Human-centered Sociotechnical Approaches to Trustworthy AI

ELHAM TABASSI

BROOKINGS INSTITUTION

Manufacturing's AI Future

Managing Risk
Leading Globally
Empowering Workers

Managing AI Risks

The NIST AI Risk Management Framework

The challenge of AI risk in manufacturing

Traditional Manufacturing Equipment

- ✓ Behaves predictably
- ✓ Follows programmed logic
- ✓ Deterministic outputs
- ✓ Traceable failures

AI Systems

- Learn from data
- Make probabilistic predictions
- Can produce unexpected outputs
- Performance varies with data

Real Manufacturing Risks

Vision system misses defects → dangerous products reach customers

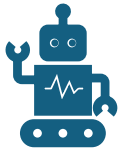
Predictive maintenance false negatives → catastrophic equipment failure

Optimization suggests parameters that cause quality issues

Manufacturing's physical consequences make managing AI risks essential, not optional

AI RMF as enabler of innovation

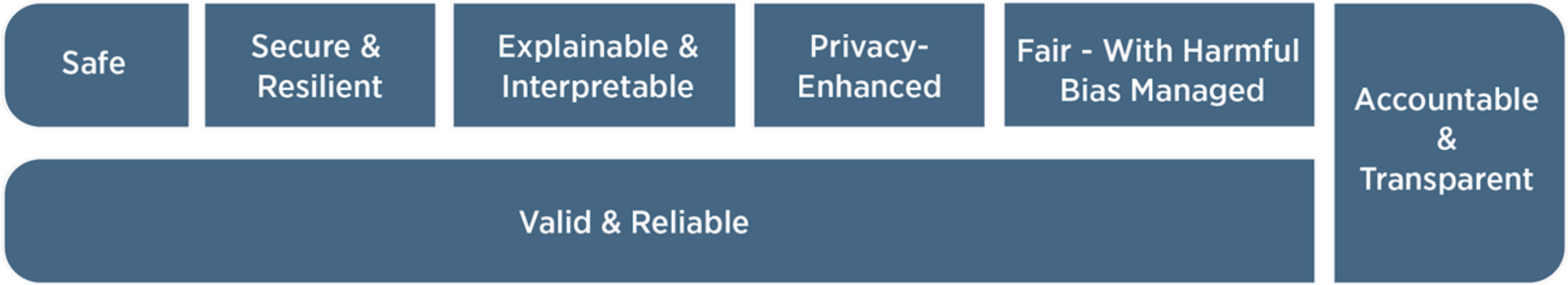
De-risk innovation for manufacturers.



Scale AI adoption with greater confidence.

