



The Compliance Fluency Gap

Why Human-Led GRC Outperforms AI-First Platforms

Published by Valdyr.io

Graham Brooks, CTO, Valdyr.io

Ernest Spicer, CPO, Valdyr.io

September 2025

Audience

Chief Information Security Officers (CISOs), Chief Compliance Officers (CCOs), GRC Managers, IT and Security Directors, and executive leadership (CEOs, CFOs) concerned with risk and operational resilience.

Abstract

The rush to adopt Artificial Intelligence in Governance, Risk, and Compliance (GRC) is creating a dangerous paradox. While promising efficiency, an over-reliance on AI for generating compliance documentation and managing GRC programs is leading to generic, ineffective security postures and fostering a superficial "checkbox compliance" culture. This whitepaper critically examines the systemic failures of AI-first GRC, citing evidence of widespread project underperformance, the erosion of essential institutional knowledge, and the fundamental unsuitability of Large Language Models (LLMs) for creating net-new, factually-grounded compliance content. We then present a superior model: a human-led, visually-defined approach that uses a tree graph structure to build deep "compliance fluency." This model, exemplified by the Valdyr platform, structures organizations as roots that branch through policies, standards, controls, evidence and processes, with controls linking directly to frameworks like SOC 2 or CMMC, providing the hierarchical context necessary for genuine risk management and operational resilience, and positions AI in its proper role as an analytical co-pilot, not an unreliable author.

Table of Contents

Part 1: The AI GRC Paradox: How Automation Obscures Risk and Erodes Expertise

- 1.0 Introduction: The Allure and the Alarming Reality of AI in GRC
- 1.1 The Pitfall of Generic Compliance: When "Best Practice" Isn't Your Practice
- 1.2 The Erosion of "Compliance Fluency": Outsourcing Thought, Retaining Risk
- 1.3 Misaligned Tools: The LLM as a Flawed Content Generator
- 1.4 The Human-led, AI Augmented Alternative

Part 2: The Valdyr Solution: Building a Resilient, Human-Led Compliance Program

- 2.0 Introduction: Pivoting to a Model of Clarity and Context
- 2.1 The Power of the Tree Graph: Visually Mapping the "Why"
- 2.2 Beyond the Audit: Integrating Processes and Standards for Real-World Context
- 2.3 The Right Role for AI: The Human-Led Co-Pilot

Part 3: Conclusion

References



Part 1: The AI GRC Paradox: How Automation Obscures Risk and Erodes Expertise

1.0 Introduction: The Allure and the Alarming Reality of AI in GRC

Organizations are betting their security postures on a dangerous gamble: that Artificial Intelligence can replace human expertise in Governance, Risk, and Compliance (GRC). This industry-wide rush toward AI-first automation promises efficiency and sophistication, yet it's creating the opposite-generic, brittle compliance programs that crumble under real scrutiny. The statistics are damning: while 78% of organizations now use AI compared to 55% the previous year, this adoption surge masks a catastrophic failure rate, with recent research revealing that the vast majority of AI implementations deliver zero measurable business value (Stanford AI Index, 2025).

