GenAI as a Catalyst for Continuous Business Capability Assessment

Master’s Thesis for Master of Science (MSc) in IT Management

Ralfe Poisson | 16 April 2026 | ralfepoisson@gmail.com

Contents

Abstract

Organisations operating in dynamic, technology-driven environments face persistent challenges in translating strategic intent into effective execution. Enterprise Architecture (EA) provides a structured approach to aligning business and technology, using business capability maturity assessment as a key mechanism for evaluating organisational readiness and guiding transformation. This study investigates using Generative Artificial Intelligence (GenAI) to support continuous business capability maturity assessment to improve strategy execution. This is a pragmatic naturalistic study using structured quantitative instruments to capture subjective perceptions through an online questionnaire of 55 practitioners across a range of industries to examine current assessment practices and perceptions of GenAI. The findings indicate that, within the sample, capability maturity assessments are generally performed infrequently and with limited scope. Constraints are driven not only by time and effort, but also by organisational factors such as lack of management buy-in and limited perceived value of current practices. While respondents recognise the potential value of more frequent assessment, perceptions of GenAI are mixed. GenAI is seen as capable of improving analytical consistency and scalability, but concerns regarding data quality, contextual understanding, and stakeholder trust significantly limit its perceived acceptance. A key insight is the gap between the realised value of current assessment approaches and the intrinsic value of more frequent, data-driven assessment. The study contributes to the literature by proposing a conceptual integration of EA and GenAI, and by introducing a distinction between realised and intrinsic value in maturity assessment practices. This reframes capability maturity assessment from a static diagnostic activity toward a continuous, strategy-aligned management mechanism. Practically, the findings suggest that GenAI is best positioned as an augmentative tool within a hybrid assessment approach, rather than as a fully autonomous solution.

1. Introduction

1.1 Background and Context

Organisations operate in environments characterised by continual technological, organisational, and competitive change, creating persistent challenges in translating strategic intent into effective execution. While strategies may be clearly articulated, their operationalisation is often hindered by complexity, misalignment, and limited visibility into organisational capabilities. Enterprise Architecture (EA) has emerged as a management discipline intended to address these challenges by providing a structured representation of the enterprise and enabling alignment between business and information technology. (Simon et al., 2013)

Within this context, business capabilities provide a key mechanism for linking strategy to execution. They describe an organisation's ability to achieve specific outcomes through the coordinated deployment of people, processes, and technology, making their current state and development trajectory critical for informed decision-making (Acar & Zehir, 2010).

Capability maturity assessments are widely used to evaluate how effectively capabilities support strategic objectives, typically through discrete maturity levels that guide improvement efforts. (Vieira *et al.*, 2014; Stoiber *et al.*, 2023) However, existing approaches exhibit structural limitations. They are often conducted infrequently, resulting in static snapshots that fail to reflect the dynamic nature of organisational change. In addition, their reliance on process-centric measures and expert judgement may introduce bias, privileging procedural compliance over outcome effectiveness (Stoiber *et al.*, 2023; Korsten et al., 2024).

These limitations create a need for more continuous and scalable approaches to capability assessment. In this context, recent advances in Generative Artificial Intelligence (GenAI) introduce new possibilities for automating and augmenting complex analytical tasks. However, its application also introduces new challenges related to reliability, bias, transparency, and stakeholder trust. This creates a fundamental tension: while GenAI may improve scalability and consistency, it may also limit acceptance and practical adoption.

This issue is particularly timely given the rapid advancement and organisational adoption of GenAI, which creates, for the first time, a practical opportunity to automate and scale aspects of capability maturity assessment that have historically depended on infrequent, expert-driven evaluation.

1.2 Problem Statement

Despite the recognised importance of capability maturity assessment in supporting strategy execution, current approaches remain constrained by several structural limitations. They are typically infrequent, resource-intensive, and dependent on subjective expert judgement, resulting in static and potentially outdated insights. (Vieira *et al*., 2014) Furthermore, existing frameworks often fail to capture the multi-dimensional nature of capabilities, leading to misalignment between what is measured and what is strategically relevant.

These limitations create a gap between the need for timely, actionable insight into capability maturity and the ability of current approaches to provide it, raising the question of whether emerging technologies such as GenAI could enable more frequent, scalable, and contextually relevant assessments.

1.3 Research Aim

The aim of this study is to investigate whether and how Generative Artificial Intelligence could address limitations in current business capability maturity assessment practices by enabling more frequent, scalable, and strategically useful forms of assessment.

1.4 Research Questions

This study adopts an exploratory, quantitative research approach guided by research questions designed to investigate the research objectives outlined above. The study is structured around the following research questions:

RQ1: How frequently do organisations currently perform business capability maturity assessments?
RQ2: What factors influence the frequency of capability maturity assessments, particularly in terms of effort and time required?
RQ3: To what extent is Generative Artificial Intelligence perceived as capable of enabling more frequent or continuous capability maturity assessments?
RQ4: How are GenAI-based capability maturity assessments perceived in terms of objectivity and stakeholder acceptance compared to traditional human-led approaches?
RQ5: How do stakeholders perceive the value of more frequent capability maturity assessments for improving strategy execution?

1.5 Significance of the Study

This study is situated at the intersection of EA, capability maturity assessment, and Generative AI, and contributes to both academic and practical domains. From an academic perspective, it addresses a gap in the literature by exploring the application of GenAI to capability maturity assessment, an area that has received limited empirical attention. By integrating insights from multiple research streams, the study contributes to the understanding of how emerging technologies may reshape established management practices.

From a practical perspective, the research is directly relevant to organisations seeking to improve strategy execution in increasingly dynamic environments. If GenAI can enable more frequent and scalable maturity assessments, it may provide decision-makers with more timely and actionable insights into organisational capabilities.

2. Project evaluation and specification

2.1 Scope and Framing of the Research

The study is deliberately scoped to focus on perceptions and feasibility, rather than the development or technical implementation of a working GenAI solution. The project is therefore positioned as an exploratory, practice-oriented investigation, examining whether GenAI could realistically address the limitations of current maturity assessment approaches. In particular, the study focuses on three core dimensions:
(i) the current frequency and limitations of capability maturity assessments,
(ii) the effort and constraints associated with existing approaches, and
(iii) stakeholder perceptions of the potential value, objectivity, and acceptance of GenAI-enabled alternatives.

2.2 Research Context and Target Population

The research is situated within the professional domain of EA, with the target population consisting primarily of Enterprise Architects and related practitioners who are directly involved in capability modelling, maturity assessment, and strategy execution activities. This focus reflects the assumption that such practitioners possess both the contextual knowledge and practical experience required to provide informed perspectives on current assessment practices and their limitations. The study leverages the researcher’s professional network, including LinkedIn and specialist EA communities, to access this population.

2.3 Feasibility, Risk & Ethics Assessment

The feasibility of the research is determined by scope, participant access, and methodological design. The study is scoped to assess the perceived feasibility and value of GenAI-enabled maturity assessment rather than implementing a technical solution, reducing complexity while preserving conceptual contribution. Access to participants is supported through professional networks and an online questionnaire enabling asynchronous participation. The structured survey design, complemented by limited qualitative input, balances comparability and interpretability within project constraints. Overall, the study is feasible within the given timeframe.

The primary risks relate to sample representativeness, response validity, researcher bias, and the evolving nature of GenAI. A specialised target population may limit representativeness, mitigated through targeted recruitment via professional networks. Self-reported survey data introduces risks of misinterpretation and conceptual ambiguity, addressed through structured questionnaire design and limited reliance on qualitative inputs. Researcher bias was mitigated through data-grounded interpretation and cross-checking of findings. Finally, rapid developments in GenAI introduce temporal risk, addressed by clearly situating the study within a defined timeframe.

The research adheres to ethical principles of voluntary participation, informed consent, and data protection. Data was collected via an online questionnaire from consenting participants, who were informed of the study’s purpose, data usage, and right to withdraw. All data was anonymised, with no personally identifiable information included in analysis. Contact details were collected only for follow-up and stored securely. The study involved minimal ethical risk.

3. Literature review

3.1 EA and Corporate Strategy

EA structures enterprise elements and their relationships, serving as a fundamental management system. A mature EA practice has been shown to reduce operating costs, improve project execution, and enhance alignment between business and Information Technology (IT). (Simon *et al*., 2013; Kappelman *et al*., 2009)

Despite extensive academic and practitioner attention, strategy implementation continues to present significant challenges. Organisations often struggle to translate strategic intent into operational outcomes, frequently due to ineffective strategy communication. (Cater & Pučko, 2010; Simon *et al*., 2013). This finding is particularly noteworthy given that one of the core functions associated with EA is the communication of organisational objectives and structural relationships across the enterprise (Kappelman *et al*., 2009). However, while these studies identify communication as a failure point, they do not explain how EA artefacts concretely improve communication effectiveness, leaving a gap between conceptual role and operational mechanism.

Market and business environments are subject to continuous technological, economic, and organisational change, rendering strategy a moving target rather than a static plan. Radeke (2011) argues that firms must therefore achieve and sustain high levels of organisational flexibility to support dynamic strategies. Persistent misalignment between IT and business functions remains a major impediment to such flexibility, reinforcing the role of EA as a mechanism for managing organisational complexity while fostering more agile forms of execution.

One challenge concerns the lack of consensus around the definition of Enterprise Architecture itself. While some authors conceptualise EA primarily as an organising logic between IT and business processes (Ross *et al*., 2006), others adopt broader definitions that encompass current and future business objectives, goals, visions, and strategies (Bernard, 2005; Pereira & Sousa, 2005), and others still emphasise EA as a management system, highlighting the tension between operational and strategic interpretations of EA. (Simon *et al*., 2013)

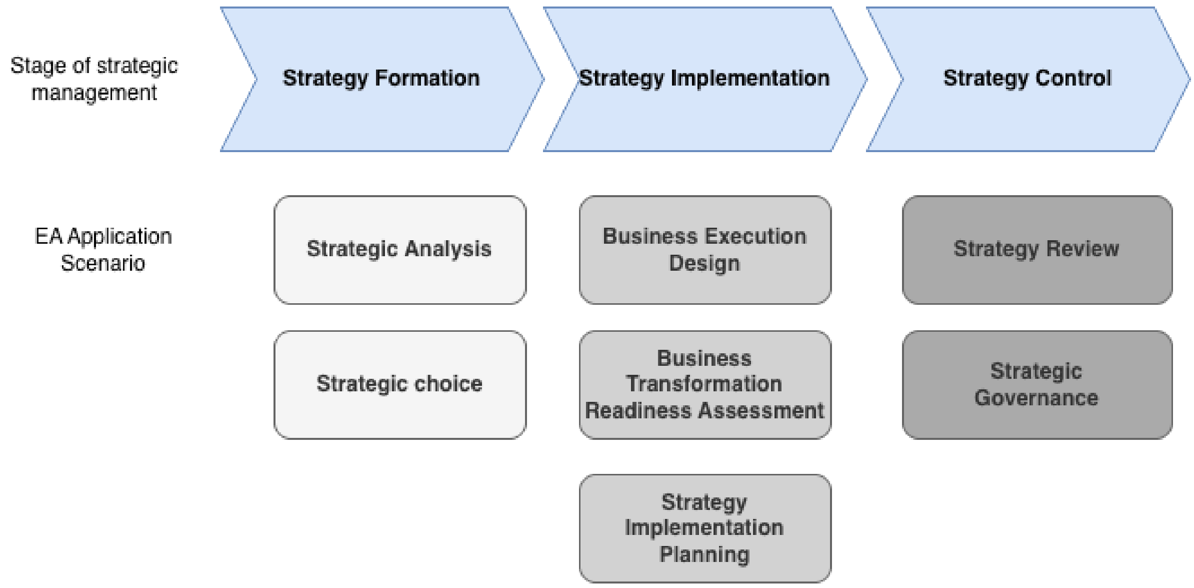


Figure 1: Strategic management stages and related EA application scenarios. Reproduced from Simon et al. (2013)

3.2 Business Capabilities

Business capabilities represent a central analytical artifact within EA and strategic management. A business capability may be defined as a firm's ability to achieve a specific business outcome through the coordinated deployment of people, processes, and technology (Keller, 2009). Unlike organisational structures or processes, capabilities provide relatively stable, outcome-oriented abstractions that enable strategic reasoning across organisational boundaries.

A substantial body of literature argues that firms achieve competitive advantage primarily through the strategic development, diversification, and deployment of capabilities (Offermann *et al*., 2017; Stoiber *et al*., 2023; Korsten *et al*., 2024). Firms differ in their ability to evolve and recombine their capabilities over time. Consequently, the diversification and strategic deployment of business capabilities is closely associated with superior organisational performance.