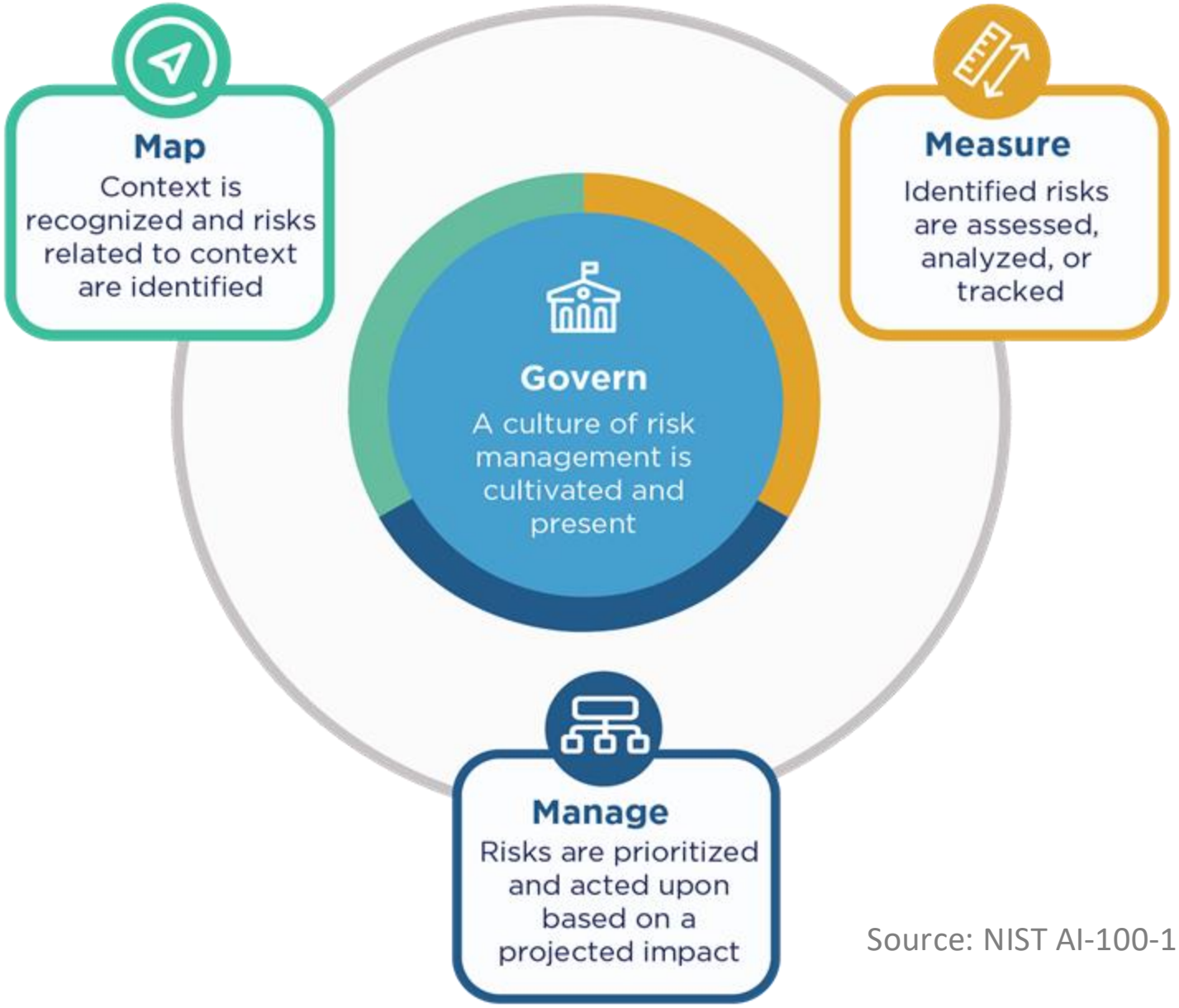
Reduce liability and build trust with workers and customers.

Key characteristics of trustworthy AI



Source: NIST AI-100-1

The AI RMF Core lays out four organizational functions to facilitate trustworthy systems and responsible practice and use.



Source: NIST AI-100-1

GOVERN: Building accountability

Example: AI Quality Inspection

Who approves deployment?

Cross-functional team: Production, Quality, IT, Safety

Who's accountable for failures?

Clear ownership when AI misses defects

How to handle disagreements?

Protocol for AI vs. human inspector conflicts

Integrate with Existing Systems

Don't create parallel AI governance. Connect to ISO 9001, ISO 45001, sector-specific standards. Same approval processes as new production equipment.

Address Workforce Concerns

When AI automates tasks, what happens to workers? Retrain, redeploy, or dismiss? Governance decides before deployment, not during crisis.

Source: NIST AI-100-1

MAP: Understanding context

Documenting what the AI system does, who it affects, what can go wrong, and what data it depends on.

