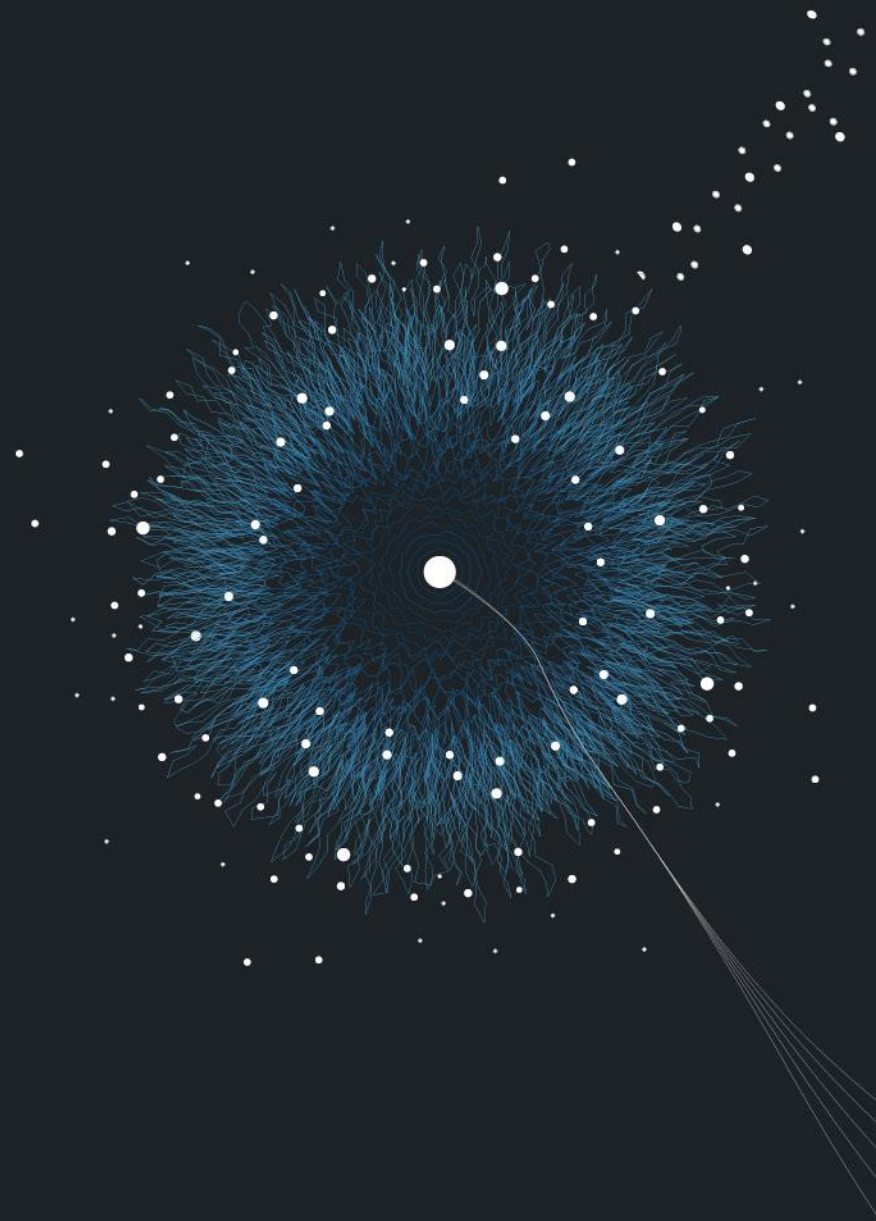
Innovation & Tech Talk Online Event 02/2026