Empirical evidence supports this relationship between capability diversification and performance. In a large-scale study of 445 manufacturing firms, Acar and Zehir (2010) demonstrate that business capabilities strongly mediate the relationship between leadership strategy and financial performance. Their findings suggest that firms which actively develop and diversify their capabilities achieve greater efficiency and higher financial performance than their rivals. This effect is attributed to the firm-specific nature of capabilities, which are often developed through learning and repetition and are therefore difficult for competitors to imitate. While Acar and Zehir (2010) emphasise capability diversification, other perspectives emphasise adaptability rather than breadth, suggesting competing interpretations of how capabilities drive performance. (Korsten *et al.*, 2024)

Business capabilities also play a critical role during periods of organisational transformation. During such transformations, organisations must understand both the current state and developmental trajectory of their capabilities in order to make informed strategic decisions (Offermann *et al*., 2017). Reflecting this importance, research highlights that organisations with mature EA practices consistently leverage capability models as a key management tool to support strategic decision-making (Brits *et al*., 2007; Riege & Aier, 2009; Keller, 2009). These findings assume that capabilities can be clearly identified and measured; an assumption challenged by the conceptual fragmentation identified by Offermann et al. (2017)

However, the existence of more than twenty distinct capability definitions (Offermann et al., 2017) raises questions about construct stability. If capabilities cannot be consistently defined, then any maturity assessment built upon them risks measuring fundamentally different constructs across organisations. This undermines both comparability and the validity of maturity scoring, an issue that remains largely unaddressed in existing frameworks.

3.3 Capability Maturity Assessments

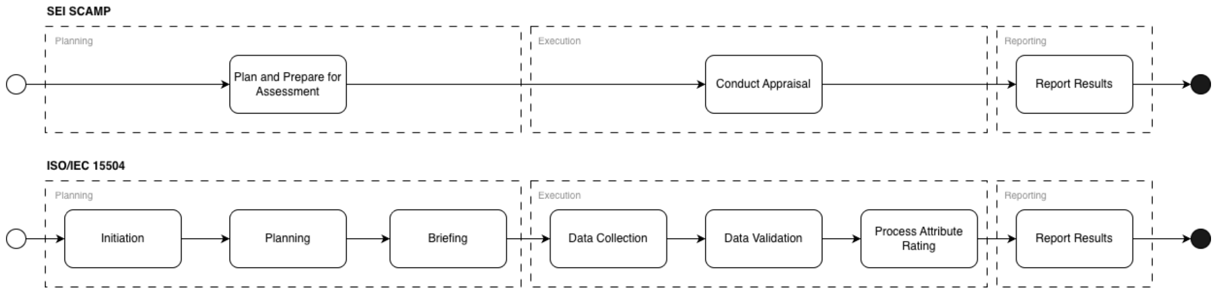


Figure 2: Assessment process comparison of SCAMPI vs. ISO 15504 (Adapted from Proenca & Borbinha, 2018)

If capabilities drive performance, then assessing their maturity becomes strategically critical. (Korsten *et al.*, 2024) Capability maturity assessments are employed to evaluate the extent to which organisational capabilities are developed, stable, and effective in supporting strategic objectives. Maturity modelling has emerged as a management practice aimed at guiding organisations in the development and improvement of business capabilities, typically through a sequence of discrete maturity levels (Korsten *et al*., 2024). Such assessments have been shown to provide valuable insights that support strategy implementation and organisational development (Stoiber *et al*., 2023). While maturity models aim to provide structured evaluation, their reliance on discrete levels contrasts with the continuous nature of capability evolution, suggesting a conceptual mismatch between model structure and organisational reality.

The lack of definitional consensus discussed in Section 3.1 further undermines comparability across maturity assessments and threatens construct validity.

A similar lack of consensus exists regarding the concept of "capability." Offermann *et al*. (2017), in a systematic literature review of 355 abstracts, identify more than twenty distinct clusters of capability definitions. This conceptual fragmentation complicates the design and interpretation of capability maturity assessments. In response, Offermann *et al*. propose a synthesised definition of capability as "a particular ability that a business may possess or exchange to achieve a specific corporate goal," which is adopted in this study.

Existing maturity assessment frameworks exhibit further limitations when applied to business capabilities. Vieira *et al*. (2014) describe maturity assessments as mechanisms that translate a domain of interest into a set of indicators collectively accepted by stakeholders as relevant for decision-making. However, they highlight a heavy reliance on competent assessors, who may be scarce or external to the organisation. This reliance raises concerns regarding the representativeness, transparency, and interpretability of assessment results for decision-makers. While Vieira *et al.* (2014) attempt to formalise assessment through indicators, this contrasts with assessor-driven approaches, highlighting a trade-off between consistency and interpretive richness.

The most prevalent maturity assessment frameworks (including CMMI, SCAMPI, ISO/IEC 15504, ISO/IEC TS 33030, and COBIT 5) are predominantly process-centric (Vieira *et al*., 2014; Stoiber *et al*., 2023). These frameworks assess maturity primarily in terms of process formalisation and control, rather than considering broader capability dimensions such as people, technology, and governance (Keller, 2009). Korsten *et al*. (2024) emphasise that misalignment between maturity level characterisations and the underlying construct being assessed undermines the validity of such evaluations, and further note that the literature lacks a theoretically grounded model for defining maturity levels specifically for business capabilities. This divergence suggests that existing maturity frameworks may systematically misrepresent capability maturity by privileging process formalisation over outcome effectiveness.

Vieira et al. (2014) propose a query-based, quantitative approach to maturity assessment in order to reduce dependence on expert assessors. While this approach shifts judgement toward formalised metrics, the authors acknowledge that it introduces a different form of complexity by placing significant responsibility on the metamodel designer. They further question whether such models would be accepted by business stakeholders and whether results would remain interpretable in practice.

A persistent tension within the maturity assessment literature arises from the divergence between process-centric and capability-centric frameworks. Established models such as CMMI and ISO/IEC 15504 conceptualise maturity primarily as the formalisation, control, and optimisation of processes, thereby equating organisational advancement with increasing procedural standardisation and compliance. By contrast, capability-centric perspectives define capabilities as outcome-oriented, integrative constructs that encompass not only processes, but also people, knowledge, governance structures, and technological resources. (Korsten et al., 2024) From this viewpoint, maturity cannot be reduced to process control alone, as the effectiveness of a capability depends on the coordinated orchestration of multiple organisational dimensions.

This reveals a conceptual inconsistency between process-centric and capability-centric interpretations of maturity, suggesting that existing models may systematically misrepresent organisational capability.

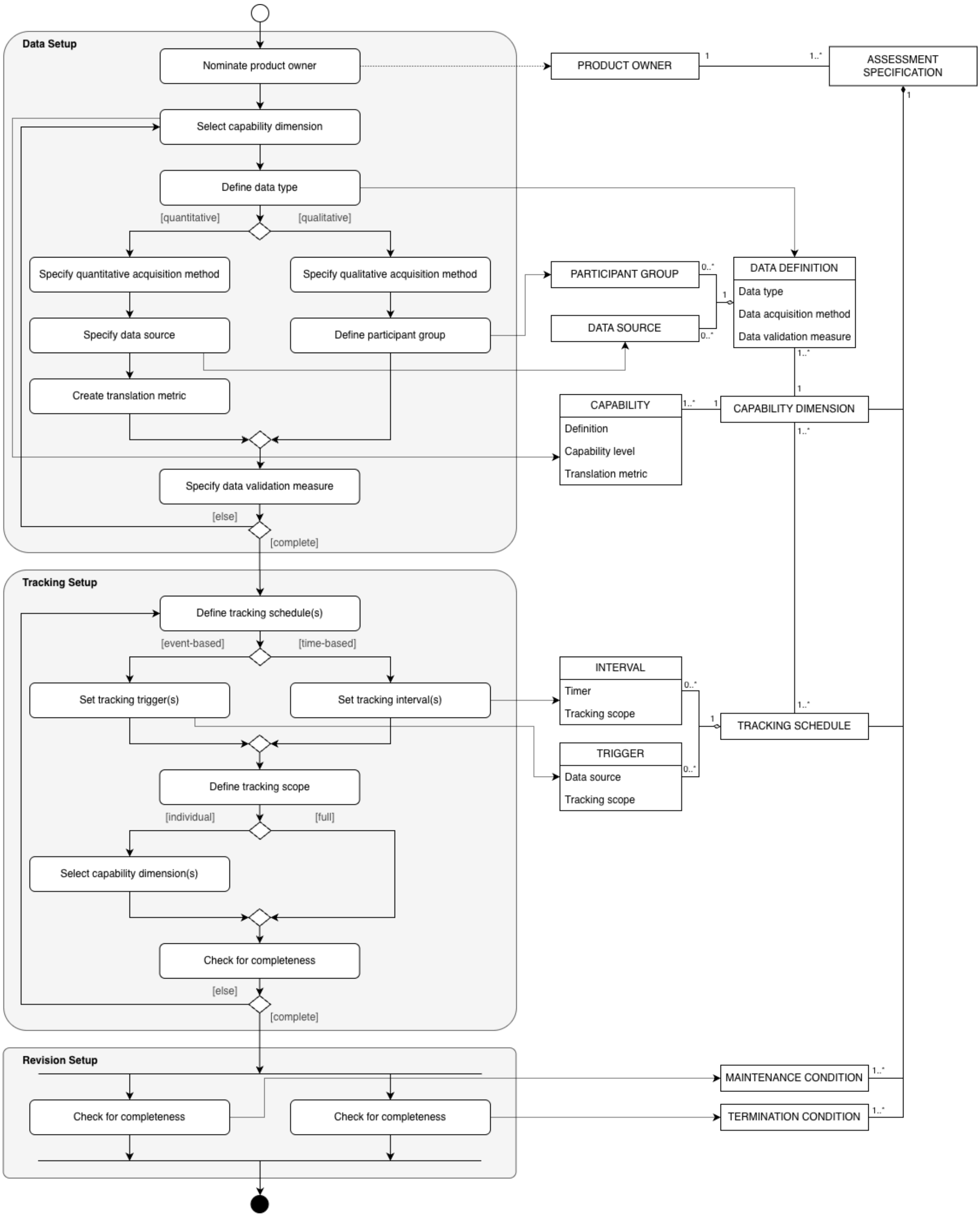


Figure 3: CMAM Assessment Specification Metamodel (Adapted from Stoiber et al., 2023)

Taken together, the literature reveals a structural misalignment between what maturity models measure and what organisations seek to understand. Process-centric frameworks equate maturity with formalisation and control, whereas capability-based perspectives emphasise outcome effectiveness and adaptability. This suggests that existing maturity models may systematically overestimate organisational capability by privileging procedural compliance over strategic impact. Consequently, the validity of maturity assessments as indicators of strategic readiness remains questionable.

3.4 The Need for Continuous Assessment

In dynamic environments, business capabilities co-evolve with technological and economic change rather than reaching a static state. Stoiber *et al*. (2023) argue that since maturity co-evolves with capability development, the notion of a one-off or infrequently repeated maturity assessment stands in tension with the fundamentally dynamic nature of organisational reality.

In their comparative assessment of seven widely used maturity frameworks, Stoiber *et al*. (2023) identify two systemic weaknesses. First, existing frameworks struggle to reconcile generality with comprehensiveness. Second, they lack mechanisms supporting iterative or continuous application. For example, SCAMPI recommends assessments at intervals of up to three years, a cadence that is poorly aligned with the modern pace of organisational change.

To address these weaknesses, Stoiber *et al*. (2003) propose the Continuous Maturity Assessment Model (CMAM), designed to support iterative assessment through traceable and partially automatable mechanisms. CMAM accommodates both qualitative and quantitative assessment methods and demonstrates significant reductions in assessment effort. In a six-month *in situ* study, assessment duration was reportedly reduced from two years to two months through the use of the CMAM framework. However, this result is not examined in detail, and several limitations remain. The authors do not clearly define what constitutes “continuous” assessment, focus primarily on software development process maturity, and make largely unsupported claims regarding inherent objectivity of continuous assessments. Furthermore, the use of Design Science Research methodology prioritises solution suitability over optimality, limiting generalisability. Despite these limitations, the CMAM framework appears to be a suitable and theoretically grounded framework in which to embed GenAI to support qualitative assessments which complement the already automated quantitative assessments. If continuous assessment requires scalable qualitative interpretation, automation of interpretive processes becomes necessary, which creates an opportunity for GenAI. Although CMAM reduces assessment duration, it does not eliminate reliance on human interpretation, distinguishing it from more fully automated approaches such as MMArch. This creates a structural requirement for technologies capable of scaling qualitative interpretation, positioning GenAI as a potential solution space rather than a direct solution.

Proença and Borbinha (2018, 2019) propose the Maturity Model Architect (MMArch), an automated maturity assessment framework that supports the definition, storage, and evaluation of maturity models using ontologies and reasoning engines. MMArch enables automated inference of maturity levels but relies on heuristic rules and does not incorporate adaptive intelligence, limiting its ability to evolve or learn from assessment outcomes.