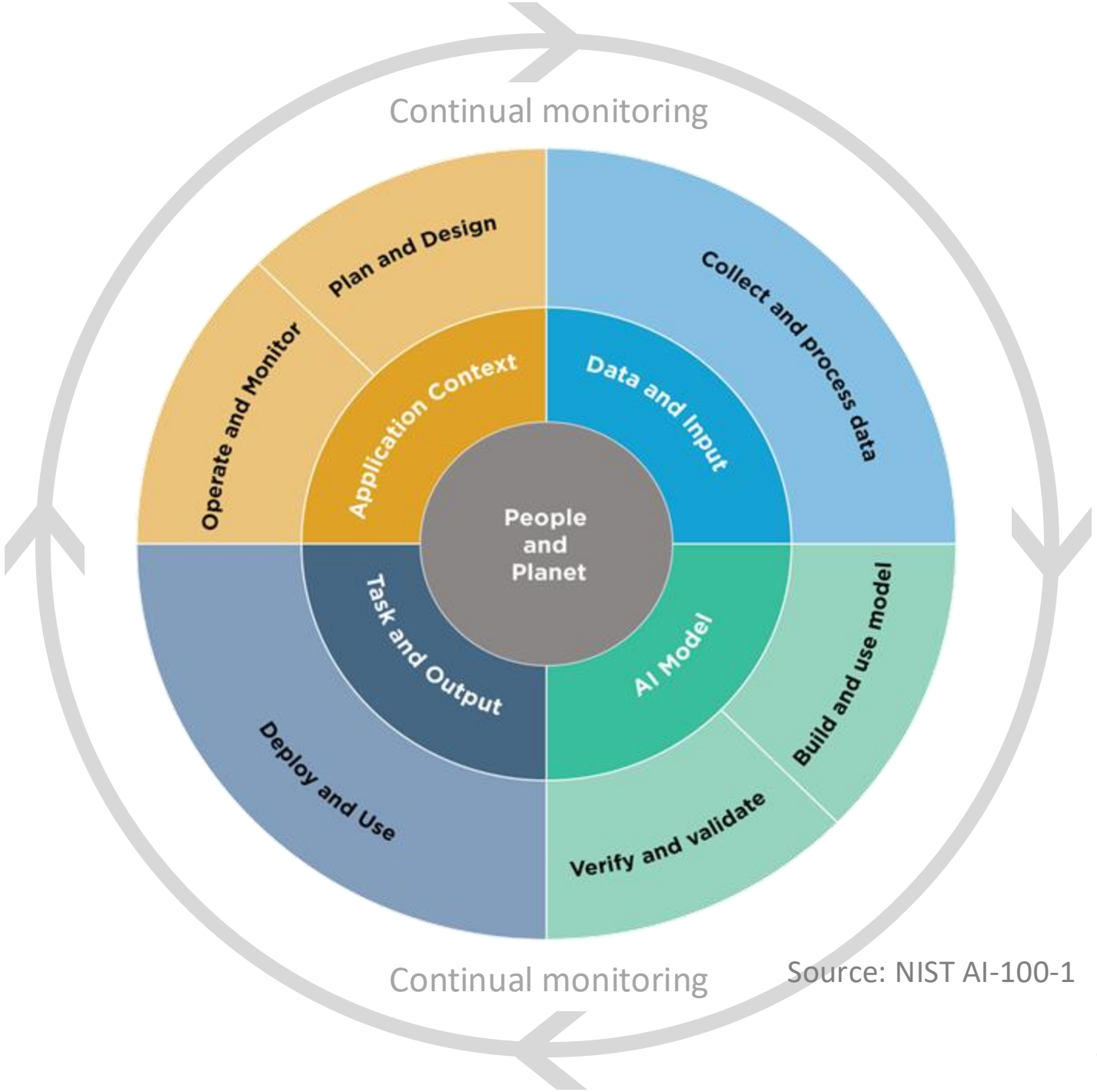
Understand failure modes. What happens if the AI system goes down? If it produces incorrect outputs? If it degrades gradually over time?

Source: NIST AI-100-1

MEASURE: Tracking performance
MANAGE: Respond to risk

Source: NIST AI-100-1

Beyond the system, a culture of responsible practice and use must pervade activities across the entire AI lifecycle.



