Preproceedings of the 21st Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 55–72

Human–AI Collaboration in Maturity Model Development: Evidence from Big Data Analytics Assessment Criteria

Marija Đukić
0000-0002-1136-4278
University of Belgrade
Faculty of Organizational Sciences
Email: marija.djukic@fon.bg.ac.rs

Jozo Dujmović
0000-0002-1715-2700
San Francisco State University
Department of Computer Science
University of Belgrade
Faculty of Organizational Sciences
Email: jozo@sfsu.edu

Ivan Luković
0000-0003-1319-488X
University of Belgrade
Faculty of Organizational Sciences
Email: ivan.lukovic@fon.bg.ac.rs

***Abstract*—The development of maturity models is a knowledge-intensive process requiring literature synthesis, domain knowledge and practical validation. In this paper, we evaluate assessment criteria for a Big Data Analytics maturity model (BDAMM) and investigate the potential of generative artificial intelligence (GenAI) as a cognitive amplifier in this context. The criteria proposed in previous work were independently evaluated by domain experts and GenAI tools. The results show agreement between human and GenAI evaluations while highlighting complementary strengths. Human experts primarily contributed to practical relevance, while GenAI proved effective in knowledge synthesis. The resulting criteria hierarchy establishes the basis for an assessment instrument based on Logic Scoring of Preference supporting explainable BDA maturity assessment. The findings demonstrate that the effectiveness of GenAI depends on iterative human–AI collaboration, domain expertise, and continuous critical evaluation. With this research, we contribute assessment criteria for BDAMM development and provide empirical evidence on the role of GenAI as a cognitive amplifier in knowledge-intensive research.**

I. Introduction

BIG DATA is commonly defined as "data that is too large, fast, and complex to be adequately managed using traditional methods or software within an acceptable time" [1]. The effective collection, processing, and analysis of large volumes of heterogeneous data provides organizations with opportunities to develop new capabilities, improve operational efficiency and decision-making, and introduce innovative products and services [2], [3]. The value organizations derive from Big Data depends on their strategic objectives but also on their ability to transform data into actionable knowledge through appropriate analytical methods and technologies [4]. These activities are encompassed by the concept of Big Data Analytics (BDA), which refers to the use of advanced analytical methods, tools, and infrastructure to process and analyze large, complex datasets to generate valuable insights and support decision-making [5].

Despite technological advances, organizations often struggle to fully realize the expected benefits of their investments in Big Data and Big Data Analytics [6]. The availability of data and enabling technologies is a prerequisite for Big Data initiatives, but their successful implementation is also determined by organization's capability to effectively manage and exploit them, which is reflected in its maturity [7]. Maturity can be defined as "a measure to evaluate the capabilities of an organization in regard to a certain discipline" [8]. Assessing Big Data Analytics capabilities (BDAC) is essential for understanding organization's current strengths and weaknesses and identifying opportunities for improvement. To support such assessments, maturity models (MMs) have emerged as "conceptual models that outline anticipated, typical, logical, and desired evolution paths towards maturity" [9]. In the context of Big Data and Big Data Analytics, maturity models enable organizations to evaluate their BDA capabilities and plan capability development toward greater value creation from Big Data [7].

The number of Big Data Analytics maturity models (BDAMMs) has increased over the years; however, several limitations remain. Models often provide only limited information about their development methodology and assessment process and are accompanied by insufficient documentation, reducing their transparency and reproducibility. Furthermore, relatively few models have undergone rigorous empirical validation or been evaluated through real-world case studies. The selection and coverage of assessment criteria also vary. Some models place strong emphasis on technological capabilities, while human and process-related aspects are often addressed only to a limited extent [10]. As a result, existing BDAMMs may not provide a comprehensive set of domain-specific assessment criteria capable of capturing the multidimensional nature of Big Data Analytics maturity.