When compared to CMAM from Stoiber *et al*. (2003), MMArch advances formal logical automation whereas CMAM advances organisationally embedded iteration. Together, they provide complementary approaches to improve rigour and repeatability of maturity assessments, though neither fully addresses the interpretive and epistemic challenges inherent in qualitative capability evaluation. In particular, Stoiber et al. (2023) do not fully resolve how qualitative judgement can be scaled without increasing complexity.

3.5 GenAI in EA

Generative Artificial Intelligence (GenAI) introduces new possibilities for automating and augmenting qualitative analysis within EA. (Haki *et al.*, 2025) Large Language Models (LLMs) are capable of processing unstructured data, identifying patterns, and generating insights, suggesting potential for improving the scalability of maturity assessments (Mahadevkar *et al.*, 2024, Nguyen & Welch, 2025). However, these capabilities are accompanied by significant risks, including hallucination, bias, lack of transparency, and limited contextual understanding (Mahadevkar et al., 2024).

Despite the increasing popularity of GenAI, it can still be challenging for firms to adopt this new technology. As with other emerging technologies, adoption barriers include the need for new technical skills, organisational resistance to change, cultural inertia, risk appetite, and vendor dependency (Kempegowda & Chaczko, 2016). Regardless, GenAI is increasingly being explored as a means of supporting EA practices. (Haki *et al.*, 2025; Jung & Wienke, 2024; Yablonsky, 2021)

Empirical studies indicate that while GenAI can support aspects of qualitative analysis, it may not yet achieve the reliability or depth of experienced human analysts (Jung and Wienke, 2024; Mehta et al., 2025). As such, GenAI is better understood as an augmentative rather than replacement technology. This creates a fundamental tension between analytical capability and epistemic reliability: while GenAI enables scalable analysis, it simultaneously introduces risks that challenge the validity and trustworthiness of outputs.

GenAI has also been investigated in the automated construction of Enterprise Capability Maps. Jung and Wienke (2024) demonstrate that pretrained language models can significantly reduce the effort required to generate capability maps from textual organisational artefacts. However, their study highlights persistent challenges related to hallucinations and inconsistent source material, concluding that expert review remains necessary.

Although direct research on GenAI-supported capability maturity assessment remains limited, GenAI appears well suited to tasks involving qualitative analysis, which constitute a substantial component of existing maturity assessment methodologies. Nguyen & Welch (2025) note that “qualitative data analysis is “seen as fertile ground for such models, due to their natural language processing capabilities”, enabling the optimisation of qualitative data analysis workflows. Modern LLM’s are positioned as autonomous digital assistants capable of high-order analytical tasks such as identifying patterns, reasoning and interpreting meaning. (Nguyen & Welch, 2025)

Beyond academic explorations, industry reports (e.g.: ISACA Now, 2024) indicate that early implementations of AI-augmented maturity frameworks in organisational settings. For instance, professional services firms are leveraging GenAI to extend traditional CMMI practices to improve delivery quality, suggesting that AI-infused maturity evaluation is an emerging trend in industry, even if systematic empirical research remains limited.

3.6 Research Gap and Positioning

While the existing literature has made significant contributions to understanding capability maturity assessment, several critical gaps remain insufficiently addressed. First, there is a lack of formalised approaches for operationalising maturity scoring. Although many frameworks define maturity levels conceptually, they provide limited guidance on how these levels should be translated into observable, measurable, and context-sensitive criteria. As a result, maturity assessments often rely heavily on interpretive judgement without a consistent methodological foundation for deriving scores, raising questions regarding comparability and practical applicability across organisational contexts.

Second, the literature largely fails to establish a clear linkage between capability maturity assessment and strategy execution. While maturity models are frequently positioned as tools for identifying improvement opportunities, there is limited research on how identified maturity gaps can be systematically translated into concrete implementation artefacts, such as requirements, projects, or operational tasks. This disconnect creates a gap between assessment outputs and actionable execution, limiting the practical value of maturity assessments as instruments for driving organisational change.

Third, there is an absence of explicit guidance regarding the temporal dimension of maturity assessment. Existing frameworks either prescribe fixed assessment intervals or do not address cadence at all, despite the dynamic nature of organisational capabilities. Empirical findings from this study indicate that assessment practices in organisations are highly inconsistent, ranging from ad hoc evaluations to infrequent periodic reviews, often covering only a subset of capabilities. This suggests a misalignment between theoretical models and real-world organisational needs, particularly in environments characterised by continuous change.

Finally, the literature does not adequately address the relationship between capability maturity assessment and benefits realisation. While maturity models are intended to support strategic improvement, there is limited research examining whether and how changes in maturity can be used to evaluate the effectiveness of transformation initiatives. In particular, the potential for capability maturity progression to serve as an intermediate or leading indicator of strategic outcomes remains underexplored.

These gaps collectively highlight that current maturity assessment approaches are underdeveloped in terms of methodological rigour, integration with execution processes, and temporal responsiveness. This provides a foundation for the present study, which investigates the role of Generative Artificial Intelligence in addressing aspects of scalability and interpretive effort, while also motivating the development of more integrated and context-aware approaches to capability maturity management.

3.7 Research Methods

Research into Enterprise Architecture and capability maturity assessment has employed a range of methodological approaches, including qualitative case studies, survey-based research, and conceptual modelling. Survey methods are commonly used to capture practitioner perspectives across organisational contexts, particularly where constructs such as perceived value, adoption, and organisational practices are not directly observable. However, the literature highlights several limitations of survey-based approaches, including reliance on self-reported data, potential response bias, and challenges in capturing nuanced interpretations of complex organisational phenomena. These limitations suggest that survey findings should be interpreted as indicative of practitioner perceptions rather than as objective representations of organisational reality.

More broadly, the literature emphasises the importance of aligning research design with the nature of the research question, particularly when investigating emerging technologies such as GenAI, where empirical evidence remains limited and practitioner perception plays a significant role in shaping adoption trajectories. These methodological considerations inform the research design adopted in this study, which is discussed in detail in Chapter 4.

4. Research design and methodology

4.1 Research Paradigm

This study adopts a naturalistic research paradigm, as it seeks to understand practitioner perceptions and interpretations of capability maturity assessment and the potential role of GenAI within organisational contexts. While naturalistic research is typically associated with qualitative methods, this study employs a structured survey instrument to capture subjective perceptions in a standardised form. The use of quantitative measures is therefore methodological rather than epistemological, enabling systematic analysis of subjective judgements across a distributed practitioner population.

The research questions focus on constructs such as objectivity, acceptance, and perceived value, which are not directly observable but are shaped by practitioner experience and organisational context. A naturalistic paradigm is therefore appropriate, as it enables interpretation of these constructs as socially situated rather than objectively measurable phenomena.

In this study, objectivity is understood not as absolute correctness, but as procedural consistency, reproducibility, and transparency of assessment outcomes. While GenAI may improve consistency by applying uniform analytical processes, outputs remain constrained by the quality and completeness of underlying data. Capability maturity assessment is therefore conceptualised as a hybrid evaluative system, in which structured criteria (potentially supported by GenAI) provide decision support, while human judgement remains necessary for contextual interpretation and accountability.

4.2 Research Approach

The research adopts a predominantly quantitative analytical approach, using structured survey data to identify patterns in practitioner perceptions. However, this does not reflect a purely positivist stance; rather, quantitative analysis is used as a tool to interpret socially constructed viewpoints. Limited qualitative inputs, including open-text responses and optional follow-up interviews, are used to support and contextualise interpretation of the primary dataset.

The study follows a deductive logic of enquiry, with research questions and survey design informed by themes and gaps identified in the literature. It is both exploratory, in examining an emerging and under-researched application of GenAI, and descriptive, in analysing current organisational practices relating to capability maturity assessment.

Alternative approaches were considered but not adopted. A purely qualitative design would have limited the breadth of perspectives, while a formal mixed-methods approach was unnecessary given the study's exploratory objectives. The chosen approach therefore represents a pragmatic balance between interpretive depth and comparative insight within the constraints of a time-limited research project.

4.3 Research Strategy

The research strategy is based on a self-report survey of practitioners, supplemented by the option for follow-up interviews. The primary data collection instrument is an online questionnaire designed to capture both structured responses and contextual insights.

The choice of a survey-based strategy is justified by several considerations. First, it enables access to a geographically distributed population of EA practitioners. Second, it allows for the collection of comparable data across multiple organisations. Third, it provides a practical means of gathering sufficient data within the time constraints.

The survey instrument was designed to take approximately ten minutes to complete and consists of approximately twenty questions, covering both current practices and perceptions of GenAI-enabled approaches . The inclusion of an option for follow-up interviews provides an opportunity for deeper qualitative exploration, although the primary analytical focus remains on survey data.

Overall, the research design can be characterised as a naturalistic, deductive study employing structured survey methods to capture and analyse practitioner perceptions. This combination reflects a pragmatic approach, balancing the need for interpretive insight with the requirement for comparable and scalable data collection. This combination of a naturalistic paradigm with structured survey methods reflects an intentional methodological compromise, recognising both the interpretive nature of the research problem and the practical need for scalable data collection.

4.4 Data Collection Methods

4.4.1 Survey Instrument Design

The primary data collection method is a self-report questionnaire administered through Microsoft Forms. The questionnaire is structured into several thematic sections (See Appendix 5 & 6 for the full inventory of questions and the coding guide):

- Participant Context (Control Variables)
  Including role, sector, organisation size, and geographic region, to enable contextual interpretation of responses.
- Enterprise Architecture Context
  Assessing the maturity and positioning of EA within the organisation, including whether capability maps exist and whether maturity is measured.
- Capability Assessment Practices
  Including frequency of assessments, scope of assessments, and reasons for limited frequency. For example, respondents are asked to indicate how often assessments are performed (e.g. annually, quarterly, monthly) and to identify key constraints such as time and effort .
- Perceived Value of Continuous Assessment
  Evaluating whether more frequent or automated assessments are perceived as beneficial for strategy execution.
- GenAI Capabilities and Limitations
  Assessing perceptions of GenAI's ability to enable more frequent assessments, as well as concerns relating to objectivity, data availability, data quality, and management acceptance.

The questionnaire primarily uses structured response formats, such as Likert scales and predefined categorical options, to facilitate comparability and efficient analysis. However, two open-ended questions are included to capture qualitative insights and contextual explanations.

This design reflects a deliberate trade-off: while qualitative depth is desirable, the structured format ensures that data can be systematically analysed within the available timeframe.

4.4.2 Sampling Strategy

The study employs a non-random, purposive sampling strategy, targeting practitioners with relevant expertise in EA and capability management. Participants were recruited through professional networks, including LinkedIn and specialist EA communities.

This approach is appropriate for this research, where the objective is not to achieve statistical generalisability but to obtain informed perspectives from knowledgeable participants. Enterprise Architects are considered particularly suitable respondents, as they are directly involved in capability modelling, maturity assessment, and strategy execution processes.

However, this sampling strategy introduces potential limitations, including the lack of representativeness of the broader population, potential bias towards organisations with more mature EA practices, and self-selection bias among respondents. These limitations are acknowledged and considered in the interpretation of findings.

4.4.3 Data Collection Process

The questionnaire was distributed electronically and completed asynchronously by participants. The survey included an explicit consent mechanism at the outset, ensuring that participation was voluntary and informed . Respondents were also given the option to provide contact details for follow-up communication or interviews, although this was not mandatory.

On completion of the Data Collection, Microsoft Forms then provides all collected survey responses in a single CSV (Comma-Separated Value) file for analysis.

During the data collection process, a limitation of the questionnaire-based approach became apparent. In follow-up interviews, several participants revised or reconsidered their responses after discussing their reasoning in more depth. This suggests that some questionnaire responses may reflect incomplete interpretation of questions rather than stable underlying perceptions. This indicates potential weaknesses in the structure and clarity of the survey instrument, particularly in ensuring consistent understanding across respondents. In retrospect, a more fully qualitative approach may have provided greater control over response validity, although this would have reduced the achievable sample size and comparability of responses. This highlights a key trade-off in the chosen approach: while structured questionnaires enable efficient data collection at scale, they limit the ability to probe and validate participant understanding in real time.

To strengthen analytical rigour, qualitative responses were examined using a light thematic analysis approach. Open-text responses and interview notes were iteratively reviewed to identify recurring patterns relating to key constructs (e.g. perceived value, organisational constraints, data readiness, and trust in GenAI). These themes were not treated as formal coded categories, but as interpretive lenses used to explain quantitative response distributions and identify underlying tensions.

4.4.4 Validity, Reliability, and Limitations

Credibility is supported through the selection of participants with relevant domain expertise and through the alignment of questions with the research objectives. Dependability is enhanced through the use of a consistent survey instrument applied uniformly across respondents.

However, several limitations remain. The reliance on self-reported data introduces the possibility of response bias, including socially desirable responses or incomplete understanding of questions. In addition, the structured nature of the questionnaire may limit the depth of qualitative insight.