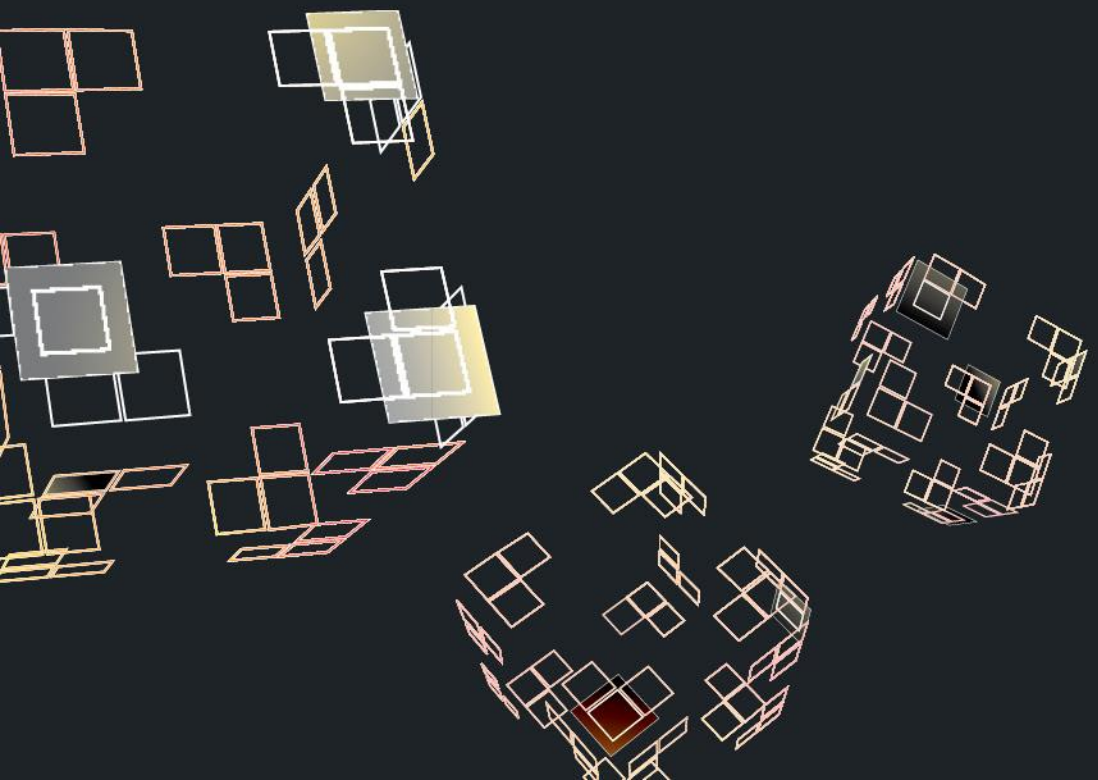
LEADING WITH AI – HOW LEADERSHIP SHAPES TRUST, EXPLAINABILITY AND GOVERNANCE IN KNOWLEDGE-INTENSIVE WORK



MARC BONGARTZ · UNIVERSITY OF TWENTE · ISACA TECH TALK · FEBRUARY 26

UNIVERSITY OF TWENTE.

TABLE OF CONTENTS



UNIVERSITY
OF TWENTE.

01 MOTIVATION, RESEARCH CONTEXT AND FRAMING

UNIVERSITY
OF TWENTE.

OWNERSHIP & ACADEMIC CONTEXT

Leading with AI -
How Leadership Shapes Trust, Explainability and Governance in Knowledge-Intensive Work

---

Marc Bongartz
MBA Management & Leadership, PhD candidate
Manager IT-Audit
m.bongartz@utwente.nl

Academic Supervision:
Dr. Pauline Weritz & Dr. Florian Schuberth

Affiliation / Context:
University of Twente
Faculty of Engineering Technology
ISACA Tech Talk | February 26

ARTIFICIAL INTELLIGENCE IN KNOWLEDGE-INTESIVE DECISION-MAKING

Artificial intelligence is increasingly embedded in organizational decision-making..

1

In these contexts, AI systems support complex judgments rather than replacing human decision-makers.
Decisions are made under conditions of uncertainty, professional discretion, and regulatory accountability..