Considering the identified limitations, we have proposed a set of assessment criteria for a Big Data Analytics maturity model [11]. The criteria were derived using a literature-based approach involving the analysis of existing maturity models and literature reviews from the Big Data and related domains, complemented by our domain knowledge and experience. This process resulted in approximately 120 assessment criteria organized into a three-level hierarchical structure comprising six domains: Organization, Human capital, Technology, Data, Processes, and Environment. The inclusion of the Environment domain addresses an important gap of existing BDAMMs, where external factors influencing the adoption and effective use of Big Data Analytics are frequently overlooked. These criteria establish the conceptual foundation of the proposed maturity model, which is being refined through an iterative process of development and validation. The first iteration focuses on evaluating the proposed BDAMM assessment criteria through a survey with industry experts. While expert judgment remains the primary source of evaluation, recent studies suggest that artificial intelligence (AI) can serve as a cognitive amplifier and support knowledge-intensive tasks [12]. In this role, AI functions as a collaborative partner, augmenting human thinking through cognitive offloading of information-processing tasks, therefore allowing human experts to focus on higher-level reasoning, critical evaluation, and decision-making. Building on this premise, we investigate the potential of generative AI (GenAI) as a cognitive amplifier through three complementary evaluation stages. First, the assessment criteria we proposed were validated by domain experts. Second, GenAI was used to generate and refine a hierarchy of assessment criteria based on literature. This hierarchy was analyzed to examine the capabilities of GenAI but was not incorporated into the proposed maturity model. Third, GenAI was used to evaluate the authors-developed assessment criteria, and these recommendations were compared with the domain expert validation results.

The objective of this research is twofold. First, we aim to evaluate the proposed assessment criteria for a Big Data Analytics maturity model through expert assessment complemented by AI-generated evaluations. Second, we investigate the potential of GenAI as a cognitive amplifier in the evaluation process. By comparing expert and AI-generated assessments, we seek to identify the similarities and differences between the two sources of evaluation and to examine the possibilities and limitations of using GenAI in this context. To support this objective, the following research questions are defined:

RQ1. To what extent can GenAI be used to provide meaningful evaluations of Big Data Analytics maturity model assessment criteria compared with human experts?

RQ2. What opportunities, limitations, and practical considerations emerge from using GenAI as a cognitive amplifier in maturity model development?

Answering these research questions provides insights into the validity of AI-assisted evaluations and the practical role of GenAI in maturity model development. Specifically, we explore whether AI-generated evaluations can complement expert judgment and identify the conditions under which GenAI can support the refinement of assessment criteria. The resulting refined criteria establish a foundation for the subsequent development and validation of the proposed BDAMM. More broadly, we contribute to the emerging body of knowledge on AI-assisted research by providing empirical evidence on the complementary roles of human expertise and GenAI and offer some practical guidance for integrating GenAI into knowledge-intensive research.

The remainder of this paper is organized as follows. Section 2 describes the development of the assessment criteria by synthesizing evidence from SLRs and existing BDAMMs. Section 3 presents the human-based evaluation of the proposed criteria, while Section 4 describes the AI-assisted evaluation. Section 5 discusses the findings by comparing human and AI evaluations and positioning the proposed model relative to existing maturity models. Finally, Section 6 concludes the paper including limitations and directions for future research.

II. Development of Assessment Criteria

One of the first steps in maturity model development is the identification and synthesis of model criteria. The assessment criteria of the Big Data Analytics maturity model we propose are presented in our previous work [11]. To provide the theoretical background and position the proposed criteria within the existing body of knowledge, we reviewed the literature in Big Data and related domains. We examined published systematic literature reviews (SLRs) that synthesized assessment criteria from maturity models, either as their primary objective or as part of a broader analysis and reviewed existing BDAMMs to identify common and distinctive characteristics. The following two subsections summarize the evidence obtained from these complementary sources.

A. Evidence from Systematic Literature Reviews

Systematic literature reviews provide a rigorous and transparent approach for identifying and consolidating evidence from existing research and are widely used in the development of maturity models. To establish the theoretical foundation for the proposed assessment criteria, we reviewed several SLRs covering the areas of Big Data, data and analytics, and artificial intelligence (Table I). The reviews covered both general-purpose and industry-specific maturity models. The MMs included in the reviews differ in the way assessment criteria are structured. Some comprise only a single level of assessment criteria, others organize them into hierarchical structures in which three to nine top-level domains are further decomposed into lower-level criteria. Similarly, the reviewed SLRs synthesize the identified criteria either as a single-level taxonomy or as a two-level hierarchical structure (Table I). In addition, different terms are used to describe the structural elements of maturity models, such as dimensions, constructs, indicators, and process areas. To be consist in the terminology, we use the term “criteria”, while referring to the highest level of the hierarchy as domains.