Finally, the use of non-random sampling restricts the ability to generalise findings beyond the sample. The results should therefore be interpreted as indicative rather than representative, providing insight into practitioner perspectives rather than definitive conclusions about the broader population.

5. Analysis and findings

5.1 Overview of the Dataset

The study generated 55 valid responses from practitioners engaged primarily in EA and related domains. The respondent group was strongly dominated by Enterprise Architects, with additional representation from IT Architects, Solution Architects, and senior roles such as CIO and executives. This composition supports the relevance of the dataset, as respondents are directly involved in capability modelling, maturity assessment, and strategy execution.

A significant proportion of respondents are drawn from organisations with more than 250 employees, including many with more than 5,000 employees . This is important because capability-based management and maturity assessment practices are typically more formalised in larger organisations.

5.2 Analytical Approach

The analysis combines descriptive statistics with interpretive qualitative analysis. Structured survey responses were analysed through frequency distributions to identify dominant response patterns, while recognising that the ordinal nature of Likert-type scales limits the validity of interval-style inference. Open-text responses and interview notes were then used analytically rather than illustratively: first, to explain why particular quantitative patterns may have emerged; second, to identify tensions not visible in the closed-response data alone; and third, to assess whether the survey findings reflected stable practitioner views or more contingent organisational positions. The findings are therefore presented not simply as response distributions, but as patterns interpreted in relation to the research questions, the literature, and the organisational context in which maturity assessment takes place.

5.3 Findings by Research Question

RQ1: Frequency of Capability Maturity Assessments

The findings indicate that capability maturity assessments are performed infrequently. Most organisations either do not conduct assessments or do so on an ad hoc basis, with only a small minority reporting annual or quarterly evaluation. This is analytically important because it suggests that maturity assessment is not embedded as a continuous management discipline, but instead activated episodically, often in response to specific initiatives or local needs. The limited scope of assessment, frequently applied only to subsets of capabilities rather than the full capability landscape, further supports the interpretation that maturity assessment is being used tactically rather than systemically. This finding aligns with the literature's critique of periodic maturity models as poorly suited to dynamic organisational environments, but goes further by suggesting that in practice the problem is not merely cadence, but the weak institutionalisation of the activity itself.

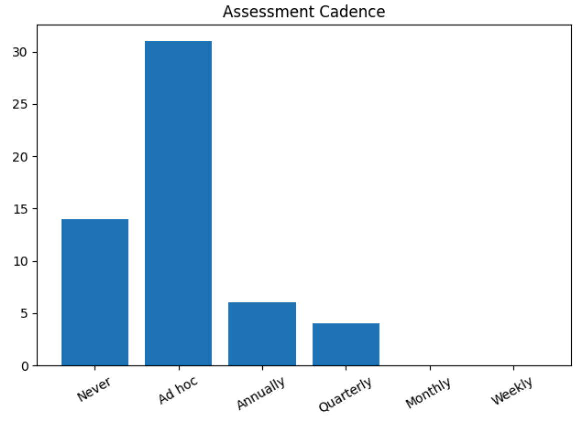


Figure 4(a): Distribution of capability maturity assessment cadence. The majority of organisations (45 out of 55) perform assessments either ad hoc or annually, with no respondents reporting continuous (monthly or weekly) assessment practices.

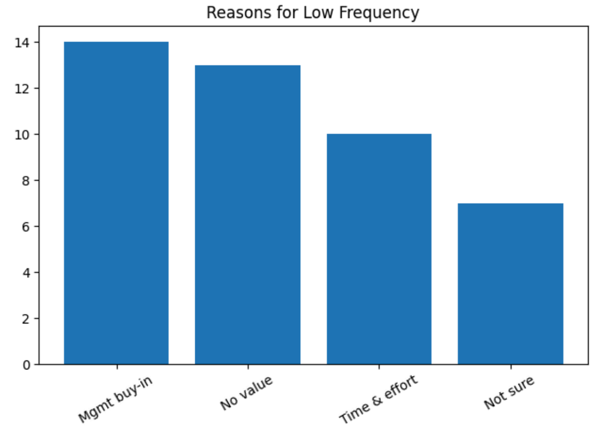


Figure 4(b): Reasons for infrequent capability maturity assessments. Contrary to expectations, organisational factors such as lack of management buy-in and perceived lack of value are more significant barriers than time and effort.

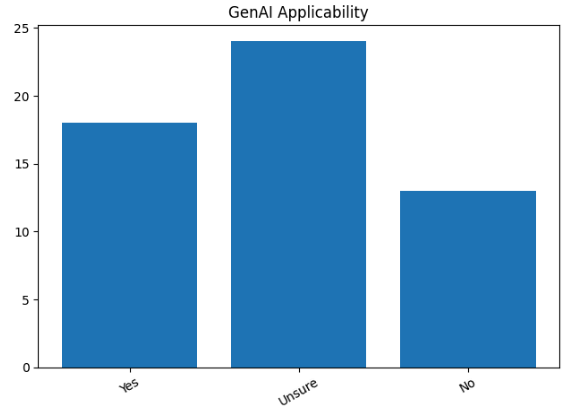


Figure 4(c): Perceived applicability of GenAI for increasing assessment frequency. While some respondents see potential, the largest group remains uncertain, indicating a lack of confidence or maturity in current GenAI capabilities.

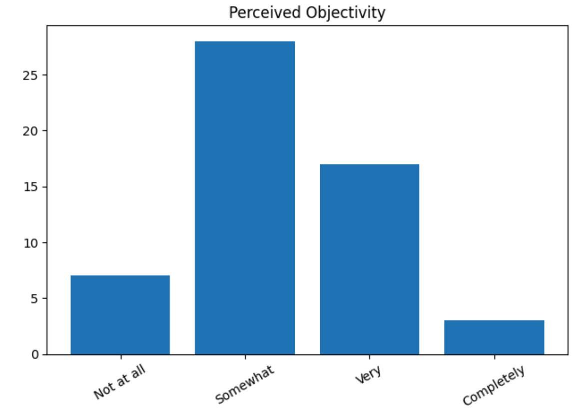


Figure 4(d): Respondents perceive GenAI assessments as at least somewhat objective. Suggests that, while GenAI is seen as capable of reducing subjectivity, confidence in fully unbiased assessment remains constrained.

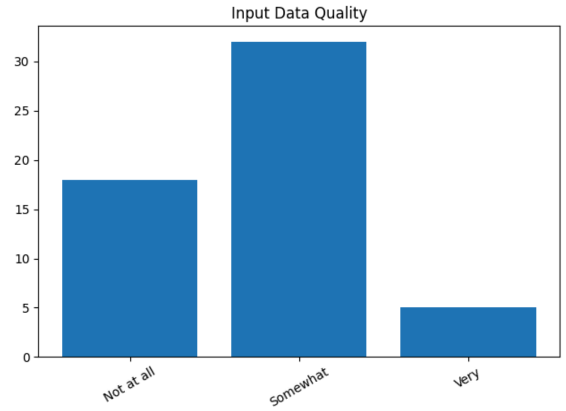


Figure 4(e): Respondents report that input data is only somewhat available and of limited quality, with a notable proportion indicating that data is not available at all. The scarcity of high-quality data represents a critical barrier to

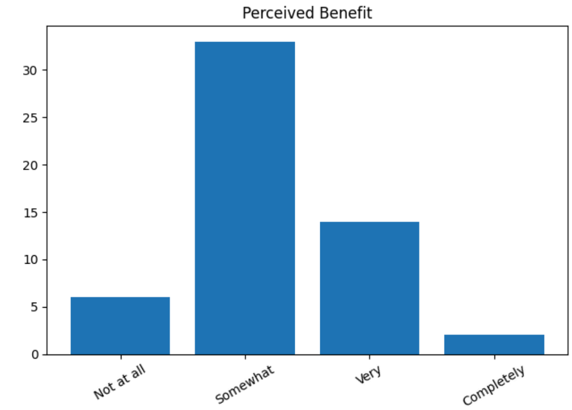


Figure 4(f): Perceived benefit of more frequent capability maturity assessments. While most respondents see some value, relatively few consider such assessments to be highly impactful, indicating moderate but not strong *the effective and trusted use of GenAI.*

support.

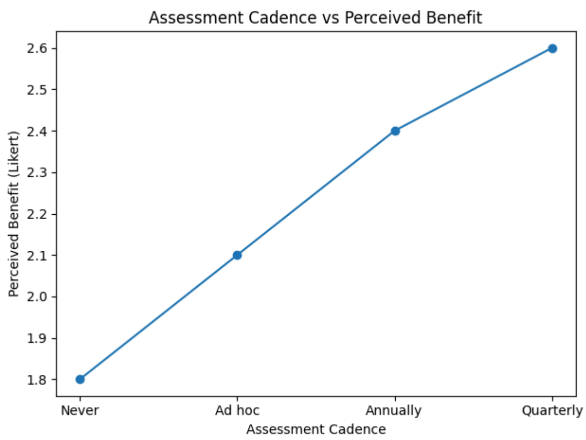


Figure 4(g): Increasing trend between more frequent capability maturity assessments and higher perceived benefit for strategy execution. Although most organisations currently assess infrequently, respondents generally recognise greater value in more regular assessments.

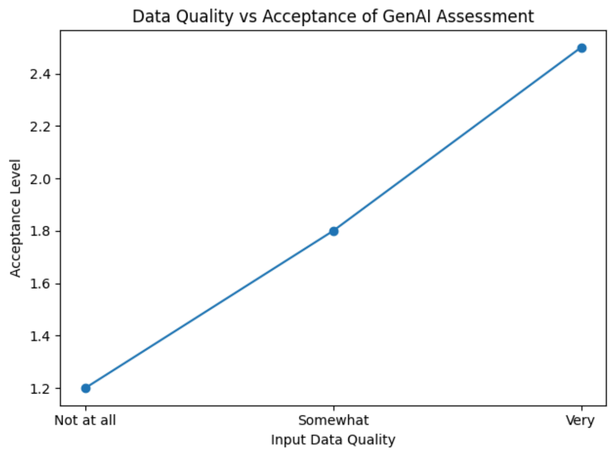


Figure 4(h): Positive relationship between perceived input data quality and stakeholder acceptance of GenAI-generated assessments. Low data quality appears to significantly reduce trust in automated assessment outcomes, despite perceptions of objectivity.

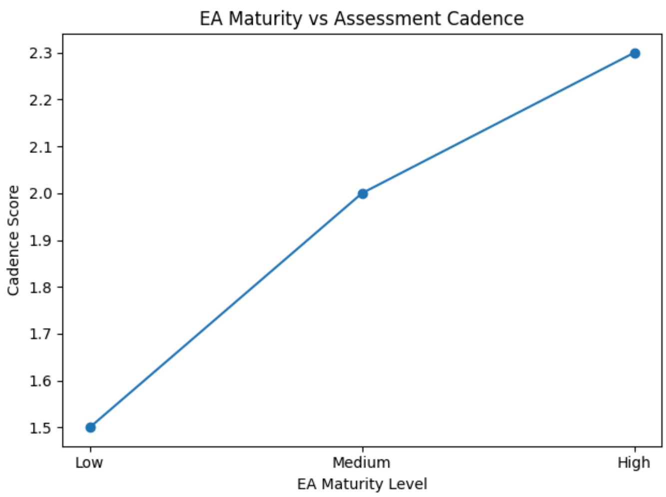


Figure 4(i): Higher levels of EA maturity may be associated with slightly more structured assessment practices. However, even among more mature organisations, assessment cadence remains relatively infrequent.

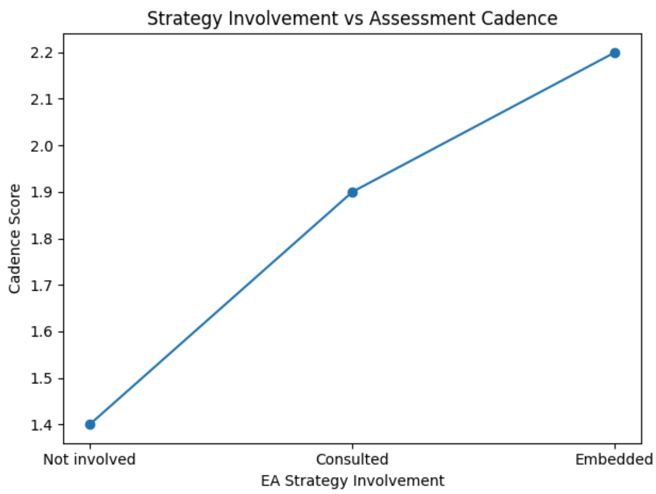


Figure 4(j): The chart indicates a modest positive association between EA involvement in strategy and assessment cadence. Note that this pattern is based on ordinal survey data and should be interpreted cautiously.

Note: The correlation visualisations are based on ordinal (Likert-scale) and categorical survey data, and do not constitute formal statistical regression analysis. Consequently, the relationships presented should be interpreted as indicative patterns that support qualitative interpretation, rather than statistically significant or causal findings.