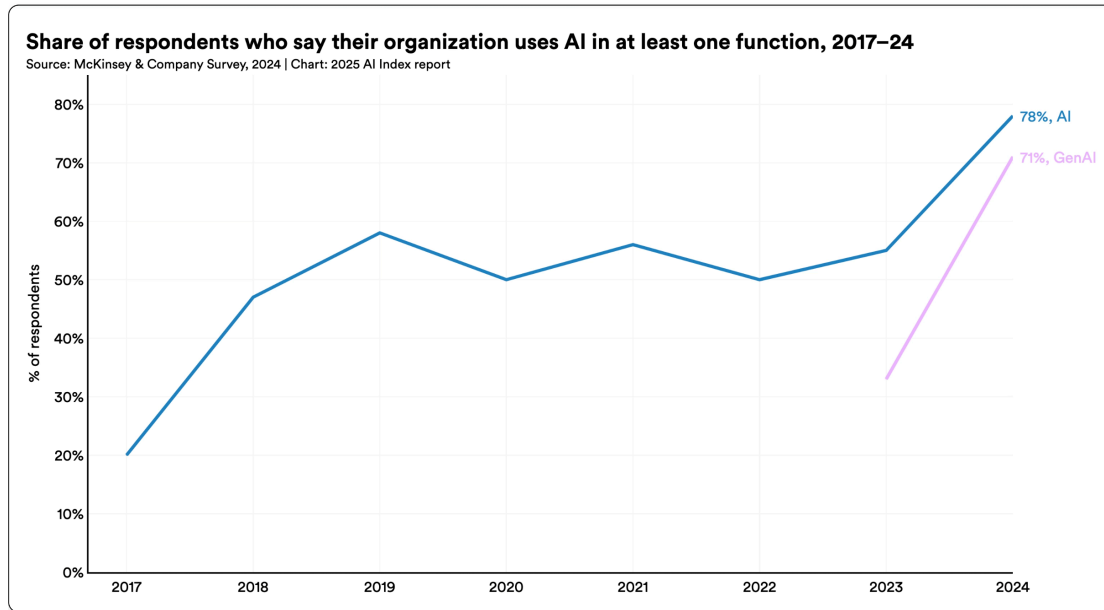


Figure 1: AI adoption rates showing 78% of organizations using AI and 71% using GenAI by 2024 (Source: Stanford AI Index 2025 Annual Report)

The strategic bet on AI-first "everything" is failing spectacularly. MIT's NANDA initiative reveals a stark finding: "95% of organizations are getting zero return" from their GenAI investments, with "just 5% of integrated AI pilots extracting millions in value, while the vast majority remain stuck with no measurable P&L impact" (Challapally, 2025). This crisis is compounded by S&P Global findings that "the share of companies abandoning most of their AI initiatives jumped to 42%, up from 17% last year" with "the average organization scrapped 46% of AI proof-of-concepts before they reached production" (Wilkinson, 2025). The RAND Corporation confirms that "more than 80 percent of AI projects fail; twice the rate of failure for information technology projects that do not involve AI" (Ryseff et al., 2024). The GRC sector faces the same crisis: tools meant to bring clarity are instead obscuring risk, eroding institutional knowledge, and fostering a fragile "checkbox compliance" culture.

95% of organizations are getting zero return from their GenAI investments, while the vast majority remain stuck with no measurable P&L impact.

This whitepaper will argue that despite the hype, an AI-first approach to GRC often creates more risk than it mitigates by generating generic documentation, fostering a dangerous lack of organizational understanding, and misapplying AI technology in ways for which it is fundamentally ill-suited. We will dissect the systemic failures of this model and present a more resilient, human-centric alternative that uses AI not as an unreliable author, but as a powerful analytical co-pilot.

1.1 The Pitfall of Generic Compliance: When "Best Practice" Isn't Your Practice

The foundational flaw in using AI to generate compliance documentation is that these systems produce generic templates by design. Research shows that AI systems consistently perform poorly in compliance tasks, generating false information and contradicting themselves when assessing documentation quality (Sovrano et al., 2024). This creates an immediate conflict with

modern compliance frameworks, which require precision and organizational specificity.

