Assessing the Maturity of Governance and Compliance Management Systems Over Time: AI *Methodology and Insights*

Kenneth Tombs, Chair, Supply Chain Business Council (ITC), Singapore[1] *kenneth.tombs@sceas.org*
and other contributors.

1. Abstract

Effective Governance, Risk, and Compliance (GRC) schemes or management systems are vital for managing risks and regulatory or standards commitments, increasingly so for micro/smaller or niche businesses. Assessing these systems for audit purposes, quality and effectiveness is challenging, especially in complex environments that demand data completeness, timeliness, and integration. Artificial Intelligence (AI) appears to offer unparalleled opportunities here by combining AI-driven data analytics with judgement making, focusing on key criteria such as completeness, improvement over time, cross-linking, granularity, and effectiveness in relation to ISO and other standards.

Keywords: Management Systems, AI, Judgement Making, Audit, Governance-Risk-Control, International Standards, Compliance, Conformance, Regulation.

2. Introduction

The growing complexity of regulatory and trading commitments with their need for robust risk management, have made Governance, Risk, and Compliance (GRC) systems a critical element for businesses of all sizes. However, smaller or niche organisations often face significant challenges in ensuring their GRC systems are not only functional but also effective and aligned with industry standards such as ISO. The central research question here is: *How can we assess the quality and effectiveness of GRC systems of any scale, in a way that is both comprehensive, repeatable, and truly representative?*

The Goal is to provide a methodology using AI-driven data analytics to evaluate key dimensions of GRC systems—such as data completeness, cross-linking, and granularity—as work towards offering organisations clear insights into enhancing their compliance schemes and improving overall system maturity. Therefore, this work pursues equipping businesses with a structured, evidence-based assessment approach to improving governance, reducing risk, and ensuring regulatory compliance which are essential for long-term success and sustainability. Anticipating the future roles of digital initiatives such as ISO 'Smart' standards is integral to meeting this Goal.

3. Literature Review

There is little publicly available research focussed on methodologies for AI-driven analytics. Two relevant publications are: *RC Systems and Organizational Risk Management* by Racz, Weippl, and Seufert (2010); and *AI-Driven Analytics in GRC Systems* by McKinsey's Consulting. Neither focus on practical implementations.

5. Hypotheses/Objectives

This work's hypothesis (or better described as assumption), is that AI can formulate valid and reliable value judgements from diverse and complex data sets representing a business's governance system; where those judgements inform analysis and decision using an AHP method, with Inconstancy values assuring the AI is consistent and representative.

6. Research Design/Methodology

A hybrid methodology utilises AI-driven functions with an iterative-with-AI approach to establishing an analytical method statement and its refinement, creating an effective and actionable method statement for GRC system evaluation. This approach transformed the foundation of improvements to Time-Cost-Quality for assessments, into a practical tool that businesses can use to enhance their governance, risk, and compliance strategies.

An overview is:

Exploration of Key Questions:

At the outset, we had identified with ISO's Customers Matter Programme the core challenges facing smaller or niche businesses in assessing the quality and effectiveness of their GRC systems. Questions such as "How can we ensure data completeness?", "What are the gaps in cross-linking between risks, controls, and actions?", and "How do we measure system maturity in alignment with ISO and industry standards?" "How do we anticipate deception and fraud?", served as guides for the research. Using AI, we were able to quickly explore volumes of example GRC data, uncover hidden patterns, and assess system weaknesses, helping to address these questions with data-driven insights.

Pursuing Interesting Analytical Paths:

As the AI analysed early datasets, various anomalies and numerous areas of interest emerged, particularly in the fields of data quality, cross-linking, and timeliness. These insights led to deeper investigations, such as identifying specific gaps in the interconnection of risks and controls or determining where data granularity was lacking. Leading to "could we use AHP to analyse data over-time alternatively to of-the-same-time"?

Formulating and Refining the Method Statement:

Throughout the research process, we formulated and continuously refined a method statement to ensure evaluation criteria and reporting were comprehensive and aligned with business need. As the method was iterative not linear, awareness of the implicit risks in iteration was maintained by a comprehensive assurance strategy – testing approximately 20 different aspects of analysis as we went! An ability to assist in the drafting of key metrics—such as data completeness, granularity, timeliness, and cross-linking—enabled us to establish clear, quantifiable measures for assessing GRC system quality. This iterative refinement was vital in shaping a methodology that was not only theoretically sound but also practical for real-world application.