RQ2: Constraints on Assessment Frequency

The findings indicate that constraints on assessment frequency are driven by both operational and organisational factors. The most frequently cited barriers were lack of management buy-in and limited perceived value, followed by time and effort required. Analytically, this matters because it shifts the explanation away from a simple resource-based account. If time and effort were the dominant barriers, automation alone would appear to solve the problem. Instead, the data suggests that infrequent assessment reflects a deeper issue: maturity assessment is often not sufficiently legitimised within organisational governance to justify more regular use.

Within the sample, organisations appear to evaluate assessment frequency through the lens of current manual practices, where the cost–benefit trade-off is perceived as weak. Qualitative responses reinforce this interpretation. One respondent noted that "if people find no value in a task, they won't use a tool to perform that task," indicating that low perceived value is not secondary to the problem, but constitutive of it. This challenges the initial assumption that low frequency is primarily caused by assessment effort, and suggests instead that the problem is partly conceptual: organisations may not yet regard maturity assessment as a sufficiently valuable management mechanism to merit continuous application.

RQ3: Feasibility of GenAI for Continuous Assessment

Perceptions of GenAI's ability to enable more frequent capability maturity assessments are mixed. While some respondents consider GenAI capable of reducing assessment effort through the analysis of unstructured data, a substantial proportion remain uncertain or sceptical. The key analytical point is that feasibility is not being judged solely in terms of model capability, but in terms of the organisational conditions required for meaningful use. In other words, respondents are not simply asking whether GenAI can analyse; they are asking whether their organisations possess assessable, trustworthy, and sufficiently integrated evidence for analysis to be credible.

A dominant theme across qualitative responses is the dependency of GenAI effectiveness on data quality and availability. Several participants emphasised that "very few companies have the data" required to support meaningful assessment, while others summarised the challenge more bluntly as "sht data in, sht data out." This suggests that the feasibility of GenAI-enabled assessment is constrained less by model sophistication than by organisational data maturity. The practical implication is significant: GenAI does not remove the need for capability discipline, codified knowledge, and traceable evidence; rather, it increases the importance of these conditions.

The qualitative responses also reveal that respondents are implicitly distinguishing between different types of feasibility. Technical feasibility refers to whether GenAI could process large volumes of unstructured information; epistemic feasibility refers to whether the resulting output would be sufficiently grounded, context-sensitive, and auditable to support decision-making. It is this second form of feasibility that appears to be the more serious limiting factor.

RQ4: Objectivity and Acceptance of GenAI-Based Assessments

The findings reveal a divergence between perceived objectivity and expected acceptance of GenAI-based assessments. Respondents generally view GenAI as capable of improving consistency and reducing subjectivity. However, this does not translate into strong expectations of stakeholder acceptance, with concerns centred on data quality, contextual understanding, and trust in AI-generated outputs. Analytically, this indicates that objectivity and acceptance are not parallel variables: improvements in procedural consistency do not automatically produce legitimacy in use.

Qualitative data makes this divergence particularly clear. One participant observed that “management will accept the results when they like them, and reject them when they don’t,” suggesting that acceptance is shaped not only by analytical quality, but by organisational politics and the perceived consequences of the result. This reframes the question of adoption. The challenge is not simply whether GenAI can produce more objective outputs, but whether those outputs are transparent, interpretable, and governable enough to be trusted by accountable stakeholders.

The implication is that the adoption barrier is partly epistemic and partly political. GenAI may reduce assessor inconsistency, yet still fail to gain acceptance if decision-makers cannot trace the assessment back to credible evidence or if the result disrupts existing power structures and expectations.

RQ5: Perceived Value of More Frequent Assessments

More frequent capability maturity assessments are generally perceived as beneficial for strategy execution, although this perception is moderate rather than strongly endorsed. This matters analytically because it suggests that respondents are not rejecting the underlying principle of frequent assessment; rather, they are expressing qualified support conditioned by context, data availability, and organisational use case. The contrast with current practice therefore does not simply indicate inconsistency. It indicates a gap between the limited value delivered by current manual approaches and the value respondents believe could exist under improved assessment conditions.

Qualitative responses suggest that this perceived benefit is context-dependent rather than universal. Participants from relatively stable organisational settings questioned whether monthly or continuous reassessment would create meaningful value where strategy and operating conditions change slowly, whereas others implied greater usefulness in transformation contexts, mergers, diversification, or fast-moving environments. This suggests that the benefit of increased assessment cadence is contingent on organisational dynamism. Continuous assessment may therefore be best understood not as a universal best practice, but as a strategically situational capability.

5.4 Cross-Cutting Insights

5.4.1 Key Insights

Three cross-cutting insights emerge when the quantitative and qualitative data are interpreted together. First, the findings reveal a structural gap between the theoretical role assigned to maturity assessment in the literature and its actual use in practice. In theory, maturity assessment is positioned as a mechanism for supporting strategy execution; in practice, it is often episodic, limited in scope, and weakly embedded in governance, with impact on strategy execution often limited to simple inputs for budget cycles.

Second, the data suggests that the core barriers to more frequent assessment are not primarily technical. While effort remains relevant, the stronger constraints are organisational: unclear ownership, weak managerial demand, low perceived value, and fragmented responsibility. This indicates that the maturity problem is not simply one of assessment design, but one of organisational positioning.

Third, the findings reveal a mismatch between what organisations currently experience and what they believe might be possible. Current manual approaches produce limited realised value, yet respondents still identify intrinsic value in more responsive, data-driven assessment under improved conditions. The significance of this distinction is that it explains why current low-frequency practice does not necessarily imply rejection of the broader concept of continuous assessment.

5.4.2 Association Analysis

The association analysis identifies several indicative patterns that support, but do not prove, the broader interpretation of the dataset. First, more frequent capability maturity assessment is associated with higher perceived benefit for strategy execution. This suggests that respondents who experience or imagine more regular assessment also tend to see greater strategic usefulness in it.

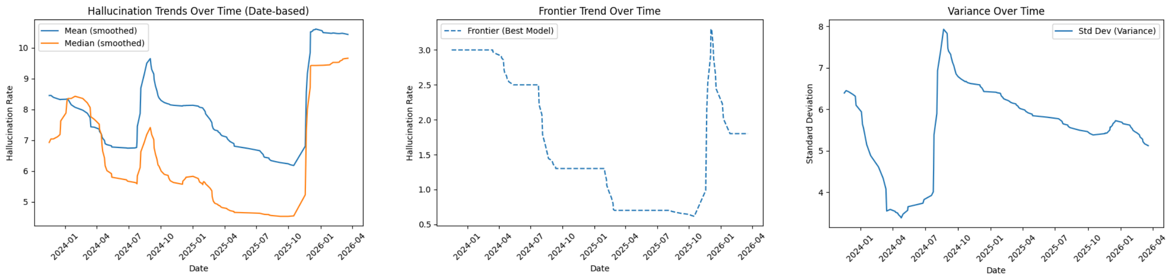
Second, higher perceived input data quality is associated with greater acceptance of GenAI-based assessments. This is analytically significant because it suggests that trust in automated assessment is mediated by the credibility of the evidence base rather than by model capability alone. In practical terms, even a more consistent GenAI process is unlikely to be accepted where the organisation does not believe the underlying data adequately represents operational reality.

Third, the weak association between greater EA maturity or stronger strategic involvement and more frequent assessment suggests that architectural maturity alone does not guarantee continuous evaluative practice. This indicates a disconnect between having formal architectural capability and using that capability to sustain ongoing measurement and review.

Taken together, these findings both support and complicate the literature. They support prior critiques of static, process-centric maturity models as poorly aligned with dynamic capability development, but they also suggest that the practical limits of assessment are not exhausted by model design alone. Organisational legitimacy, governance, data maturity, and perceived value emerge as equally important explanatory factors. In this respect, the study extends the literature by showing that the barriers to continuous assessment are not only methodological, but socio-organisational.

5.4.3 Longitudinal Hallucination Trends

As the research consistently showed that hallucinations are currently a significant factor in recommending against fully automated assessment, it could be asked if this situation is likely to change in the short to near-term future.



(a) *(b)* *(c)*

<u>Figure 5</u>: (a) Temporal evolution of mean and median hallucination rates across LLMs (smoothed), (b) Frontier hallucination rate (best-performing model) over time, (c) Temporal evolution of variance (standard deviation) in hallucination rates

To provide indicative insight into whether GenAI hallucination behaviour is improving or deteriorating over time, the Vectara dataset (See Appendix 3) was analysed. (Vectara, 2026) Temporal analysis of hallucination rates reveals a non-linear trajectory. (See Figures 5a-c) Initial improvements are followed by a stabilisation phase and subsequently by a divergence phase characterised by increased variance and a deterioration in average performance. This suggests that while model capability improves at the frontier, ecosystem-wide reliability is influenced by model proliferation and heterogeneity. Note that this analysis is exploratory and based on secondary leaderboard data, and should therefore be interpreted as indicative rather than conclusive.

6. Conclusions

6.1 Answer to Research Questions

The research questions can now be addressed directly as follows:

RQ1: Capability maturity assessments are performed infrequently, typically on an ad hoc or annual basis, and are rarely conducted across the full capability landscape.

RQ2: The frequency of assessments is constrained by a combination of operational effort, lack of management buy-in, and limited perceived value of current assessment practices.

RQ3: GenAI is perceived as potentially capable of enabling more frequent assessments, but this capability is conditional on data availability, data quality, and organisational readiness.
RQ4: GenAI-based assessments are perceived as moderately objective, but are unlikely to be widely accepted by management without improvements in data quality, transparency, and contextual interpretability.
RQ5: More frequent assessments are generally perceived as beneficial for strategy execution, particularly when enabled through automation or improved analytical approaches.

6.2 Theoretical Contributions

This study contributes to the literature by integrating three previously separate domains of EA, Capability Assessment, and GenAI.

First, it reinforces the argument that capability maturity assessment is inherently dynamic and that periodic assessment approaches are misaligned with the evolving nature of organisational capabilities. This supports and extends existing work on continuous assessment models by providing empirical evidence of the limitations of current practice.

Second, the study introduces a conceptual distinction between realised value and intrinsic value in capability maturity assessment. While existing approaches deliver limited value in practice, the concept of frequent and continuous assessment is perceived as valuable under improved conditions. This distinction helps explain the gap between current organisational behaviour and perceived future potential.

Third, the study highlights a critical tension between analytical capability and organisational trust in the context of GenAI. While GenAI may enhance analytical consistency and scalability, its effectiveness is constrained by epistemic factors such as data quality, interpretability, and stakeholder acceptance. This contributes to broader discussions in the literature regarding the role of AI in qualitative analysis and decision support.

A distinction emerging from the interviews is between benchmarking-driven and strategy-driven approaches to capability maturity assessment. In the first, maturity assessment is used primarily for external comparison, standardisation, or consultant-led review. In the second, it is used as an internally anchored mechanism for identifying capability gaps relative to strategic intent. This distinction is analytically important because it suggests that maturity assessment is not one coherent practice but at least two overlapping ones, each with different assumptions about value, cadence, and actionability. Where benchmarking logic dominates, periodic assessment may be sufficient. Where strategy-execution logic dominates, more continuous and context-sensitive assessment becomes more compelling. This distinction also helps explain why respondents differ in their views of the value of higher-frequency assessment.

6.2.1 Proto-Framework for Continuous Capability Maturity Management

The findings of this study, together with the limitations identified in the literature, indicate the need to reconceptualise capability maturity assessment. Existing approaches treat assessment as a periodic diagnostic activity, producing static snapshots of organisational capability. However, both the empirical results and the dynamic nature of organisational environments suggest that such approaches are insufficient for supporting effective strategy execution. This research therefore proposes a proto-framework for continuous capability maturity management, repositioning assessment as an integrated, iterative, and governance-driven process. As a proto-framework, it represents a structured conceptualisation grounded in empirical findings, but not yet fully validated or generalised across organisational contexts.

At its core, the proto-framework emphasises vertical alignment between strategic objectives and operational execution. Strategy is translated into required capabilities and target maturity states, defined through observable, context-specific criteria across dimensions such as people, process, technology, and data. This shifts maturity from an abstract ordinal construct to a concrete and actionable representation of capability performance.