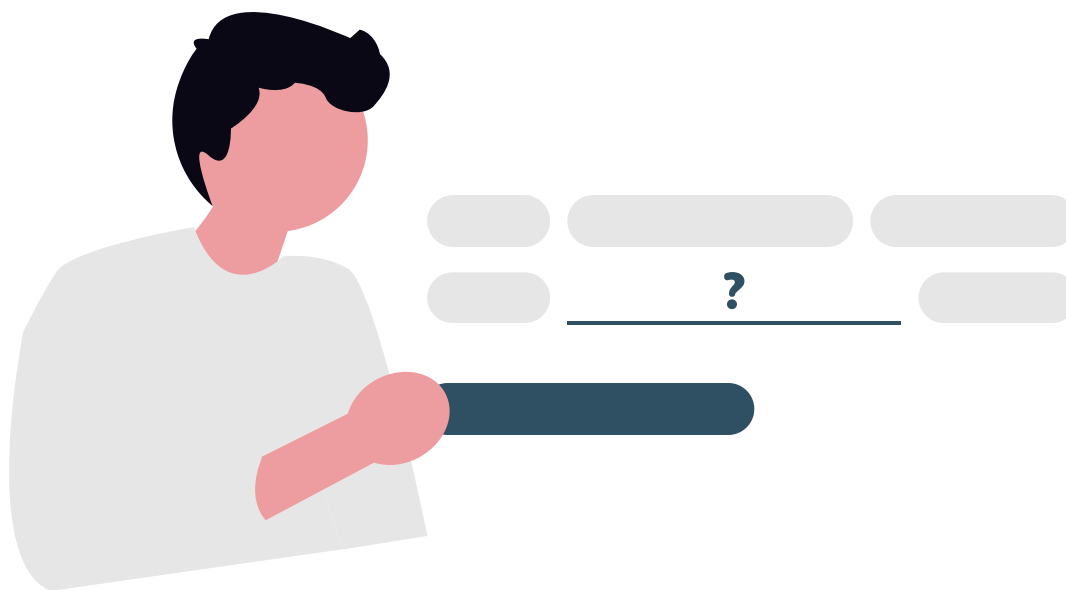
Frameworks like SOC 2 and CMMC are intentionally non-prescriptive. Their value comes from flexibility, requiring organizations to scope audits and define controls based on their unique business processes and risk profiles. Current AI templates fail because they assess systems broadly rather than focusing on specific organizational contexts (Bogucka et al., 2024). Effective compliance isn't about adopting generic "best practices"-it's about creating controls precisely tailored to your company's technology stack, personnel structure, and data handling procedures.

When organizations use AI to generate policies, they're essentially adopting someone else's architecture. This creates a critical disconnect where official policies don't reflect how the organization actually operates. The result is documentation that's useless during real incidents and indefensible during audits-completely undermining the goal of building customer trust through frameworks like SOC 2 (Winecoff & Bogen, 2024).

Industry stakeholders often misunderstand-or miscommunicate-what problem needs to be solved using AI.

1.2 The Erosion of "Compliance Fluency": Outsourcing Thought, Retaining Risk

Beyond generic content, AI-generated compliance documents create a more insidious risk: they encourage "cognitive offloading," allowing leaders to bypass the critical thinking necessary to understand their own GRC program. AI systems often achieve technical correctness while missing what humans actually need (Gabriel et al., 2023). The legal principle of due diligence requires corporate leaders to maintain active oversight of risk systems-a responsibility that cannot be delegated to algorithms. Writing compliance documentation builds institutional knowledge, forcing leaders to confront regulatory complexity. When this work is outsourced to algorithms, organizations fail to build "compliance fluency" - the deep knowledge of why controls exist and how they impact the business.



This lack of fluency inevitably fosters a superficial "checkbox compliance" culture. Research shows that automated compliance systems work only in narrow, low-interpretation scenarios and cannot change organizational culture (Bello y Villarino & Bronitt, 2024). The organizational goal shifts from genuinely securing systems to simply possessing documents that pass cursory review. Teams can point to policies but cannot articulate their rationale or defend implementation under scrutiny. This creates a dangerously fragile security posture that may satisfy checklists but will crumble under sophisticated audits or fail completely during real-world security incidents.

1.3 Misaligned Tools: The LLM as a Flawed Content Generator

The systemic underperformance of AI-first GRC platforms stems from fundamental technological misalignment. These platforms incorrectly deploy Large Language Models as content generators: a task for which they are demonstrably unreliable. LLMs have proven applications, but significant limitations persist when applied to tasks they weren't designed for (Kaddour et al., 2023).