Ensuring Practical Application:

Working through real-world representations of data and exploring various configurations of risks, controls, and actions, we were able to tailor the methodology to provide actionable insights for any organisation or standard. The method offered a possible roadmap for improving GRC system maturity, integrating traditional audit practices with modern, data-driven insights.

7. Results/Model Analysis

With a problematic level of detail to work with, for this paper the focus is on how well has the AI coped with comparisons over time rather than using AHP itself.

AHP generally excels at evaluating static hierarchies of criteria by breaking complex decisions into smaller, manageable comparisons. AHP's strength lies in its structured-modular approach. As the dataset or system evolves over time, it is straightforward for the AI to update the hierarchy or weights without having to completely reconstruct the model. This ensures that the methodology remains flexible and responsive to ongoing changes while preserving its consistency and clarity. Major strengths were seen as:

Flexibility in Criteria Weighting: AHP easily updates the importance of different criteria as the system evolves. If the content and priorities of a management system change (e.g., timeliness becomes more critical than completeness over time), AHP can quickly adapt by reweighting criteria without overhauling the entire structure.

Pairwise Comparisons Adapt to New Data: AHP's pairwise comparison process is scalable and adaptable. As new data or criteria emerge, you can add them to the hierarchy and conduct new comparisons. This ensures the method remains relevant as the dataset evolves.

Consistency Checks for Evolving Data: AHP's built-in consistency checks (like the Consistency Ratio) ensure that judgments remain reliable even when introducing new elements. This feature helps maintain accuracy and coherence in evaluations, even when datasets are in flux.

Dealing with Complexity: AHP is designed to handle complex, multi-criteria decisions, which makes it suitable for tracking often myriads of items denoting progress over time.

Reusability: AHP models are reusable. Once the hierarchy is built it can be adapted over time by adding new criteria/data, updating weights, or introducing new datasets, allowing it to handle longitudinal changes without needing to start from scratch.

8. Conclusions

There has been significant progress in this work of structuring a robust method for assessing GRC systems. By refining key definitions, implementing a dual-level reporting strategy, and emphasising the importance of measurable criteria/data, there are strong foundations for future AI judged assessments. However, the ongoing refinement of thresholds and objective measures will be crucial to further improving the methodology's accuracy and consistency. The world of business is not static, nor can this method be so.

Broadly the Goal and objectives were achieved; its now a matter of turning this into a scheme deployable at scale and usable by all.

Out of the box the AI could perform the tasks and analysis, what followed was how to be confident in that analysis, with of any recommendations and opinions that followed.

From this ongoing work important conclusions were drawn about how to assess and improve Governance, Risk, and Compliance (GRC) systems, as well as how to structure an assessment methodology for maximum effectiveness:

The Method Statement Rules: The temptation to allow the AI to go-its-own-way as a 'black box' is a hidden bias, like an employee, its sometimes problematic to ensure they follow the business rules and desired analytics. The same applies to AI.

Dual-Level Reporting Improves Usability: By creating two levels of reporting based on the same analysis —one for business management and another for compliance and technicians. This enhances the overall usability of subsequent analysis and ensures that recommendations are actionable at both levels.

Clear Definitions and Thresholds Prevent Misclassification: The need for robust category definitions and explicit thresholds was highlighted when an early dataset was misclassified. Resulting in the AI challenging a working assumption, leading to an improvement! This underlined the importance of having clear, quantifiable criteria for classifying datasets finalised here as: Incomplete, Improving, Functional, or Professional. This ensured consistent analysis, reduced subjectivity, and improved reliability.

Granularity and Cross-Linking Are Critical for Maturity: Through testing and refinement, detailed descriptions were identified; (granularity) showing strong cross-linking between risks, controls, data richness, field numbers, and actions are key indicators of system maturity. Datasets lacking these elements struggle to rise above the Improving category, regardless of other aspects like timeliness or completeness.

Timeliness and Control Effectiveness Need Objective Measures: Timeliness and effectiveness were shown to be areas where human subjective judgment created inconsistencies. Establishing objective measures, such as delay thresholds and clear criteria for control effectiveness, will improve the accuracy and fairness of the assessments.

Continuous Improvement Loop is Essential: The importance of a feedback loop was reinforced, ensuring that insights from each assessment feed into improving both the GRC system and the AI method statement itself. This continuous feedback mechanism helps evolve the system over time, preventing stagnation and driving progressive improvement.