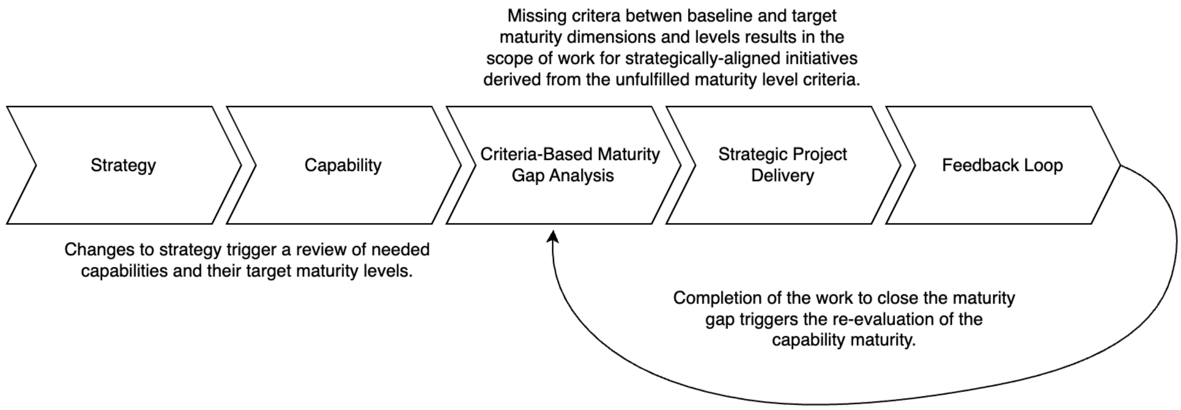


Figure 6: Proposed Proto-Framework for Continuous Capability Maturity Management

The gap between baseline and target maturity defines a set of capability gaps, which can be decomposed into missing or partially satisfied criteria. These can be translated into requirements and improvement initiatives, linking maturity assessment directly to delivery artefacts such as projects, epics, and tasks. In this way, assessment evolves from a descriptive tool into an operational mechanism for coordinating transformation and monitoring strategic capability development.

A further key element of the proto-framework is the introduction of event-driven assessment cadence. Traditional models rely on fixed intervals, which are poorly aligned with continuous organisational change. The proto-framework proposes reassessment triggered by state-based events, such as the completion of capability-improving project work, enabling evaluation to reflect actual change rather than arbitrary timelines.

The proto-framework incorporates a hybrid evaluation model combining structured objectivity with accountable subjectivity. Predefined criteria and, potentially, GenAI support improve consistency and scalability, while human judgement remains essential for contextual interpretation and accountability. This reflects the divergence observed between perceived objectivity and stakeholder acceptance, positioning GenAI as an augmentative rather than authoritative component of the assessment process.

From a governance perspective, the proto-framework recognises that capabilities are realised across multiple stakeholders rather than single organisational units. This implies the need for clearly defined ownership structures and alignment on capability definitions, maturity criteria, and target states. It also enables maturity assessment at both local and enterprise levels, supporting granular insight alongside strategic oversight.

Finally, the proto-framework introduces a conceptual linkage between capability maturity progression and benefits realisation. While maturity improvements do not directly equate to business outcomes, they may serve as intermediate indicators of capability enablement, providing a structured basis for evaluating the effectiveness of strategic initiatives.

Taken together, the proto-framework conceptualises capability maturity management as a continuous, closed-loop system integrating strategy, measurement, execution, and governance. Rather than an isolated activity, maturity assessment becomes embedded within organisational processes, supported by both human judgement and AI-enabled analysis. As a conceptual model, it provides a foundation for future research to evaluate its feasibility, effectiveness, and practical implications.

6.3 Practical Implications

The findings suggest that improving strategy execution requires embedding capability maturity assessment within governance structures, ensuring clear ownership and alignment with decision-making processes. GenAI should be introduced incrementally as an augmentative tool, supporting analysis while maintaining human oversight.

Organisations should prioritise data quality and transparency to enable trust in assessment outputs. Capability maturity progression may serve as an intermediate indicator of capability development, although it should be interpreted as a leading indicator rather than a direct measure of business value.

The qualitative findings further support the proposed hybrid model of assessment. Participants consistently emphasised that while GenAI may enhance analytical scalability, human judgement remains essential for contextual interpretation. As one respondent noted, “GenAI results would still need a human to review,” reinforcing the need for human-in-the-loop governance.

6.4 Limitations of the Study

This study has several limitations. First, the use of a non-random, purposive sample limits generalisability. The dataset is skewed towards EA practitioners and organisations with some degree of EA maturity, and may therefore not reflect the broader population. Second, the reliance on self-reported data introduces potential bias, including subjective interpretation of questions and socially desirable responses. The structured survey design also limits the depth of insight that might have been obtained through more extensive interviews. Third, the study examines perceptions of GenAI rather than implemented systems, meaning that the findings reflect expectations and concerns rather than observed organisational outcomes.

Several conceptual limitations should also be acknowledged in relation to capability maturity assessment and its proposed evolution towards more continuous and integrated approaches. First, attributing changes in capability maturity to specific initiatives remains difficult, since organisational change is typically non-linear and shaped by multiple concurrent factors. Second, event-driven reassessment depends on the definition of meaningful triggers and thresholds; without these, reassessment may become either excessive and inefficient or too infrequent to remain useful. Third, maturity levels are typically represented as ordinal scales, so the distance between levels is not necessarily uniform, which limits the validity of aggregation and comparison over time or across capabilities. Finally, the proposed link between maturity progression and benefits realisation should be treated cautiously. While maturity improvement may indicate enhanced organisational capability, it does not directly demonstrate business outcomes, and is therefore better understood as a leading or intermediate indicator rather than a definitive measure of realised strategic value.

In addition, the qualitative responses revealed potential limitations in questionnaire-based data collection, as some participants revised their views during follow-up discussion. This suggests that structured survey responses may not fully capture nuanced perspectives.

6.5 Recommendations

Based on the findings, several recommendations can be made.

1. Organisations should establish stronger governance around capability management, including clear ownership, consistent definitions, and integration with strategic & delivery processes.
2. Capability maturity assessment should be progressively embedded into regular management practices, rather than treated as a one-off or project-specific activity.
3. GenAI should be introduced incrementally, starting with support for data aggregation and preliminary analysis, while maintaining human oversight.
4. Investment in data quality and integration should be prioritised to enable more effective use of AI-driven approaches.
5. Organisations should develop frameworks for evaluating and validating AI-generated insights to support trust and adoption.

6.6 Future Research

Future research could extend this study in several directions. First, empirical case studies of organisations implementing GenAI-supported maturity assessment would provide valuable insight into real-world effectiveness and adoption challenges.

Second, longitudinal studies could examine how perceptions and practices evolve over time as GenAI technologies mature and become more widely adopted.

Third, further research could explore the design of hybrid assessment models that combine AI-driven analysis with structured human evaluation, potentially addressing the tension between scalability and contextual understanding.

The relationship between capability maturity assessment and measurable strategy execution outcomes could be investigated in greater depth, providing stronger evidence of the practical value of continuous assessment approaches.

A further important direction for future research is the empirical validation and refinement of the proposed capability maturity assessment proto-framework. This could involve applying the proto-framework in real organisational settings to assess its practical usability, consistency of scoring, and impact on decision-making and strategy execution. Comparative evaluation against established maturity models would help determine whether the proto-framework addresses identified limitations relating to scoring transparency, contextual sensitivity, and actionability.

6.7 Concluding Remarks

This study demonstrates that while the concept of continuous capability maturity assessment is both relevant and valued, its realisation is constrained by organisational, methodological, and epistemic factors. GenAI offers promising capabilities to address some of these challenges, particularly in relation to scalability and data analysis. However, its effective application depends not only on technological capability, but also on data maturity, governance structures, and stakeholder trust.

Rather than representing a complete solution, GenAI should be understood as an enabling technology within a broader transformation of how organisations assess, manage, and evolve their capabilities in support of strategy execution.

References

Acar, A. Z., Zehir, C. (2010) “The Harmonized Effects of Generic Strategies and Business Capabilities on Business Performance”. *Journal of Business Economics and Management* 2010, 11(4): 689-711. doi:10.3846/jbem.2010.34

Bernard, S. (2005) “An introduction to enterprise architecture”. *AuthorHouse*: Bloomington

Brits, J., Botha, G. H. K., Herselman, M. E. (2007) “Conceptual Framework for modeling business capabilities”. *Proceedings of the 2007 Information Science and IT Education Joint Conference*.

Cater, T., Pucko, D. (2010) “Factors of effective strategy implementation: empirical evidence from Slovenian business practice”. *J East Eur Manag Stud* 15(3):207-236.

Dillman, D. A., Christian, L. M. (2014) “Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method: 4th ed.” Wiley & Sons Ltd.

Fowler, F. J, (2012) “Survey Research Methods: 4th Edition”. Sage Publications, Inc. https://doi.org/10.4135/9781452230184

Haki, K., Safaei, D., Magan, A., Griffiths, M. (2025) “Integrating Generative AI Into Enterprise Platforms: Insights from Salesforce”. *Information Systems Journal*. Wiley & Sons Ltd. doi:10.1111/isj.12593

ISACA Now. (2024) “Integrating AI with CMMI: Cognizant’s Path to Elevating Excellence”. Available at https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2024/integrating-ai-with-cmmi-cognizants-path-to-elevating-excellence (Accessed on 8 February 2026)

Jung, J., Wienke, P. (2024) “Generating Business Capability Maps using GenAI: A Case Study”. *Companion Proceedings of the 17th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling Forum*: Stockholm.

Kappelman, L., McGinnis, T., Petite, A. (2009) “Enterprise Architecture: Charting the territory for academic research”. *Proceedings of 14th Americas conference on Information Systems (AMCIS)*

Keller, W. (2009) “Using capabilities in enterprise architecture management”. *Object architects*

Kempegowda, S. M., Chaczko, Z. (2016) “Adoption of Emerging Technologies established on Comprehensive Capability Maturity Model Framework: A new practical model”.

Korsten, G., Ozkan, B., Aysolmaz, B., Mul, D., Turetken, O. (2024) “Understanding Capability Progression: A Model for Defining Maturity Levels for Organizational Capabilities”, in H. van der Aa *et al* (Edds.), *Enterprise, Business-Process and Information Systems Modelling - 25th Internal Conference BPMDS 2024.* Cyprus: Springer. doi:10.1007/978-3-031-61007-3_26a

Mahadevkar, S., Patil, S., Kotecha, K., Soong, L. W., Choudhury, T. (2024) “Exploring AI-driven approaches for unstructured document analysis and future horizons”. *Springer Open*. doi:10.1186/s40537-024-00948-z

Mehta, S. D., Paul, S., Awiti, E., Young, S., Zulaika, G., Otieno, F. O., Phillips-Howard, P. A., Mason, L., Bhaumik, R. (2025) “Evaluation of large language models within GenAI in qualitative research”. *Nature*. doi:10.1038/s41598-025-18969-w

Nguyen, D. C., Welch, C. (2025) “Generative Artificial Intelligence in Qualitative Data Analysis: Analyzing - Or just chatting?”*Organizational Research Methods* 1-37. Sage doi:10.1177/10944281251377154

Offerman, T., Stettina, C. J., Plaat, A. (2017) “Business Capabilities: A Systemic Literature Review and a Research Agenda”. IEEE Xplore.

Pereira, C. M., Sousa, P. (2005) “Enterprise architecture: business and IT alignment”. *Proceedings of SAC* 2005.

Proenca, D., Borbinha, J. (2018) “Maturity Model Architect: A Tool for Maturity Assessment Support”. *IEEE Xplore*. doi:10.1109/CBI.2018.10045

Proenca, D., Borbinha, J. (2019) “Information Governance Maturity Assessment using Enterprise Architecture Model Analysis and Description Logic”, in T. Aalberg *et al*. (eds.) *Digital Libraries for Open Knowledge*. Switzerland: Springer International Publishing AG, pp. 265–279. Available at: https://doi.org/10.1007/978-3-030-30760-8_23.

Radeke, F. (2011) “Toward Understanding Enterprise Architecture Management’s Role in Strategic Change: Antecedents, Processes, Outcomes”. *Wirtschaftsinformatik Proceedings 2011(62)*. Available from http://aisel.aisnet.org/wi2011/62

Riege, C., Aier, S. (2009) “A contingency approach to enterprise architecture method engineering”. *J Enterp Architect 5(1):36-48*

Ross, J. W., Weill, P., Robertson, D. C. (2006) “Enterprise architecture as strategy” *Harvard Business School Press*: Boston

Simon, D., Fischbach, K., Schoder, D. (2013) “Enterprise architecture management and its role in corporate strategic management”. *Inf Syst E-Bus Manage* (2014) 12:5-42. doi:10.1007/s10257-013-0213-4

Stoiber, C., Stoter, M., Englbrecht, L., Schonig, S. (2023) “Keeping Your Maturity Assessment Alive: A method for the continuous tracking and assessment of organizational capabilities and maturity”. *Bus Inf Syst Eng* 65(6):703-721. doi:10.1007/s12599-023-00805-y

Vectara, Inc. (2026) “Hallucination Leaderboard”. Available at: https://github.com/vectara/hallucination-leaderboard (Accessed on 29 March 2026)

Vieira, R., Cardoso, E., Becker, C. (2014) “A traceable maturity assessment method based on Enterprise Architecture modelling”. *IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations.* doi:10.1109/EDOCW.2014.44

Yablonsky, S. (2021) "AI-driven platform enterprise maturity: from human led to machine governed" Kybernetes 50(10) pp. 2753-2789 Emerald Publishing, Ltd.