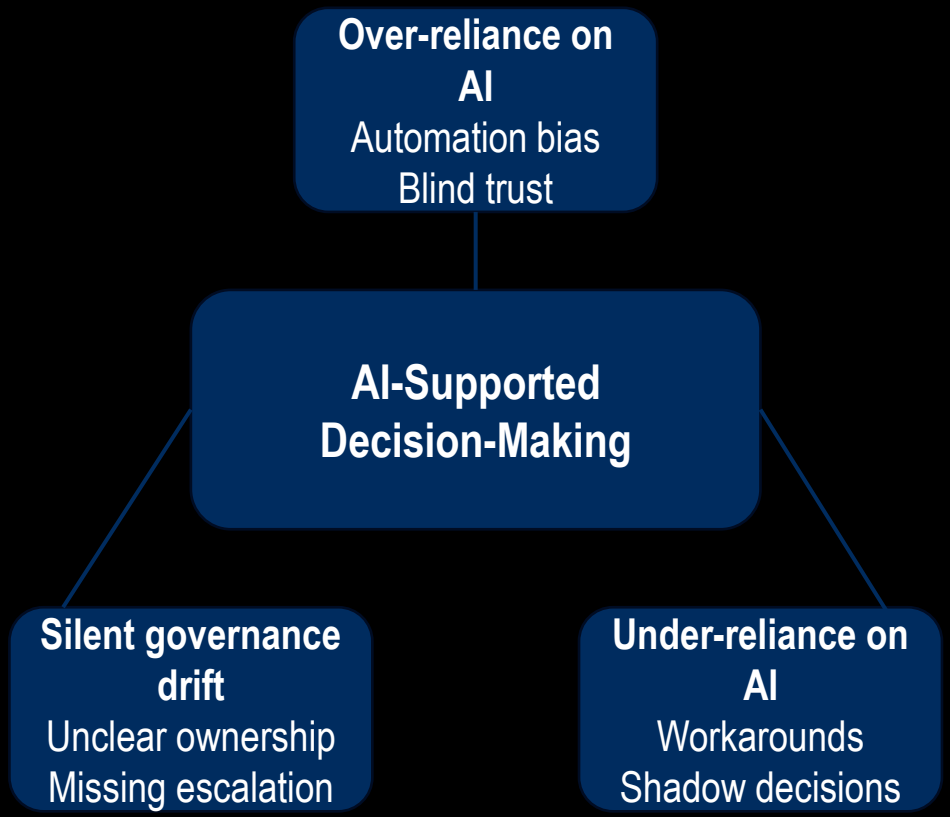
2

Explainable AI has therefore emerged as a prominent design principle, widely framed as a means to enable transparency, understanding, and trust in AI-supported decisions.

3

WHY THIS MATTERS FOR ISACA PRACTITIONERS

- AI-supported decision-making rarely fails because of missing explanations or poor models alone.In practice, failures emerge when trust in AI is miscalibrated.
- Leadership shapes whether organizations drift toward blind reliance on AI, systematic distrust, or unclear responsibility. These patterns are not technical errors, but governance failures.
- The key role of leadership is therefore to actively calibrate trust by structuring decision rights, framing explanations, and defining accountability and escalation.



THE TRUST PROBLEM BEHIND EXPLAINABLE AI

Explainable AI is often assumed to be sufficient for building trust in AI-supported decisions.
However, empirical research shows that transparency does not reliably result in appropriate trust.

In practice, explanations may lead either to persistent skepticism or to uncritical over-reliance on AI outputs.
Trust judgments are shaped not only by explanation quality, but by organizational context, professional norms, and accountability structures.

As a result, explainability can obscure responsibility, shift decision ownership, and create false confidence if it is not embedded in leadership and governance practices.

UNIVERSITY
OF TWENTE.

RESEARCH QUESTION & ANALYTICAL SCOPE

Explainable AI is widely discussed as a technical solution for fostering trust in AI-supported decisions.

However, existing research offers limited insight into how trust actually emerges and stabilizes in organizational contexts.