An LLM's true strength lies in acting as a "universal data processor" - parsing unstructured data, translating between formats, and identifying correlations. In these applications, the LLM structures existing information rather than creating new knowledge, allowing output validation against external sources. Hybrid models that combine different memory types provide the validation and transparency that pure generative models lack (Lewis et al., 2020).

Using an LLM to generate security policies fundamentally misuses the technology; here, the LLM functions as a "digital oracle," generating content through statistical probability rather than genuine comprehension. This process is inherently flawed because "LLMs can potentially fabricate erroneous" content, and "the false information hallucinated by LLMs often appears highly plausible, to the extent that even humans may feel hard to detect" (Zhang et al., 2023). This stems from training on "trillions of tokens obtained from the web, making it difficult to eliminate fabricated, outdated or biased information." Deploying this unreliable function as the foundation for compliance is a strategic error that directly contributes to high AI project failure rates.

The false information hallucinated by LLMs often appears highly plausible, to the extent that even humans may feel hard to detect.

1.4 The Human-led, AI Augmented Alternative

The profound failures of AI-first approaches don't require rejecting technology—they demand reframing its role. The most effective path forward combines AI's speed and scale with human judgment and accountability. This framework prioritizes augmentation over replacement, seeking "high levels of human control and high levels of computer automation" to create technologies that "amplify, augment, enhance, and empower people" rather than replacing them (Shneiderman, 2020). Research shows that combining AI systems with human expertise creates hybrid solutions that exceed the sum of their parts (Fabri et al., 2023).



In this model, AI handles what it does best: processing vast datasets, identifying patterns, and automating repetitive monitoring. This frees human experts for irreplaceable tasks requiring contextual understanding, ethical reasoning, and accountability. The Human-in-the-Loop framework ensures experts validate critical outputs, correct errors, and provide feedback that improves reliability over time. Rather than creating bottlenecks, strategic human intervention amplifies value by transforming raw AI output into trusted, legally defensible decisions (Fareedi et al., 2025; Shneiderman, 2020).

Part 2: The Valdyr Solution: Building a Resilient, Human-Led Compliance Program

2.0 Introduction: Pivoting to a Model of Clarity and Context

Having established the profound risks of an AI-first GRC strategy (from generic documentation to the erosion of institutional knowledge), the path forward requires a transition to a stronger and more intelligent alternative. The critical failures of current platforms reveal a clear need for a new approach, one that prioritizes clarity, context, and human expertise over the flawed promise of pure automation. Valdyr.io was engineered to be this solution, born from the recognition that effective compliance requires not just automation, but intelligent orchestration of human insight and technological capability.

Valdyr offers a human-led GRC model that builds deep compliance fluency through a visually-defined tree graph, providing the rich context that AI-first platforms lack. Unlike traditional platforms that treat compliance as a collection of disconnected documents and checklists, Valdyr recognizes that true security emerges from understanding the interconnected relationships between policies, controls, and operational reality. By structuring all GRC elements in a clear organizational hierarchy from root to branches, Valdyr transforms abstract compliance requirements into a tangible, navigable, and understandable forest, empowering teams to see, manage, and master their entire security posture.