Despite structural differences, a high degree of consistency can be observed regarding the capabilities included in maturity assessments. Data and Technology appear in every reviewed study, reflecting their foundational role as assessment domains. Organization is also present across all reviews, indicating

TABLE I
ASSESSMENT CRITERIA IDENTIFIED IN SYSTEMATIC LITERATURE REVIEWS

Reference	Area of MMs	No. of levels	Top-level criteria
[13]*	Big Data	2	Organization and People, Strategy and Business, Governance and Trust, Management, Technology, Data and Analytics
[7]	Big Data	2	Organization, People, System Architecture, Data, Process
[14]	Data and Analytics	up to 4	Technology, Data, Organization
[15]*	Data and Analytics	1	Organization, Infrastructure, Data Management, Data Governance, Analytics Operations
[16]	Artificial Intelligence	1	Organization, Governance, People, Technology and Tools, Analytics, Data, Intelligent Automation
[17]	Artificial Intelligence	2	Organization, Strategy, Management/Leadership, People, Technology, Data, Resources, External Influences

*The criteria structure is not explicitly presented in the paper. The structure shown here represents the authors' interpretation based on analysis of the source.

that maturity extends beyond technological infrastructure to encompass organizational and managerial capabilities. People emerge as a consistently represented domain, highlighting the importance of employee competencies in data-driven transformation. Strategy and Governance receive increasing attention in more recent studies, extending assessment beyond operational capabilities to include ethical considerations, regulatory compliance, privacy, and security. Therefore, Technology, Data, Organization, People, Strategy, and Governance can be regarded as the core domains underpinning Big Data Analytics maturity assessment. At the same time, the reviewed literature demonstrates notable differences in the conceptual organization of these domains. Some models integrate strategy, governance, leadership, and management into broader domain of Organization, while others use them as distinct domains. Similarly, human and process-related capabilities may be incorporated within the domain of Organization or represented as independent assessment area. In contrast, criteria capturing external influences are rarely included, indicating that the external environment remains an underrepresented aspect of existing maturity models.

B. Evidence from Existing BDA Maturity Models

Existing Big Data Analytics maturity models provide insight into how assessment criteria are operationalized in practice. To identify common assessment domains and criteria, we analyzed the maturity models, readiness models, and capability frameworks included in our previous systematic literature review [10]. The comparison of top-level criteria is presented in Table II, while a detailed mapping of domains and lower-level criteria is provided as supplementary material.

The reviewed models exhibit considerable diversity, primarily in their scope but also in their structure. Most models organize assessment criteria into two hierarchical levels. Some models [4], [21], [26] employ three-level hierarchies, where the lowest level consists of assessment statements rather than additional criteria. The number of domains ranges from three to eight. Organization, Technology, Data, and People are the most frequently represented capability areas, although described using different terminology and levels of abstraction. Strategy and Governance are also very common, either as independent domains or as components of broader organizational constructs. At a more detailed level, recurring criteria relate to organizational culture, employee competencies, technological infrastructure, analytical technologies, data management and processes. These observations are consistent with the evidence synthesized from the systematic literature reviews, confirming the existence of a conceptual core underlying BDA maturity assessment. At the same time, the reviewed models differ in the conceptual organization of assessment criteria. Although some have expanded their scope to acknowledge environmental considerations, explicit assessment of external influences remains relatively uncommon.

III. HUMAN-BASED EVALUATION

The first stage of the criteria evaluation represents assessment of the initial set of proposed BDAMM criteria through a survey with human domain experts. The evaluation aims to determine whether the proposed criteria adequately capture the capabilities required to assess Big Data Analytics maturity and whether they are formulated in a clear and understandable manner. The human-based evaluation provides the foundation for the subsequent refinement of the criteria and serves as a benchmark for comparison with the AI-assisted evaluation presented later in the paper.

A. Initial Assessment Criteria Set

The evidence synthesized from SLRs and BDAMMs was consolidated into an initial set of assessment criteria complemented by our own experience in the research area. This version was presented in our previous work [11] and serves as the starting point for the evaluation and refinement process presented in this paper. The initial hierarchical structure of the proposed assessment criteria is summarized in Table V, which presents the domain and subdomain levels.