Leadership is often acknowledged as contextual background, but rarely theorized as a central mechanism shaping trust in explainable AI.

This research examines how leadership practices shape trust in explainable AI within knowledge-intensive work environments.

02 METHODOLOGICAL FOUNDATION: SYSTEMATIC LITERATURE REVIEW

UNIVERSITY OF TWENTE.

METHODOLOGICAL APPROACH: SYSTEMATIC LITERATURE REVIEW

Systematic literature review following established methodologies.

Focus on leadership, trust, and explainable AI across disciplines.

Structured database searches with transparent selection criteria.

Analysis of key concepts, theoretical perspectives, and recurring patterns.

Foundation for identifying research gaps

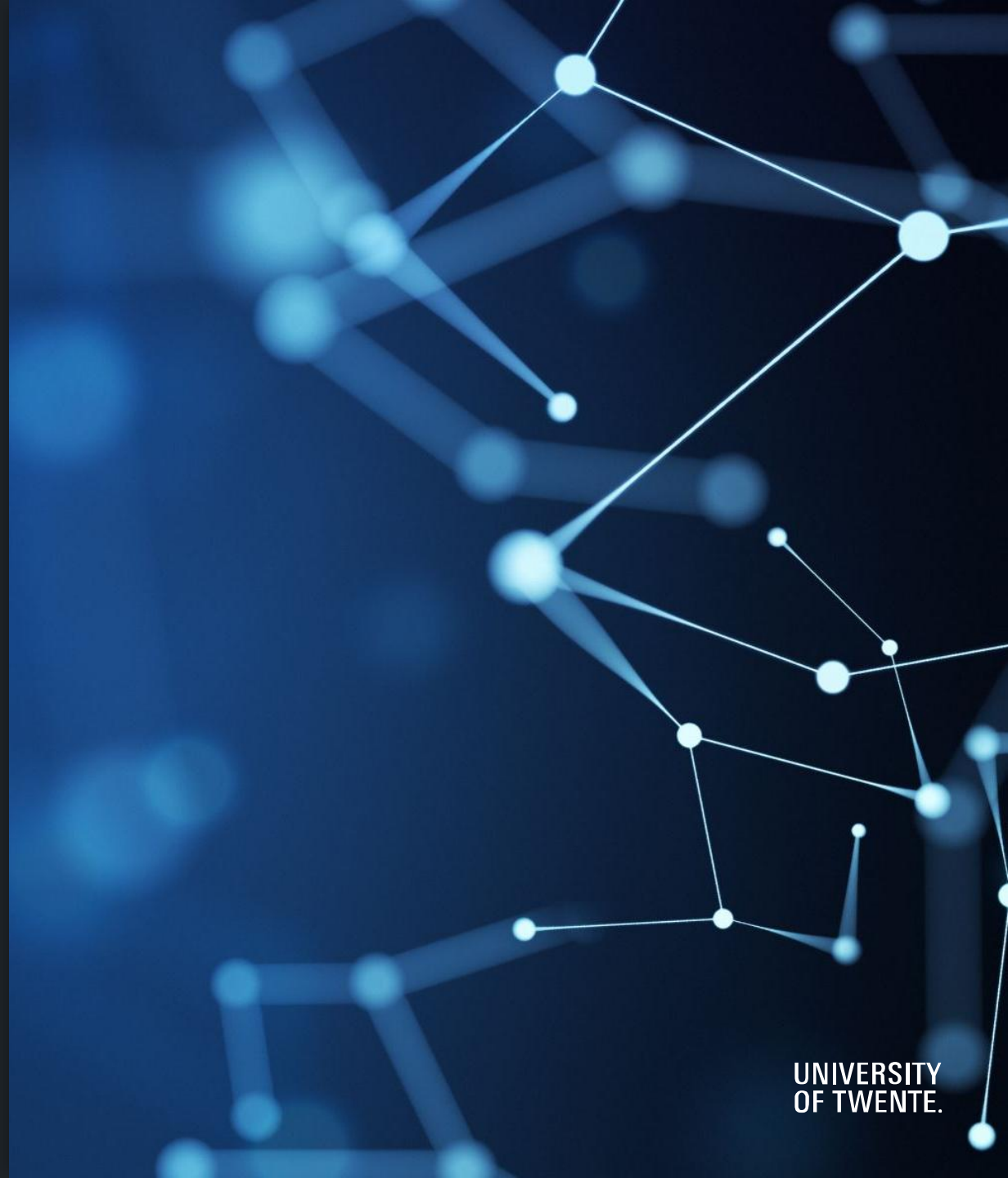
and developing a leadership-centered perspective.