APPENDICES

APPENDIX 1: Acronyms

Acronym	Definition
AI	Artificial Intelligence
CMAM	Continuous Maturity Assessment Model
CMMI	Capability Maturity Model Integration
EA	Enterprise Architecture
GenAI	Generative Artificial Intelligence
IT	Information Technology
ISO	International Organization for Standardization
LLM	Large Language Model
MMArch	Maturity Model Architect
RAG	Retrieval-Augmented Generation
RQ	Research Question
SCAMPI	Standard CMMI Appraisal Method for Process Improvement

APPENDIX 2: Glossary of Terms

Business Capability

A business capability is an organisation’s ability to achieve a specific business outcome through the coordinated deployment of people, processes, technology, and governance structures. Capabilities are outcome-oriented and relatively stable abstractions that enable strategic reasoning across organisational boundaries.

Capability Maturity

Capability maturity refers to the degree to which a business capability is developed, stable, and effective in achieving its intended outcomes. It is typically represented through discrete maturity levels, although the definition and interpretation of these levels vary across frameworks.

Capability Maturity Assessment

A structured process used to evaluate the maturity level of business capabilities, typically involving qualitative judgement and/or quantitative indicators to identify strengths, weaknesses, and improvement opportunities.

Continuous Assessment

An approach to maturity evaluation in which capabilities are assessed iteratively or continuously over time, rather than through periodic, one-off assessments. Continuous assessment aims to provide timely and evolving insights aligned with organisational change.

Enterprise Architecture (EA)

A management discipline that provides a structured representation of an organisation and the relationships between its components, enabling alignment between business strategy, processes, information, and technology.

Generative Artificial Intelligence (GenAI)

A class of artificial intelligence systems, typically based on large language models, capable of generating, interpreting, and synthesising content from structured and unstructured data. In this study, GenAI is considered as a tool for augmenting qualitative analysis in capability maturity assessment.

Large Language Model (LLM)

A type of machine learning model trained on large datasets of text to perform tasks such as natural language understanding, generation, and reasoning. LLMs underpin many GenAI applications.

Objectivity (in this study)

Objectivity is defined as the degree of consistency, reproducibility, and transparency in assessment outcomes, rather than absolute correctness. It reflects the ability to produce stable results across comparable inputs, while acknowledging dependence on underlying data quality.

Perceived Value

The extent to which stakeholders believe that capability maturity assessment contributes to improved decision-making and strategy execution. This may differ between current practices (realised value) and potential future approaches (intrinsic value).

Realised Value

The practical value currently obtained from existing capability maturity assessment practices, often constrained by effort, frequency, and organisational adoption.

Intrinsic Value

The theoretical or potential value of capability maturity assessment when performed more frequently, systematically, or with enhanced analytical support (e.g. GenAI).

Stakeholder Acceptance

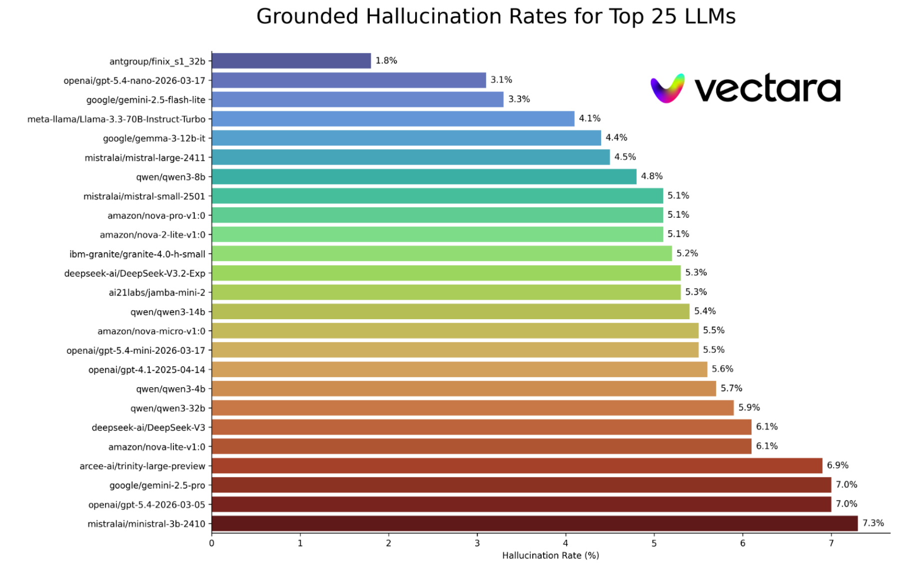
The degree to which organisational decision-makers trust and are willing to use assessment outputs, particularly those generated or supported by AI-based systems.

Data Quality

The degree to which input data used for assessment is accurate, complete, consistent, and representative of organisational reality. Data quality is a key determinant of trust in AI-generated outputs.

APPENDIX 3: Analysis of Longitudinal Hallucination Data

To the best of the author’s knowledge, no existing academic literature was identified that provides a longitudinal analysis of hallucination rates in large language models over time. While prior research has focused on point-in-time benchmarking and evaluation of model performance, there appears to be a gap in examining how hallucination behaviour evolves across successive model generations and ecosystem changes. To address this gap, this study leverages data from the Vectara Hallucination Leaderboard, an industry-maintained benchmark that evaluates large language models on factual consistency and hallucination-related metrics. Vectara is a company specialising in retrieval-augmented generation (RAG) and search technologies, and its leaderboard provides a continuously updated, standardised comparison of model performance across multiple dimensions. Although not peer-reviewed, this source constitutes a form of grey literature that is particularly valuable due to its recency, consistency of evaluation methodology, and coverage of a wide range of models over time. As such, it offers a pragmatic and uniquely suitable dataset for exploratory longitudinal analysis of hallucination trends, while requiring careful critical interpretation in line with established research guidance on non-academic sources.



Model	Hallucination Rate	Factual Consistency Rate	Answer Rate	Average Summary Length (Words)
antgroup/finix_s1_32b	1.8 %	98.2 %	99.5 %	172.4
openai/gpt-5.4-nano-2026-03-17	3.1 %	96.9 %	100.0 %	144.4
google/gemini-2.5-flash-lite	3.3 %	96.7 %	99.5 %	95.7
microsoft/Phi-4	3.7 %	96.3 %	80.7 %	120.9
meta-llama/Llama-3.3-70B-Instruct-Turbo	4.1 %	95.9 %	99.5 %	64.6

Figure 7: Screenshot of the Vectara Leaderboard (https://github.com/vectara/hallucination-leaderboard) accessed on 29.03.2026

The dataset[1] was reconstructed from the version history of the Vectara hallucination leaderboard README.md by extracting the leaderboard table from each committed README snapshot. Commit dates were obtained from the Git history of README.md and recorded at day level, meaning that multiple leaderboard updates on the same day share the same date value. The analysis therefore reflects documented leaderboard states over repository time rather than a controlled longitudinal series with uniform temporal intervals. In addition, the composition of models changes over time, and rows for removed models were generated during data extraction as missing-value markers rather than being directly observed in the original leaderboard tables. Consequently, the results should be interpreted as indicative ecosystem-level trends shaped by both temporal change and leaderboard composition, rather than as a fixed-cohort longitudinal benchmark.

Where multiple README commits occurred on the same day, results were aggregated to the date level, which may obscure intra-day leaderboard changes.

Central Tendency of Hallucination Rates Over Time

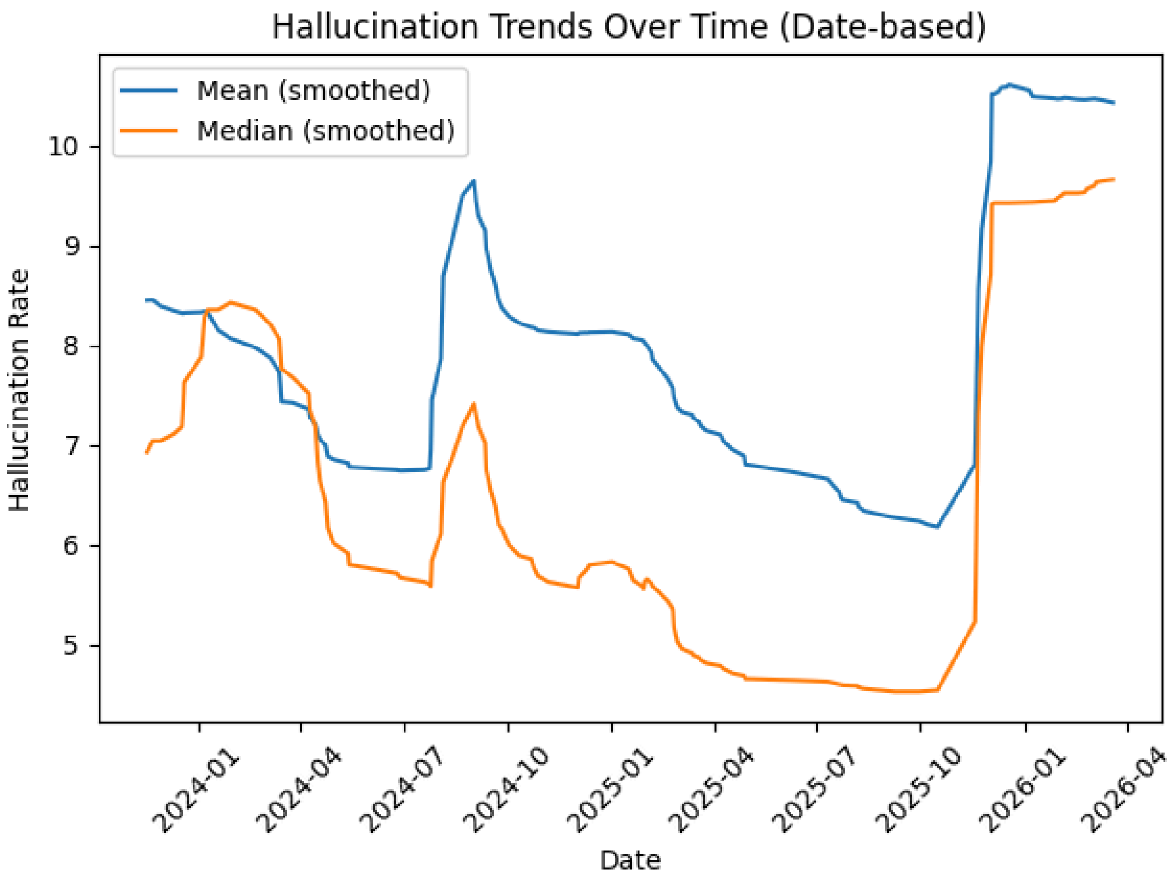


Figure 5a: Temporal evolution of mean and median hallucination rates across LLMs (smoothed)

This figure presents the smoothed mean and median hallucination rates across all models in the Vectara leaderboard over time. The median exhibits a gradual decline from late 2023

[1] The derived dataset has been made publicly available here: https://github.com/ralfepoisson/llm-hallucination-trends

through mid-2025, indicating modest improvements in the typical model's factual reliability. However, the mean remains consistently higher than the median, reflecting a right-skewed distribution driven by lower-performing models. Notably, a sharp increase in both metrics is observed in late 2025, suggesting a deterioration in aggregate performance coinciding with an influx of new or experimental models. This divergence between mean and median highlights the importance of distinguishing between typical and aggregate model behaviour when assessing hallucination trends.