The proposed criteria are organized into a three-level hierarchy consisting of six domains: Organization, Human capital, Technology, Data, Processes, and Environment. These domains are further decomposed into 29 subdomains that represent major capability areas relevant to Big Data Analytics maturity. The third level contains assessment criteria for each subdomain and provides the basis for maturity assessment. In total, the initial criteria set comprises 121 assessment criteria. The hierarchical structure was designed to provide comprehensive coverage of the organizational capabilities required

TABLE II
ASSESSMENT CRITERIA IDENTIFIED IN EXISTING BIG DATA MATURITY MODELS

Reference	Type	No. of levels	Top-level criteria
[18]	Readiness model	2	Big Data Strategy; Policy and Collaboration; Infrastructure and Human Resources
[19]	Readiness model	1	Infrastructure; Governance; Culture; People; Process
[20]	Staged maturity model	2	Strategic Alignment; Organization; Governance; Data; Information Technology
[21]	Staged maturity model	2	Organization; Governance and Security; Infrastructure; Data Management; Data Analytics
[22]	Capability framework	2	Customer Relationship Management; Partner Life Cycle Management; Product/Service Life Cycle Management; Enterprise Risk Management; Strategy Development; Transformation Competence; Enterprise Architecture and Process Management; Information Management
[23]	Staged maturity model	1	Big Data Strategy; Benefits; Costs; Risks
[24]	Readiness framework	2	Data Ecosystem; Organization Specification; Resources; Data Management; Data Specification
[25]	Staged maturity model	2	Organization; Strategy; People; Technology; Data
[4]	Capability framework	2	Tangible Resources; Human Skills; Intangible Resources
[26]	Staged maturity model	2	Strategic Alignment; Organization; Data; Technology
[27]	Staged maturity model	1	Organizational and Attitude Factors; Information Technology; Technology; People Readiness
[28]	Staged maturity model	2	Strategy Alignment; Governance; People; Technology; Data; Methodology
[29]	Staged maturity model	1	Data/Knowledge; IT Solutions; Functionalities; Sustainable Development Context
[30]	Staged maturity model	1	Organization; Leadership; Data Governance and Integration; Analytics
[31]	Staged maturity model	2	Organizational; Human; Technological; Business

for successful Big Data Analytics maturity assessment while maintaining a clear separation between conceptually distinct capability areas. Compared with existing maturity models, the proposed structure explicitly distinguishes Organization, Human capital, and Processes as separate domains and incorporates Environment to capture external factors influencing organizational maturity.

B. Expert Validation Survey

The initial criteria set was evaluated through a survey completed by experts from industry to examine the relevance and clarity of the proposed criteria and identify areas requiring refinement. The survey targeted professionals with practical experience in Big Data Analytics, such as technical specialists and managers involved in the BDA initiatives. The survey instrument consisted of two sections. The first section was used to collect data about the respondents, such as industry sector, organization size, current position, professional experience and level of involvement in BDA initiatives (Appendix A). These questions were used to characterize the respondent sample (Table III) and assess the suitability of participants for evaluating individual domains.

The second section was used to evaluate the proposed criteria. Each assessment domain was presented separately with its hierarchical structure and descriptions of the corresponding criteria. Before evaluating it, respondents indicated their familiarity with the domain. Respondents mostly reported high familiarity with the five domains: Organization, Human capital, Technology, Data, and Processes. Familiarity with the Environment domain was often reported as low. For each criterion, respondents evaluated two attributes: relevance, reflecting the importance of the criterion for assessing Big Data Analytics maturity, and clarity, reflecting the comprehensibility and precision of its name and definition. Both attributes were assessed using a 10-point scale, where higher values indicate greater relevance or clarity. To reduce respondent fatigue associated with evaluating many criteria, the survey platform allowed participants to pause the survey and continue their responses later.

We collected six complete answers from experts representing diverse industries, such as healthcare, telecommunications, information technology, finance, and manufacturing. The organizations vary in sizes, whereas all participants had similar professional experience and experience working with BDA. Four respondents were directly responsible for BDA initiatives, and the remaining two were highly involved in such initiatives. Self-assessed expertise in BDA ranged from intermediate to advanced, indicating sufficient practical knowledge to evaluate the proposed assessment criteria.

The survey results are analyzed using descriptive statistics for the relevance and clarity scores of each assessment criterion. The analysis aimed to identify criteria requiring revision while also examining the overall distribution of scores across the proposed assessment domains. To support the subsequent refinement process, the criteria were classified into four groups based on predefined thresholds: Strong Keep (both relevance and clarity scores ≥ 9), Keep (both scores between 8 and 9), Need Intervention (scores between 7 and 8 or uneven relevance and clarity), and Candidate for Removal (both scores below 7). The average relevance and clarity scores for each assessment criterion are presented in Figure 1.