ANALYTICAL FRAMEWORK AND CODING LOGIC

The identified literature was analyzed using a structured analytical framework.
The analysis focused on how trust, explainability, and leadership are conceptualized and related.

Key concepts, theoretical perspectives, and underlying assumptions were systematically coded.
Special attention was given to how leadership is framed: as a contextual factor or as an active mechanism.

The coding process enabled a cross-study synthesis and the identification of recurring patterns and gaps.



UNIVERSITY OF TWENTE.

03

KEY FINDINGS: LEADERSHIP, TRUST & EXPLAINABILITY

OVERVIEW OF DOMINANT RESEARCH STREAMS

Research on trust in explainable AI follows a small number of dominant perspectives. Most studies focus on technical explainability and user understanding as primary trust drivers.

A second stream conceptualizes trust mainly as an individual and psychological construct.

Organizational and leadership-related perspectives appear less frequently and remain fragmented.

As a result, practical guidance for managing trust in high-stakes organizational contexts is limited.

TECHNICAL EXPLAINABILITY PERSPECTIVE

The dominant research stream conceptualizes trust primarily as a function of technical explainability.

Trust is commonly assumed to increase as AI models become more transparent and understandable.

Explanations are treated as cognitive tools that enable users to assess model behavior and outputs.

Leadership and organizational context are largely treated as secondary or implicit factors rather than central mechanisms.

Selected research foundations:
Ribeiro et al. (2016); Doshi-Velez & Kim (2017); Guidotti et al. (2018)

TRUST BEYOND EXPLAINABILITY: WHY LEADERSHIP MATTERS

Explainable AI increases transparency, but transparency alone does not ensure appropriate trust.

Prior research indicates that explanations can lead to both under-reliance and over-reliance on AI.

In organizational practice, trust depends on how explanations are framed, legitimized, and embedded in decision processes.

Leadership shapes whether AI outputs are treated as contestable reasons or as authoritative system recommendations.

Trust in AI therefore emerges from leadership practices that structure accountability, decision rights, and appropriate reliance.

This shifts the focus from explainability as a technical feature to trust as an organizational outcome.