Source: NIST AI-100-1

Implementation guides

Start small and scale

Integrate with existing processes

Build internal capability

Document everything

Engage suppliers

Check out AI RMF Playbook

Leading Globally

How Standards Shape Manufacturing Competitiveness

Standards enable interoperability and trade, *and* shape competitive advantage

The Strategic Reality

The countries and companies that shape AI standards for manufacturing will influence the industry's future. This isn't about compliance—it's about competitiveness.

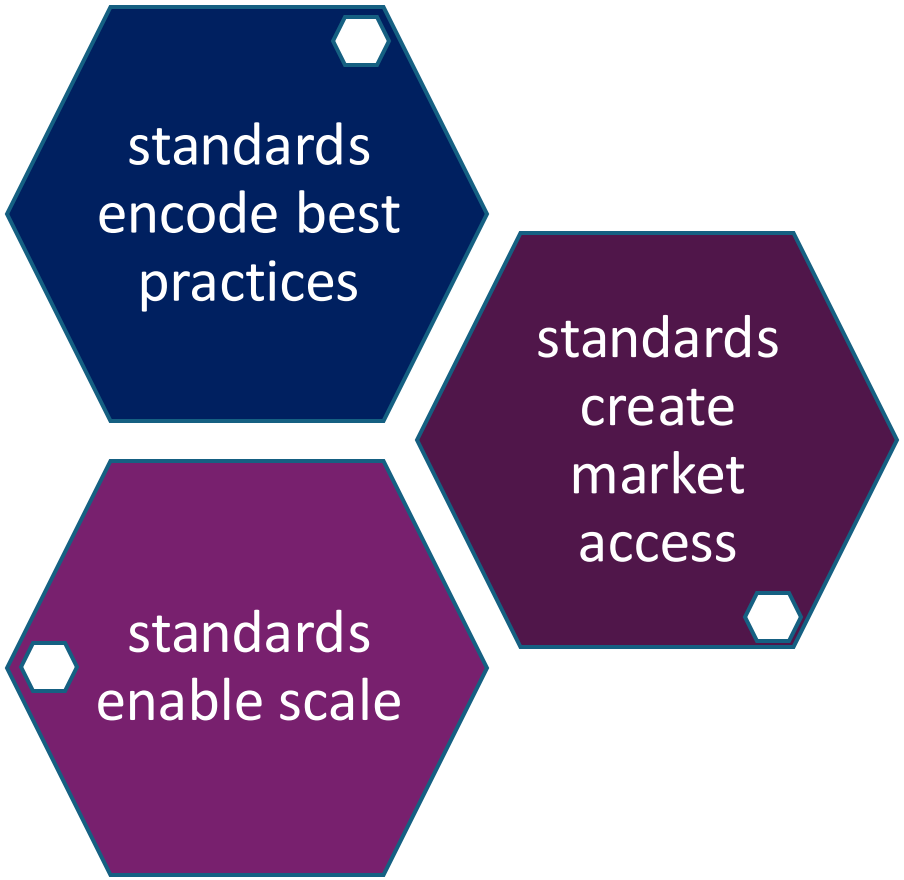
What Standards Determine

- What AI systems can do
- How they interoperate
- What testing they must pass
- What data formats they use
- What competencies workers need

Why Participation Matters

- ✓ Help write rules → align with your strengths
- ✓ Don't participate → adapt to others' rules
- ✓ First-mover advantage in implementation
- ✓ Shape ecosystems that attract investment

Standards bridge between innovation and trust.



Empowering workers

Building Effective Human-AI Teams

Technical Robustness and Societal Resilience

“AI does not happen to us.” The future of AI is not inevitable—it’s being built in rooms like this.

The question is not just what AI *can* do, but what we *want* it to do—for people, society, and the planet.

THANKS FOR YOUR ATTENTION!