Context-Sensitive Weighting of Criteria Adds Flexibility: The suggestion to allow the AI flexibility in how criteria are weighted based on context (e.g., industry risk profile, organizational goals) has been critical in ensuring the method is adaptable. This allows for a more tailored analysis that reflects the unique challenges and priorities of different sectors.

Fit-for-Purpose Evaluation Highlights the Need for Refinement: While the method is fit for service beta testing purposes, the work surfaced many areas for refinement, particularly around the use of quantitative measures, such as percentages of missing data or overdue items. These refinements will prevent subjective interpretation from undermining the accuracy of the classification process long-term.

9. Confidence in Method Statement

With the test strategy in place, the method statement approach provides a solid, reliable framework for assessing GRC systems with a high degree of accuracy, especially in identifying data quality issues and guiding system improvement. Here two separate AI were used, one as tester (Claude) and the other as testee (Leonard), broadly the two tools were interchangeable in reporting on their counterpart.

Here is a brief overview of how well the method statement approach performs in terms of accuracy and reliability:

Accuracy:

Clear Criteria: The method uses well-defined criteria (e.g., completeness, cross-linking, granularity, timeliness), which helps to ensure assessments are objective and based on measurable data points.

Thresholds and Scoring: By introducing specific thresholds (e.g., percentages for completeness or timeliness), these reduced the risk of misinterpretation, improving the accuracy of classifications such as Incomplete, Improving, Functional, and Professional.

Granularity and Detail: The method statement captures detailed aspects of data quality, such as the quality of descriptions and the presence of cross-linking between risks, controls, and actions, ensuring that subtle gaps[i] and edges[ii] in the dataset are detected.

Opportunities for Improvement (OFIs): The identification of targeted OFIs ensures the method not only assesses status, it also provides specific, actionable recommendations for improving data quality and management system performance.

Reliability:

Consistency Across Assessments: The method is designed to produce consistent results by using standardised criteria and a scoring/grading system. This minimises subjectivity-risks and ensures that different analysts could apply the method with similar outcomes.

Dual-Level Reporting: By creating both high-level summaries for management and detailed technical reports for compliance managers and system developers, we ensure all levels of stakeholders receive reliable, tailored information suited to their needs.

Feedback Loop: The inclusion of a continuous improvement loop ensures lessons learned from each assessment cycle are incorporated into future evaluations, improving reliability over time.

Scalability: While there are some limitations regarding scalability for very large datasets, the method is reliable when applied to structured, well-defined datasets and can be adapted to larger systems through automation.

Potential Sources of Variability:

Subjectivity in Certain Criteria: There remains some subjectivity, particularly in areas like GRC controls effectiveness and written-description-professionalism, where qualitative judgment is required. However, this has been minimised by adding guidelines and thresholds to the method statement.

Data Quality Dependency: The method's accuracy is highly dependent on the quality of the data input. If data is missing, incomplete, or inconsistent, the results of the analysis ‘might’ be skewed, this is for further research.

10. Limitations

While this method is well-structured and robust for many scenarios, these limitations highlight the need for flexibility, scalability, and support for various user levels. By addressing these areas, the method statement currently of 19 A4 pages, can simplified and further refined to reduce subjectivity, improve scalability, and ensure consistent application across different contexts:

Subjectivity in Judgments

Limitation: Despite efforts to define clear criteria and thresholds, subjective interpretation remains, especially in areas like GRCs control effectiveness, description-professionalism, or the evaluation of cross-linking.

Mitigation: Stronger quantitative measures (like scoring systems) and clearer documentation of assumptions can reduce subjectivity, but subjectivity cannot be eliminated entirely as for humans.

Over-Reliance on Thresholds

Limitation: While thresholds help define levels like Incomplete, Improving, etc., they can sometimes oversimplify complex data quality issues. For example, a dataset could meet a threshold for completeness but still lack sufficient depth in key areas.

Mitigation: Complement thresholds with qualitative checks or allow more flexibility in how thresholds are applied based on the context of the management system being assessed.

Granularity of Feedback

Limitation: The method might produce very granular feedback, especially for the technical development report. While this is beneficial for system builders, it might overwhelm teams with too many details or suggestions, particularly in large or complex GRC systems.

Mitigation: Prioritising recommendations and highlighting the most critical areas first can help prevent information overload.