Frontier (Best Model) Trend

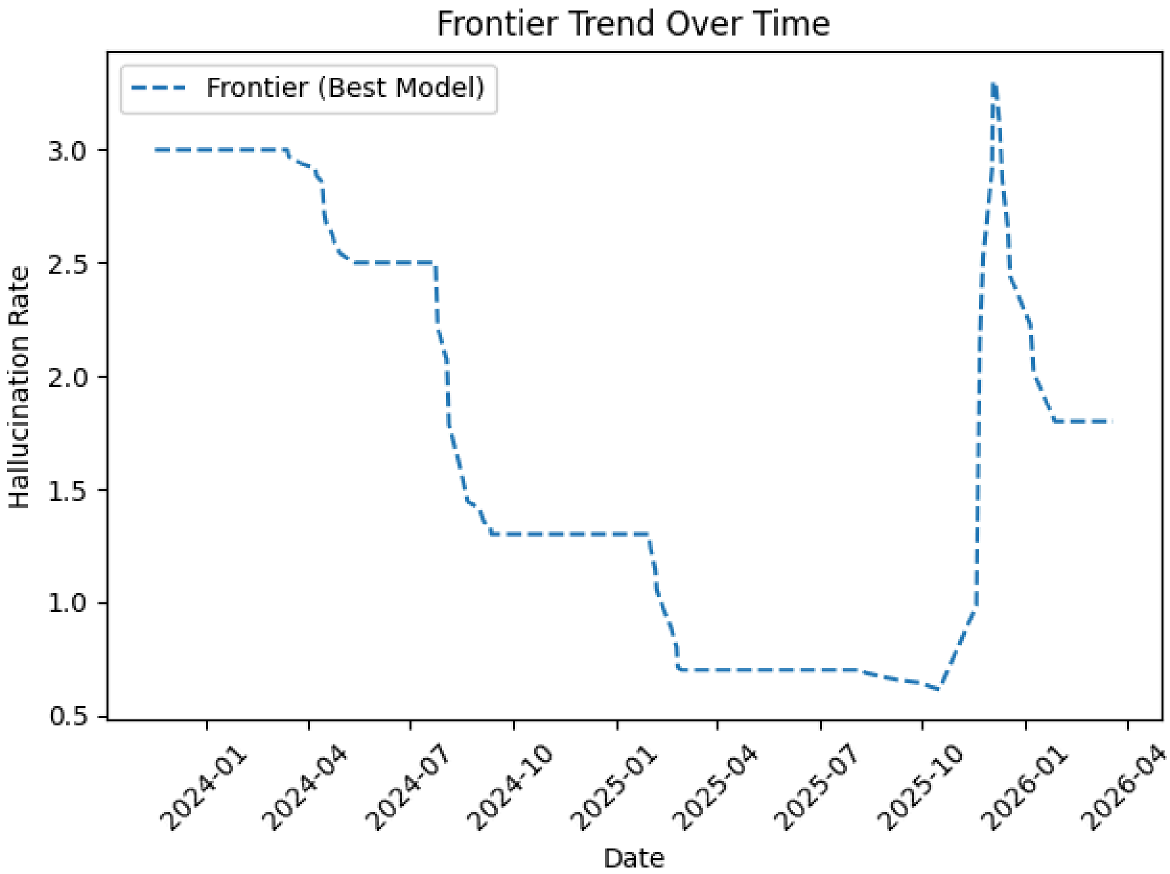


Figure 5b: Frontier hallucination rate (best-performing model) over time

This figure illustrates the evolution of the lowest hallucination rate observed among models at each point in time, representing the performance frontier. The data reveals a stepwise improvement in frontier performance throughout 2024 and early 2025, followed by a plateau at low hallucination levels. A temporary regression is observed in late 2025, after which performance partially recovers. These findings suggest that improvements in state-of-the-art models occur in discrete increments rather than as a continuous trend, and that frontier performance is not strictly monotonic. While leading models achieve increasingly low hallucination rates, occasional regressions indicate sensitivity to model turnover and evaluation conditions.

Variance of Hallucination Rates

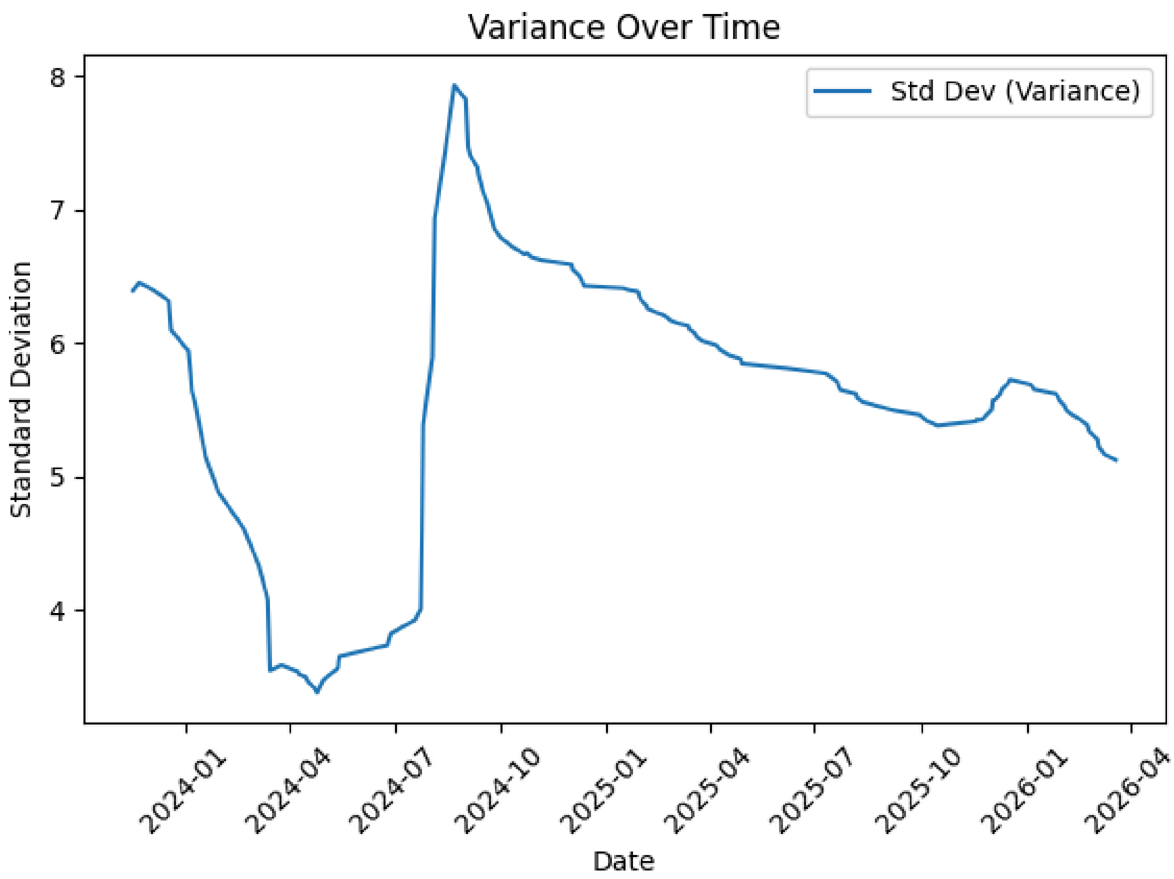


Figure 5c: Temporal evolution of variance (standard deviation) in hallucination rates

This figure shows the standard deviation of hallucination rates across models over time, serving as a proxy for ecosystem variability. Variance decreases during 2024 and early 2025, indicating convergence in model performance as hallucination rates become more consistent across models. However, a subsequent increase in variance is observed in late 2025, suggesting renewed divergence driven by the introduction of heterogeneous models with varying performance levels. This pattern indicates that while model quality initially converges during periods of technological maturity, it may diverge again as the ecosystem expands and diversifies.

Interpretation

Taken together, these analyses reveal a multi-layered evolution of hallucination performance. While frontier models demonstrate substantial improvements and eventual stabilisation at low hallucination rates, the broader ecosystem exhibits more complex dynamics. Median performance improves gradually, indicating steady gains in typical model quality, whereas increasing variance and rising mean values in later periods reflect growing heterogeneity. This suggests that improvements in state-of-the-art models do not necessarily translate into uniform gains across the model landscape, reinforcing the need for continuous and context-specific evaluation of model reliability.

APPENDIX 4: List of Figures

APPENDIX 5: Questionnaire

The following questions were included in the online self-reported questionnaire using Microsoft Forms:

#	Question	Type	Mandatory
1	What is your role?	drop-down	Yes
2	What sector do you work in?	drop-down	Yes
3	How many employees in your organization?	drop-down	Yes
4	In which region do you work?	drop-down	Yes
5	How established is Enterprise Architecture in your company?	drop-down	Yes
6	Is Enterprise Architecture involved in business strategy development and/or execution?	drop-down	Yes
7	Is a Capability Map developed for your company?	drop-down	Yes
8	Does each capability have a clear owner?	drop-down	Yes
9	Is the maturity level of each capability measured?	drop-down	Yes
10	How often are assessments of capability maturity performed?	drop-down	Yes
11	What is the scope of the assessments, if and when they are performed.	drop-down	Yes
12	Why are capability maturity assessments not performed more frequently? (Select the most applicable answer)	drop-down	Yes
13	In your estimation, would upper management find benefit in automated, continuous assessment of capability maturity to support strategy execution efforts?	drop-down	Yes
14	How established is Generative AI (GenAI) in your company?	drop-down	Yes
15	Do you believe that Generative AI (GenAI) could enable regular capability assessments at higher frequencies (monthly, for instance)?	drop-down	Yes
16	What is the reason for your previous choice (regarding GenAI enabling regular assessments)?	Free text	No

17	How objective do you think a GenAI-based business capability maturity assessment would be?	drop-down	Yes
18	How likely is it that required input data will be available for a GenAI model to use for an automated assessment?	drop-down	Yes
19	How likely is it that required input data will be of sufficient quality for the assessment to be well accepted by management?	drop-down	Yes
20	How likely is it that upper management would accept assessment results performed by GenAI?	drop-down	Yes
21	Do you have any thoughts on GenAI-enabled continuous assessment of capability maturity for supporting strategy execution?	Free text	No
22	If you would like to receive information about the findings of the research once concluded, please provide an email address to which the information will be shared.	Free text	No

APPENDIX 6: Coding Table

The following was used to code the closed-form responses selected from drop-downs into numerical values to support analysis:

Question	Value	Code
Q01	Unspecified	0
Q01	Enterprise Architect	1
Q01	CIO	2
Q01	CEO	3
Q01	IT Architect	4
Q01	Executive	5
Q01	Manager	6
Q01	IT Employee	7
Q01	Other	8
Q02	Unspecified	S00
Q02	Accommodation activities	S01
Q02	Arts, entertainment and recreation	S02
Q02	Construction and real estate activities	S03
Q02	Disaster risk management	S04
Q02	Education	S05
Q02	Energy	S06
Q02	Environmental protection and restoration activities	S07
Q02	Financial and insurance activities	S08
Q02	Forestry	S09
Q02	Human health and social work activities	S10
Q02	Information and communication	S11

Q02	Manufacturing	S12
Q02	Professional, scientific and technical activities	S13
Q02	Services	S14
Q02	Transport	S15
Q02	Water supply, sewerage, waste management and remediation	S16
Q02	Other	S17
Q03	Undefined	E00
Q03	>10	E01
Q03	10-49	E02
Q03	50-249	E03
Q03	250-4,999	E04
Q03	>5,000	E05
Q04	Unspecified	0
Q04	Africa	1
Q04	Asia	2
Q04	Europe	3
Q04	North America	4
Q04	South America	5
Q04	Other	6
Q05	Unspecified	0
Q05	There is no EA function	1
Q05	There are some architects scattered around the company	2
Q05	There is a well-defined EA function	3
Q05	EA is embedded in all strategic initatives	4
Q06	Unspecified	0
Q06	Not at all	1
Q06	EA is consulted for some strategic/transformation initiatives	2
Q06	EA is embedded into strategy execution	3
Q06	EA is consulted for strategy development	4

Q06	EA is a core pillar of strategy development	5
Q07	Unspecified	0
Q07	No	1
Q07	Partially	2
Q07	Yes	3
Q08	Unspecified	0
Q08	No	1
Q08	Partially	2
Q08	Yes	3
Q09	Unspecified	0
Q09	No	1
Q09	Partially	2
Q09	Yes	3
Q10	Unspecified	0
Q10	Never	1
Q10	Ad hoc or less frequently than annually	2
Q10	Annually	3
Q10	Quarterly	4
Q10	Monthly	5
Q10	Weekly	6
Q11	Unspecified	0
Q11	N/A - Assessments are not conducted.	1
Q11	A small number of capabilities involved in a project/initiative	2
Q11	At least 25% of the capability map	3
Q11	At least 50% of the capability map	4
Q11	At least 75% of the capability map	5
Q11	100% of the capability map	6
Q12	Unspecified	0
Q12	N/A - Assessments are not conducted	1
Q12	Not sure	2

Q12	No value is seen in more frequent assessments	3
Q12	Lack of management buy-in	4
Q12	Time & Effort involved	5
Q12	Other	6
Q13	Unspecified	0
Q13	Not at all	1
Q13	Somewhat	2
Q13	Very	3
Q13	Completely	4
Q14	Unspecified	0
Q14	Not at all	1
Q14	Somewhat	2
Q14	Very	3
Q14	Completely	4
Q15	Unspecified	0
Q15	Yes	1
Q15	Unsure	2
Q15	No	3
Q16	-	-
Q17	Unspecified	0
Q17	Not at all	1
Q17	Somewhat	2
Q17	Very	3
Q17	Completely	4
Q18	Unspecified	0
Q18	Not at all available	1
Q18	Somewhat available	2
Q18	Mostly available	3
Q18	Completely available	4
Q19	Unspecified	0

Q19	Not at all	1
Q19	Somewhat	2
Q19	Very	3
Q19	Completely	4
Q20	Unspecified	0
Q20	Unlikely: Don’t trust AI	1
Q20	Unlikely: Don’t trust the input data quality	2
Q20	Unlikely: Don’t trust the input data completeness	3
Q20	Unlikely: AI can’t understand the nuances of the business context	4
Q20	Unlikely: Other	5
Q20	Likely that they will trust the results	6
Q21	-	-
Q22	-	-