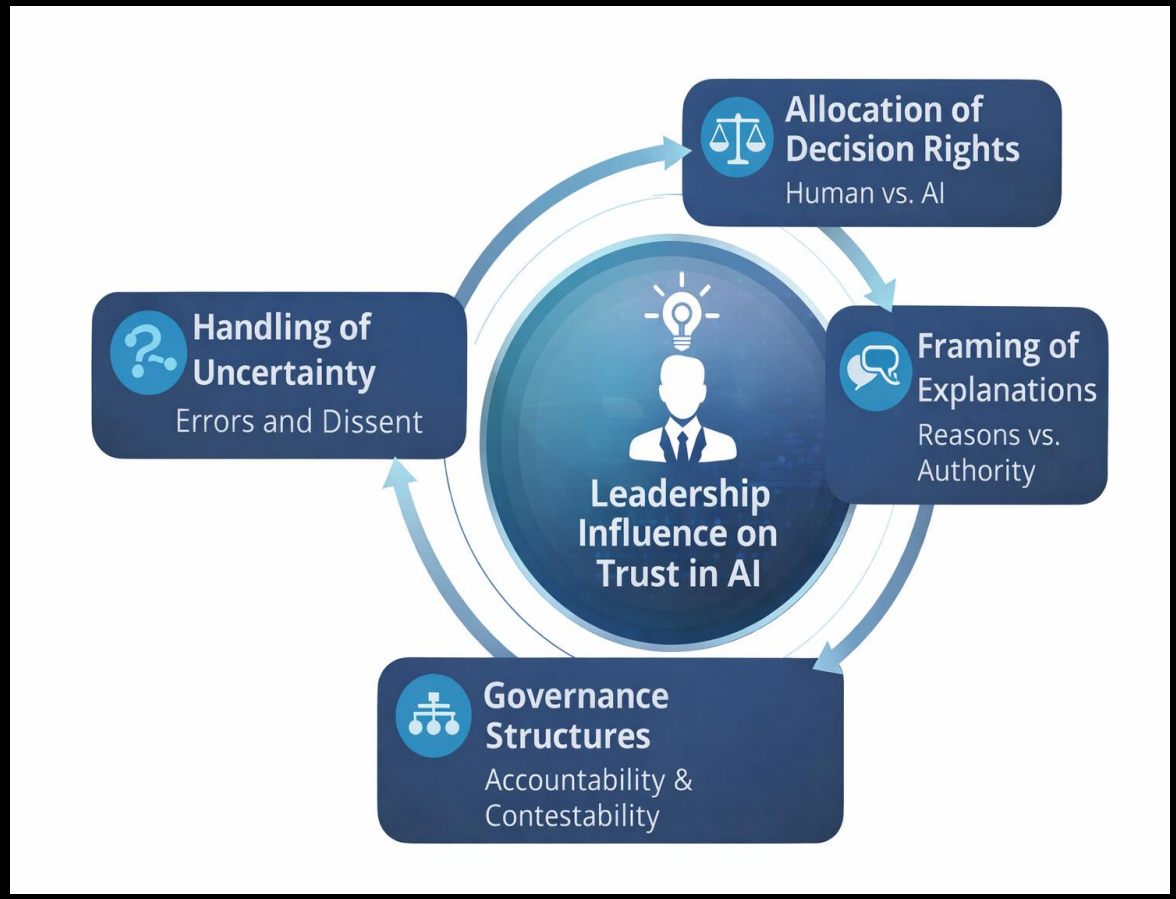
Selected research foundations:
Mayer et al. (1995); McKnight et al. (2002); Langer et al. (2021); Bevilacqua et al. (2025)

KEY LEADERSHIP INFLUENCE FACTORS ON TRUST IN AI

Leadership shapes trust in AI through a small number of recurring mechanisms, including the allocation of decision rights.

These mechanisms structure how AI is integrated into decision-making, how uncertainty is handled, and how accountability is assigned in practice.

The following framework synthesizes the key leadership influence factors identified across the literature.



Selected research foundations:
Weick (1995); Gioia & Chittipeddi (1991); Vial (2019); Hanelt et al. (2021); Bevilacqua et al. (2025).

UNIVERSITY OF TWENTE.

ALLOCATION OF DECISION RIGHTS: HUMAN VS. AI

The allocation of decision rights between humans and AI is a central leadership decision shaping trust in AI. Leaders define whether AI systems support human judgment, provide recommendations, or effectively replace human decision-making
.
When decision rights remain implicit or ambiguous, organizations face increased risks of blind reliance on AI outputs or systematic distrust and workarounds.

Clear allocation of responsibility is therefore essential to ensure accountability, auditability, and effective AI governance.

Selected research foundations:
Parasuraman et al. (2000), Vial (2019), Hanelt et al. (2021), Bevilacqua et al. (2025)

UNIVERSITY
OF TWENTE.

FRAMING OF EXPLANATIONS: REASONS VS. AUTHORITY

Framing matters

How AI explanations are framed fundamentally shapes how they are interpreted and acted upon in organizational decision-making.