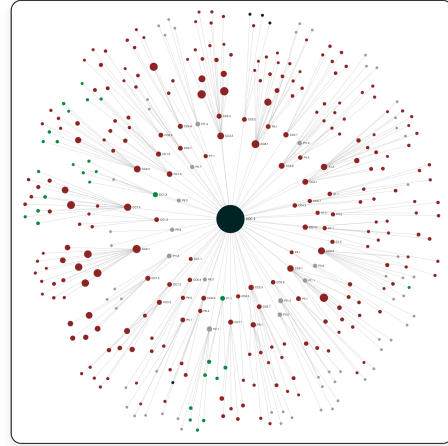


Figure 3: Valdyr Mapping of SOC2

2.1 The Power of the Tree Graph: Visually Mapping the "Why"

At its core, Valdyr is built upon a "Unified Compliance Forest"-a tree graph data model where all GRC elements are structured in a clear hierarchical relationship with the organization as the root. Policies branch from the organizational root, followed by standards, then controls, then evidence and processes, with controls linking directly to frameworks of choice like SOC 2 or CMMC. This creates a living organizational tree that maps the entire compliance program. This architecture provides the fundamental solution to the problems of data silos and context-free documentation that plague traditional GRC tools.

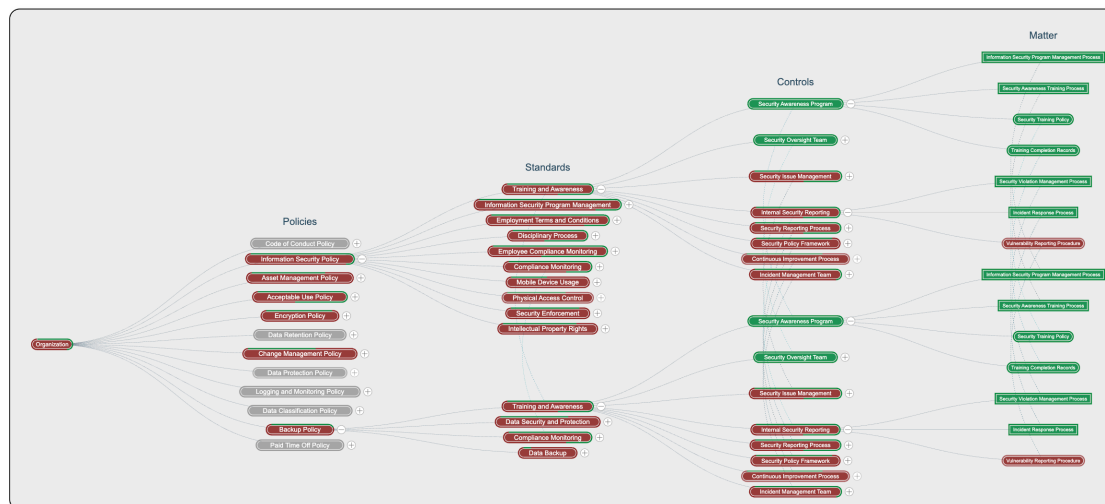


Figure 2: Valdyr's Unified Compliance Forest showing the hierarchical relationship from organizational root through policies, standards, controls, and evidence

This hierarchical approach immediately eliminates the silos inherent in document-centric systems. Instead of existing in separate folders or modules, each policy branches down through its supporting standards, which in turn branch to their implementing controls, which finally connect to the evidence and processes that prove their effectiveness. Controls also branch directly to compliance frameworks like SOC 2 or CMMC, creating clear traceability from organizational root to framework requirements. This creates a clear, navigable tree structure that reflects the true, hierarchical nature of a GRC program.

What's more, the Interactive Compliance Map transforms risk assessment from an abstract exercise into a visual, intuitive process. By representing the entire GRC program graphically, leaders can perform intuitive gap analysis that would be impossible in a spreadsheet or a linear list of controls. It becomes immediately obvious, for example, when a critical asset processing PII is not connected to a required encryption control, or when a high-risk vendor lacks a connection to a signed DPA. These are the kinds of critical, real-world gaps that AI-generated templates and checklist-oriented platforms consistently miss. By allowing users to visually traverse the relationships between every component of their program, this model builds genuine "compliance fluency," empowering them to understand not just that they are compliant, but precisely why and how.