Overall, 70 of the 121 criteria (57.8%) received both relevance and clarity scores of at least 8. Forty-eight criteria (39.7%) were classified as “Need Intervention”, indicating that they should be reviewed and refined before the next iteration of the model. Only three criteria (2.5%) were identified as “Candidates for Removal”, having received both relevance and

TABLE III
CHARACTERISTICS OF THE INDUSTRY EXPERT RESPONDENTS

ID	Industry (No. of employees)	Position/Role	Professional experience (Experience with BDA)	Involvement in BDA initiatives (BDA expertise)
1	Healthcare/Pharmaceuticals (250–499)	Technical role	3–5 years (3–5 years)	Highly involved (Intermediate)
2	Telecommunications (1000 or more)	Technical role	3–5 years (3–5 years)	Directly responsible (Intermediate)
3	Information Technology/Software (500–999)	Technical role	3–5 years (3–5 years)	Highly involved (Advanced)
4	Finance/Banking (50–249)	Manager (IT-related)	6–10 years (6–10 years)	Directly responsible (Advanced)
5	Information Technology/Software (50–249)	Technical role	6–10 years (3–5 years)	Directly responsible (Advanced)
6	Manufacturing/Production (1000 or more)	Technical role	3–5 years (3–5 years)	Directly responsible (Advanced)

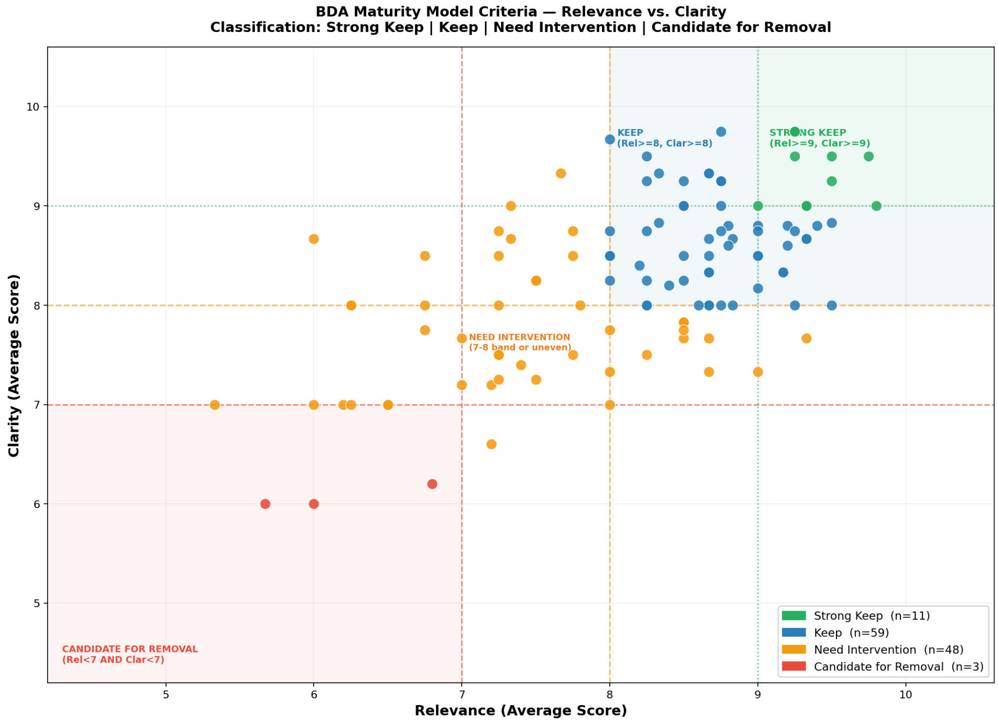

Fig. 1. Expert Evaluation of Assessment Criteria: Relevance and Clarity Scores

clarity scores below 7. The cross-analysis presented in Figure 2 illustrates the relationship between the mean relevance and clarity scores assigned to each assessment criterion. Each cell represents the number of criteria corresponding to a particular combination of relevance and clarity scores. The concentration of criteria in the upper-right region indicates that most assessment criteria were perceived as highly relevant and clearly formulated. Relatively few criteria received low scores for both attributes simultaneously, suggesting that only a limited subset exhibited substantial deficiencies. A more detailed examination reveals that criteria receiving lower evaluations were more frequently associated with lower relevance than lower clarity. In other words, respondents were generally able to understand the intent and wording of the proposed criteria but questioned their importance for assessing Big Data Analytics maturity. This pattern suggests that the primary challenge concerns the selection and positioning of certain assessment criteria rather than their definitions. In practice, this involves reassessing whether these criteria should be retained, reorganized or replaced by criteria that more accurately reflect BDA capabilities. This observation is further supported by qualitative feedback which questioned more the suitability of certain capability areas rather than the clarity of their descriptions.