Authority framing

When explanations are framed as authoritative, decision-makers tend to defer responsibility to the system. This increases automation bias and weakens accountability, particularly in high-stakes contexts.

Reasons framing

In contrast, framing explanations as reasons supports sensemaking, dialogue, and contestability.

Decision-makers integrate AI outputs with domain expertise and professional judgment

Leadership implication

Framing explanations is a leadership practice that directly shapes trust calibration, accountability, and auditability in AI-supported decisions.

Selected research foundations: Weick (1995), Gioia & Chittipeddi (1991), Langer et al. (2021), Bevilacqua et al. (2025)

HANDLING OF UNCERTAINTY, ERRORS, AND DISSENT

AI-supported decisions operate under conditions of uncertainty, incomplete information, and model limitations.

Leadership determines how uncertainty is acknowledged, how errors are interpreted, and whether dissent is permitted.

When uncertainty and errors are suppressed, organizations tend to conceal model limitations and discourage critical reflection.

This increases the risk of silent failures, automation bias, and unmanaged model risk.

In contrast, leadership practices that encourage transparency, error reporting, and constructive dissent support learning, oversight, and appropriate reliance on AI.

They are essential for auditability, incident management, and trustworthy AI governance..

Selected research foundations: Weick (1995), Edmondson (1999), Reason (1997), Bevilacqua et al. (2025).

FROM LEADERSHIP PRACTICES TO ACCOUNTABILITY STRUCTURES

Leadership practices shape how accountability structures around AI are designed and enacted in organizations. Decision rights, framing of explanations, and handling of uncertainty jointly determine how responsibility, oversight, and escalation are organized.

When leadership practices remain implicit, accountability becomes fragmented and control responsibilities are blurred.
This increases the risk of unmanaged model risk, unclear ownership, and ineffective auditability.

In contrast, explicit leadership practices enable clear role definitions, traceable decision paths, and effective escalation mechanisms.
Such accountability structures form the foundation of trustworthy, governable, and auditable AI systems.

Selected research foundations: Weick (1995), Vial (2019), Hanelt et al. (2020), Bevilacqua et al. (2025)

UNIVERSITY
OF TWENTE.

A LEADERSHIP-CENTERED MODEL OF TRUSTWORTHY AI GOVERNANCE

Trustworthy and governable AI emerges from the interaction between leadership practices, organizational structures, and technical systems.Leadership acts as the central mechanism linking explainability, accountability, and appropriate reliance on AI.

Decision rights, framing of explanations, and the handling of uncertainty shape how AI governance structures are enacted in practice.These leadership practices determine how responsibility, oversight, and escalation are organized around AI-supported decisions.

As a result, trust in AI is not a system property, but an outcome of leadership-centered governance design that integrates technical safeguards with human judgment.

Selected research foundations: Weick (1995), Vial (2019), Hanelt et al. (2021), Bevilacqua et al. (2025).

UNIVERSITY OF TWENTE.

04 LITERATURE-BASED IMPLICATIONS FOR AUDIT, IT GOVERNANCE & INFORMATION SECURITY

UNIVERSITY OF TWENTE.

IMPLICATIONS FOR AUDIT

Leadership-centered governance reshapes what becomes auditable in AI-supported decision-making.

Audit focus shifts from technical model properties to decision architectures, accountability structures, and governance routines.

1

Auditors must assess not only whether controls exist, but how decision rights are allocated, how explanations are used, and how overrides are governed.

This requires evaluating leadership practices that structure responsibility, escalation, and professional judgment.

2

Key audit challenges emerge around traceability of decisions, attribution of responsibility, and detection of automation bias.
Effective audit therefore depends on the integration of technical evidence with organizational context and leadership behavior.

3

Leadership-centered governance transforms audit from control verification to the assessment of decision accountability in AI-supported systems.

4

IMPLICATIONS FOR IT GOVERNANCE

1. Leadership-centered AI governance reshapes how IT governance structures are designed and enacted.
Governance moves from managing systems to governing decision architectures and accountability relationships.