2.2 Beyond the Audit: Integrating Processes and Standards for Real-World Context

Most GRC platforms fail by hyper-focusing on narrow audit objectives. In rushing to satisfy SOC 2 or CMMC requirements, they ignore organizational context and internal processes. This creates compliance programs that are "a mile wide and an inch deep": fragile facades that pass cursory checks but lack operational substance. Traditional systems operate in silos with outdated processes, struggling against evolving regulatory requirements, while integrated approaches enable proactive, adaptive risk management (Bello y Villarino & Bronitt, 2024; Oluoha et al., 2022).

Valdyr's architecture corrects this flaw by treating process documents and internal standards as integral branches within the compliance forest; not attachments, but core structural elements. This provides multi-layered, context-rich compliance views impossible with checklist tools. Current templates lack regulatory grounding and ignore real-world system uses (Bogucka et al., 2024). Valdyr creates an unbroken, auditable path from high-level policy through supporting standards to specific process actions and execution evidence. This integrated approach transforms compliance challenges into strategic advantages, enabling continuous monitoring and faster adaptation to changing regulations (Oluoha et al., 2022). By growing the "how" of process from the "what" of policy, Valdyr builds compliance that's both resilient and reflects actual organizational operations.

2.3 The Right Role for AI: The Human-Led Co-Pilot

Valdyr's philosophy amplifies rather than replaces human expertise by positioning AI as an analytical co-pilot, not a content author. Effective human-AI collaboration creates systems that empower users while maintaining human oversight (Fareedi et al., 2025). The Valdyr AI Co-Pilot uses Retrieval-Augmented Generation (RAG), ensuring outputs are trustworthy and contextually relevant. RAG approaches are more factually grounded, hallucinate less, and provide better control compared to purely generative models (Lewis et al., 2020).


 **Health Assessment**

Health Summary

The "Acceptable Use Policy" is currently in an **unhealthy** state due to several underlying issues. A total of 6 out of 7 standards and numerous controls are unhealthy, primarily because their evidence lacks a history of updates. This indicates that while policies and processes might be in place, there's a lack of ongoing monitoring and record-keeping.

Figure 4: Real-time AI gap analysis providing clarity into why the Acceptable Use Policy is non-compliant

This human-led model eliminates AI hallucination risk. The Co-Pilot exclusively searches the organization's own secure, human-authored compliance forest rather than generating new content from internet sources. This guarantees every answer is grounded in company-specific data and procedures. Since AI hallucinations significantly undermine reliability by appearing plausible when false, systems must support rather than replace human judgment (Zhang et al., 2023; Shneiderman, 2020). The result empowers compliance managers to ask complex questions in plain English and get reliable, contextually grounded answers.



Encryption Policy Review

Based on the provided hierarchy and content, several areas of the encryption policy require attention to ensure effective encryption and compliance with relevant standards.

Unhealthy Standards and Controls

The hierarchy indicates that several standards and controls related to encryption are **Unhealthy**, including:

- Training and Awareness
- Compliance Monitoring
- Data Security and Protection
- Encryption Key Management

Specific Controls to Address

To improve the encryption policy, focus on the following controls:

- Security Awareness Program: Ensure that employees understand the importance of encryption and how to properly use encrypted systems.
- Security Policy Framework: Review and update the encryption policy to reflect current best practices and regulatory requirements.
- Security Oversight Team: Establish a team to oversee encryption practices and ensure compliance with the policy.
- Data Backup Implementation: Verify that encrypted data is properly backed up and can be recovered in case of a disaster.

Action Plan

To address the identified issues, consider the following steps:

1. Conduct a thorough review of the encryption policy and update it to reflect current best practices and regulatory requirements.
2. Develop a training program to educate employees on proper encryption practices and the importance of security awareness.
3. Establish a security oversight team to monitor encryption practices and ensure compliance with the policy.
4. Verify that encrypted data is properly backed up and can be recovered in case of a disaster.

Next Steps

By addressing the identified areas and implementing the recommended controls, you can improve the effectiveness of your encryption policy and reduce the risk of non-compliance.