Scalability for Large Datasets

Limitation: Applying the method to very large datasets with thousands of risks, controls, and actions might be time-consuming and difficult to manage. The process could become cumbersome if the datasets are not well-structured or if automated tools are absent.

Mitigation: Automating parts of the pre-analysis checks (e.g. completeness checks, cross-linking verification), would improve scalability and reduce manual effort. A hybrid of workflow or process driven analytics may alleviate this need for scale by using ‘micro-bites’ of AI to build the final matrices.

Dynamic and Evolving Systems

Limitation: GRC systems should be dynamic and evolve over time. The method may not always account for rapidly changing data or shifting priorities, particularly in high-risk or fast-moving environments where new risks, controls, or actions are frequently introduced.

Mitigation: Establish a more frequent review cycle of the method statement to capture these changes, or integrate real-time monitoring where possible.

Data Quality Dependencies

Limitation: The method heavily relies on the quality of the data input into the GRC system. If the data is inconsistent, poorly maintained, or incomplete, the analysis might flag as issues symptoms of data entry problems rather than deeper management system flaws.

Mitigation: Emphasise the need for inbuilding regular data quality audits and integrate data cleansing steps as part of the GRC system’s pre-analysis operational routine.

Handling of Context-Specific Risks

Limitation: The method might struggle to account for industry-specific or organisation-specific nuances in how GRC risks, controls, and actions are evaluated. For instance, what is considered timely or granular in one industry, may not be appropriate in another.

Mitigation: Adapt the criteria weighting and assessment framework to reflect the specific business context of the GRC system, allowing for learning by industry or organisation-specific customisation.

Integration with Other Management Tools

Limitation: The method is focused on GRC systems, but modern risk and compliance management often integrates with other tools and systems (e.g., incident management, audit platforms as iGRC). The method may not fully capture the dependencies and interconnections between these systems.

Mitigation: Expand the method to include an assessment of integration quality between the GRC system and other related tools, ensuring a GRC system isn't assessed in isolation.

Time-Consuming for Smaller Teams

Limitation: For smaller teams or organisations with limited resources, applying both the high-level and detailed analysis could be too time-consuming. They may find it difficult to maintain a consistent review cycle with limited staff.

Mitigation: Offer a streamlined version of the method for smaller teams, where only critical areas are assessed unless a more in-depth analysis is necessary.

Training and Expertise Requirements

Limitation: The method requires a certain level of human expertise to apply correctly, especially for system developers or administrators handling the detailed technical analysis. If the human assessors are inexperienced or unfamiliar with GRC frameworks, there could be misinterpretations or inconsistencies by tending to always believe the AI's reports.

Mitigation: Provide training materials and example datasets for teams to practice with, or develop standardized templates to guide less experienced users through the process.

11. Contributions and Thanks

Special thanks to:

Prof. Luis Vargas for grounding the later stages of work for how you assure that when the AI says 'X' , it is actually 'X'.

Shashi Bhattarai for his observations on the method overall using AHP.

Individuals who assisted with thinking, experience and sense-checks for the work:

David Rogers – Deputy Chair of SCBC – maturity modelling and structure.
Rafael Molho – Deputy Chair of SCBC - service approach.
Ron Nolasco – Senior Trade Commissioner – keeping this aligned with ITC and business need.
James Young – CEO ICTTM – developing the e-certification and e-assurance fit.
Marcus Tipler – Early AI experiments and practice.

Steve Brown – Architecture and deployment.
RedNao – AIO Forms package and speciality code.
John Dudmesh – Development approach and Q&A.
Ashleigh Stevens - Word Press usage.
Dr Nathaniel Payne – Record management by block chain methods.
Barry Welsford-Lippiatt – Usability and function.

A special thanks to:

BusinessOptix Limited an Anglo-American global leader in process management, knowledge and management systems; and

GRC One Limited another global leader in corporate management systems for GRC purposes.

Both businesses together provided insights and experience that made this work possible.

8. Key References

ISO Customers Matter Programme: https://www.iso.org/strategy2030/key-areas-ofwork/standard-users-needs.html

Prompts, Workflow and Method Statements – Making AI Work For Using AHP For Governance Risk Controls Assessment Purposes – December 2024 – Supply Chain Business Council – accompanying paper.

[i] Gaps literally means missing data or fields that might otherwise be expected.
[ii] Edge means where a fractional change in analysis cause a major change in reporting for no apparent reason.