2. Effective IT governance requires explicit allocation of decision rights, clearly defined roles, and formal escalation paths for AI-supported decisions.
Leadership practices determine whether governance principles are translated into enforceable routines or remain symbolic.

3. Key governance challenges arise around ownership of AI decisions, integration of AI into existing governance frameworks, and alignment between business, IT, and risk functions.
Leadership-centered governance enables coherent decision-making across organizational boundaries and reduces fragmentation of responsibility.

IMPLICATIONS FOR INFORMATION SECURITY

Leadership-centered AI governance reshapes how information security is operationalized in AI-supported decision environments. Security moves beyond protecting systems to safeguarding decision integrity, accountability, and controlled use of AI.

Governance & Responsibility
Leadership Oversight · Escalation · Accountability

Effective information security depends on how leaders define acceptable use, handle uncertainty, and respond to errors and incidents. Leadership practices influence whether security controls support professional judgment or enforce blind reliance on AI outputs

Decision Layer
Decision Integrity · Human Judgment · Contestability

Key security challenges emerge around automation bias, silent failure modes, model misuse, and inadequate incident escalation. Leadership-centered security governance integrates technical safeguards with human oversight and organizational responsibility.

Security Operations
Monitoring · Logging · Incident Response

05

KEY TAKEAWAYS AND DISCUSSION

UNIVERSITY OF TWENTE.

KEY TAKEAWAYS FOR ISACA PRACTICE

Takeaway 1:
Leadership is a governance mechanism for AI.

Trustworthy AI depends on how decision rights, accountability, and escalation are enacted.

Takeaway 2:
Explainability alone is insufficient.

Leadership determines whether explanations support judgment or enforce compliance.

Takeaway 3:
Audit, governance, and security converge around decision accountability.

AI governance requires integrated oversight of decisions, not isolated controls.

Takeaway 4:
Established governance and audit frameworks gain relevance when applied to decision architectures.

Their value lies in enabling leadership-centered governance practices.

CLOSING & DISCUSSION

Closing Remarks & Discussion
Trustworthy AI is not achieved through technology alone.
It emerges from leadership-centered governance that aligns decision-making, accountability, and oversight.

Thank you for your attention.
I look forward to your questions and discussion.

Marc Bongartz
Manager IT Audit & Information Security
PhD Candidate – Leadership & AI Governance
m.bongartz@utwente.nl
+49 176 22274614

Advisory & Assurance Expertise
- IT Audits & Information Security
- C5 Audits
- NIS-2 Readiness & Compliance
- ISMS Implementation
- Critical Infrastructure (KRITIS) Assessments
- AI Audits & AI Governance
- Business Transformation & AI Integration

UNIVERSITY OF TWENTE.

REFERENCES/ LITERATURE

- Behl, A., & Bevilacqua, M. (2025). Explainable artificial intelligence, trust, and human–AI interaction: Governance and leadership perspectives.
- Bevilacqua, M., et al. (2025). Leadership-centered AI governance: Trust, accountability, and decision-making in AI-supported organizations.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning.
- Edmondson, A. (1999) Psychological safety and learning behavior in work teams.
- Gioia, D. A., & Chittipeddi, K. (1991). Sensemaking and sensegiving in strategic change initiation.
- Guidotti, R., et al. (2018). A survey of methods for explaining black box models.
- Hanelt, A., et al. (2020). Digital transformation and strategy: A systematic literature review.
- Hanelt, A., et al. (2021). Why digital transformation succeeds or fails: A systematic review.
- Langer, M., et al. (2021). What do we want from explainable artificial intelligence? A stakeholder perspective.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce.
- NIST AI RMF (2023). National Institute of Standards and Technology. (2023). AI Risk Management Framework (AI RMF 1.0). NIST.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation.
- Reason, J. (1997). Managing the risks of organizational accidents.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier.
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda.
- Weick, K. E. (1995). Sensemaking in organizations.