidence that shows what an AI system was intended to do, what data it used, how it was tested, when it changed, what decisions or recommendations it produced, and the humans who were accountable for approving or acting on them.

8. Continuous Assurance, Monitoring, and Incident Readiness

Principle: Monitor performance, drift, abnormal behavior, security events, and incidents across the full AI lifecycle.

Action Guidance: Organizations should establish lifecycle assurance for AI systems, including pre-deployment testing, runtime surveillance, post-deployment monitoring, model drift detection, vulnerability management, incident reporting, and periodic reauthorization.

9. Workforce Readiness and Human-Machine Teaming

Principle: Use AI to augment industrial workers' judgment, capability, safety, and productivity without eroding critical operational competence.

Action Guidance: Organizations should train engineers, operators, technicians, cybersecurity teams, executives, and frontline supervisors to understand AI's capabilities, limitations, failure modes, and appropriate use in industrial contexts.

10. Vendor Accountability, Interoperability, and Hidden AI Risk

Principle: Require vendors to disclose AI capabilities, data use, model changes, security roles, audit rights, and disablement options.

Action Guidance: Organizations should require vendors, system integrators, managed service providers, and cloud providers to disclose AI capabilities, data usage, model update practices, external connections, security responsibilities, audit rights, fallback modes, and disablement options.

11. Testbeds, Pilots, and Evidence-Based Deployment

Principle: Move from sandbox to simulation, shadow mode, supervised pilot, limited production, and scaled deployment with evidence.

Action Guidance: Organizations should use staged deployment pathways: sandbox, simulation, digital twin, offline evaluation, shadow mode, supervised pilot, limited production, then scaled deployment.

12. Standards Alignment and Regulatory Readiness

Principle: Map AI governance to recognized AI, cyber, safety, sector, and emerging regulatory frameworks.

Action Guidance: Organizations should align AI governance with recognized frameworks and sector obligations, including NIST AI RMF, NIST Cybersecurity Framework, ISA/IEC 62443, sector-specific rules, safety standards, AI-in-OT guidance, and applicable U.S., EU, and international AI requirements.

13. Ethical Use, Public Trust, and Societal License

Principle: Assess impacts on workers, communities, reliability, affordability, human rights, environmental outcomes, and public confidence.

Action Guidance: Organizations should assess the broader impacts of AI on workers, communities, customers, safety, affordability, reliability, access, environmental outcomes, and public confidence.

14. Cross-Sector Collaboration and Shared Learning

Principle: Share noncompetitive lessons across sectors to strengthen resilience across interconnected critical infrastructure.

Action Guidance: Organizations should contribute to a trusted community of practice that shares noncompetitive lessons about AI opportunities, failures, incidents, governance practices, test results, workforce needs, vendor issues, and policy gaps.

These principles are intended to help industry move faster by moving smarter - accelerating AI adoption while preserving the safety, reliability, resilience, and public trust on which critical infrastructure depends.

Coming next: a member implementation guide with maturity model, No-Regrets Controls, and adoption tools.