Sources:

Policy: Encryption Policy

Health: Unhealthy

Standard: Encryption Key Management

Health: Unhealthy

Standard: Data Security and Protection

Health: Unhealthy

Figure 5: The Valdyr AI Co-Pilot in action, helping users identify and address gaps in encryption policy implementation

This approach transforms daily compliance management by using AI for sophisticated analysis rather than unreliable content generation. The Co-Pilot provides daily insights by comparing the real-time compliance forest state against target attestations like SOC 2 or CMMC. It flags gaps by identifying missing organizational tree branches, providing clear, actionable updates on true compliance posture. Well-designed systems positively influence employee decision-making and reduce corruption (Meyer, 2020). This leverages AI's core strength: data correlation to accelerate human insight and enable proactive risk management.



Part 3: Conclusion

The evidence is clear: an AI-first approach to GRC, built on the flawed premise of generative content, is a failing strategy. It creates a dangerous illusion of security, furnishing organizations with generic "checkbox compliance" that is detached from their operational reality. This model actively erodes the critical, institutional knowledge required for a resilient security posture and exposes the business to unacceptable risks stemming from the inherent unreliability of generative AI. The result is a compliance program that is not only ineffective but fragile, incapable of withstanding a rigorous audit or a real-world security event.

Valdyr presents a fundamentally different path forward. Our human-led, visually integrated model is designed to build deep "compliance fluency" across an organization. By structuring the entire GRC landscape as a unified compliance forest, we provide the true operational context that connects high-level organizational policies through their supporting standards, controls, and processes down to the evidence of daily work. In this approach, AI is used responsibly and powerfully; not as an unreliable author, but as an analytical co-pilot, securely grounded in your organizational tree, that accelerates human expertise.

The legal principle of due diligence requires corporate leaders to maintain active oversight of risk systems: a responsibility that cannot be delegated to algorithms.

For organizations seeking a resilient, defensible, and genuinely understood GRC program, the strategic choice is clear. It is time to move beyond the empty promises of AI-generated content and the false security of automated checklists. It is time to invest in a platform built for clarity, context, and the enduring value of human-centric control.



References

- Bogucka, E., Menzel, A., & Imbsweiler, J.** (2024). Co-designing an AI impact assessment report template with AI practitioners and AI compliance experts. *arXiv preprint*. <https://arxiv.org/html/2407.17374v2> (<https://arxiv.org/html/2407.17374v2>)
- Challapally, A.** (2025). *The GenAI Divide: State of AI in Business 2025*. MIT NADA Initiative. https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf (https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf)
- Fabri, L., Häckel, B., Oberländer, A. M., Rieg, M., & Stohr, A.** (2023). Disentangling human-AI hybrids: Conceptualizing the interworking of humans and AI-enabled systems. *Business & Information Systems Engineering*, 65(6), 623-641. <https://doi.org/10.1007/s12599-023-00810-1> (<https://doi.org/10.1007/s12599-023-00810-1>)
- Shneiderman, B.** (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *arXiv preprint*. <https://arxiv.org/pdf/2002.04087> (<https://arxiv.org/pdf/2002.04087>)
- Gabriel, I., Ghazavi, A., Weinberg, A., Birhane, A., Schroeder, R., Pinto, A., & Ding, J.** (2023). Large Language Models as Fiduciaries: A Case Study Toward Robustly Communicating With Artificial Intelligence Through Legal Standards. *arXiv preprint*. <https://arxiv.org/abs/2301.10095> (<https://arxiv.org/abs/2301.10095>)
- Fareedi, A., Mulgun, D., Chaudhry, M. A., & Navas, J. F.** (2025). Enriching Human-AI Collaboration: The Ontological-Service Framework for Value Creation in Conversational AI. *Mediterranean Conference on Information Systems*. https://www.researchgate.net/publication/395352686_Enriching_Human-AI_Collaboration_Value_Creation_in_Conversational_AI (https://www.researchgate.net/publication/395352686_Enriching_Human-AI_Collaboration_Value_Creation_in_Conversational_AI)
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D.** (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *arXiv preprint*. <https://arxiv.org/abs/2005.11401> (<https://arxiv.org/abs/2005.11401>)
- Meyer, M.** (2020). The effectiveness of compliance management systems: An experimental approach. *ResearchGate*. https://www.researchgate.net/publication/275544916_The_Effectiveness_of_Compliance_Management_Systems_-_An_Experimental_Approach (https://www.researchgate.net/publication/275544916_The_Effectiveness_of_Compliance_Management_Systems_-_An_Experimental_Approach)
- Slattery, P., Gordon, R., & Thompson, N.** (2024). Global AI adoption is outpacing risk understanding, warns MIT CSAIL. *MIT Computer Science and Artificial Intelligence Laboratory*. <https://www.csail.mit.edu/news/global-ai-adoption-outpacing-risk-understanding-warns-mit-csail> (<https://www.csail.mit.edu/news/global-ai-adoption-outpacing-risk-understanding-warns-mit-csail>)
- Oluoha, O. M., Odeskina, A., Reis, O., Okpeke, F., Attipoe, V., & Orieno, O. H.** (2022). Artificial intelligence integration in regulatory compliance: A strategic model for cybersecurity enhancement. *Journal of Frontiers in Multidisciplinary Research*, 3(1), 35-46. https://www.researchgate.net/publication/391901838_Artificial_Intelligence_Integration_in_Regulatory_Compliance_A_Strategic_Model_for_Cybersec (https://www.researchgate.net/publication/391901838_Artificial_Intelligence_Integration_in_Regulatory_Compliance_A_Strategic_Model_for_Cybersec)
- Bello y Villarino, J.-M., & Bronitt, S.** (2024). AI-driven corporate governance: A regulatory perspective. *Australian Journal of Corporate Law*, 39(1), 1-25. <https://www.tandfonline.com/doi/full/10.1080/10383441.2024.2405752> (<https://www.tandfonline.com/doi/full/10.1080/10383441.2024.2405752>)
- Winecoff, A. A., & Bogen, M.** (2024). Improving governance outcomes through AI documentation: Bridging theory and practice. *arXiv preprint*. <https://arxiv.org/html/2409.08960> (<https://arxiv.org/html/2409.08960>)
- Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R.** (2023). Challenges and applications of large language models. *arXiv preprint*. <https://arxiv.org/abs/2307.10169> (<https://arxiv.org/abs/2307.10169>)
- Ryseff, J., De Bruhl, B. F., & Newberry, S. J.** (2024). *The root causes of failure for artificial intelligence projects and how they can succeed: Avoiding the anti-patterns of AI*. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA2680-1.html (https://www.rand.org/pubs/research_reports/RRA2680-1.html)
- Sovrano, F., Vitali, F., & Palmirani, M.** (2024). Simplifying software compliance: AI technologies in drafting technical documentation for the AI Act. *Empirical Software Engineering*. <https://link.springer.com/article/10.1007/s10664-025-10645-x> (<https://link.springer.com/article/10.1007/s10664-025-10645-x>)
- Wilkinson, L.** (2025). *AI project failure rates are on the rise: Report*. Construction Dive. <https://www.constructiondive.com/news/AI-project-fail-data-SPGlobal/742934/> (<https://www.constructiondive.com/news/AI-project-fail-data-SPGlobal/742934/>)
- Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Lotufo, J. B., Rome, A., Shi, A., & Oak, S.** (2025). The AI Index 2025 Annual Report. *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University*. https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf (https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf)
- Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Lotufo, J. B., Rome, A., Shi, A., & Oak, S.** (2025). The AI Index 2025 Annual Report. *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University*. https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf (https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf)
- Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S.** (2023). Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *arXiv preprint*. <https://arxiv.org/abs/2309.01219> (<https://arxiv.org/abs/2309.01219>)