Figure 3 summarizes the average relevance and clarity scores for each assessment domain. Organization achieved the highest average relevance score (8.75), followed by Data (8.37), while Processes (7.40) and Environment (7.67) received the lowest relevance evaluations. A small number of subdomains received lower relevance ratings, such as Process digitalization, Process capability level, and Market, which reduced the overall scores of the Processes and Environment domains. These indicate that the lower domain score reflects concerns about specific capability areas rather than with the

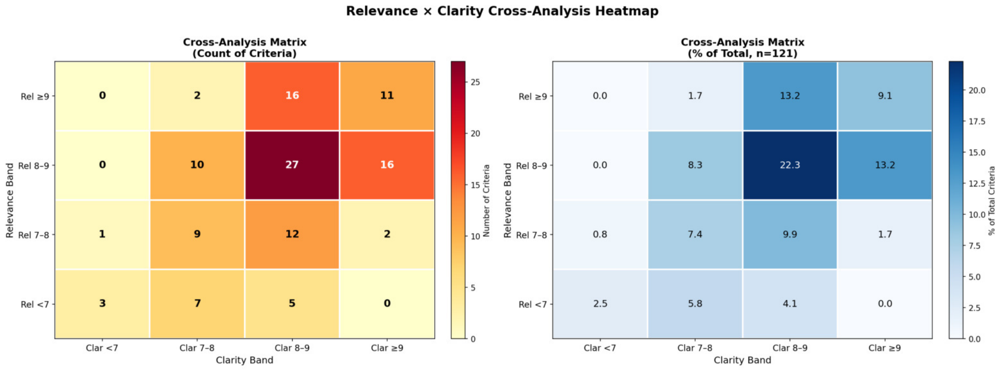

Fig. 2. Distribution of Assessment Criteria by Relevance and Clarity Score Bands

domains as a whole and therefore point to targeted refinement rather than comprehensive restructuring. Clarity scores were generally higher and more consistent across domains, ranging from 8.03 for Human Capital to 8.71 for Technology. Notably, the Environment domain achieved a relatively high clarity score (8.54) despite its lower relevance score, which may reflect respondents' comparatively lower familiarity with this domain. Criteria are generally perceived as clear and relevant, while the lower relevance scores observed for the Processes and Environment domains suggest that these domains warrant further validation and refinement in subsequent iterations of the model.

IV. AI-Assisted Evaluation

The second stage of the criteria evaluation investigates the potential of GenAI to support the validation of Big Data Analytics maturity model assessment criteria. GenAI is explored as a cognitive amplifier that can assist in information-processing tasks. To reduce dependence on a single tool, Microsoft Copilot and SciSpace were used in this research. The selection of these tools was based on the authors' familiarity and practical experience rather than on predefined selection criteria. Although the results are presented for each tool, the objective is not to compare their performance but to examine the possibilities and limitations of GenAI in general. To explore the boundaries of current GenAI capabilities, the AI-assisted evaluation comprises two parts. In the first, we explore the ability of GenAI to generate a hierarchical set of assessment criteria from the existing literature and subsequently review and refine its own output. In the second part, we validate the initial assessment criteria developed through literature synthesis and our domain knowledge, allowing comparison between AI-generated evaluations and the human expert assessments presented in the previous section.

A. AI-Assisted Criteria Generation

In the first part of the AI-assisted evaluation, we explore the capability of GenAI to generate and refine a criteria tree for BDAMM. The aim is to examine whether current GenAI tools can synthesize evidence from existing literature into a coherent maturity model structure, for later comparison. The same prompting strategy was applied to both Microsoft Copilot and SciSpace to ensure consistency of the evaluation. The complete prompts are provided in Appendix B and Appendix C. In the first step, each tool was instructed to generate a three-level hierarchical criteria tree based on the available literature on Big Data Analytics and related maturity models. In the second step, the generated hierarchy was reviewed by the same tool using an evaluation prompt. The generated and refined criteria were analyzed using three complementary perspectives: (1) structural analysis - examining the number of domains and subdomains, hierarchical depth, and overall organizational logic; (2) thematic evolution - tracing conceptual changes between the initial and refined hierarchies, including additions, removals, relocations, and semantic refinements; (3) BDA-specificity - evaluating the extent to which the generated hierarchies incorporate concepts and terminology specific to the Big Data Analytics domain. Table IV summarizes the initial and refined hierarchies generated by each tool.

The Copilot-generated hierarchy maintained a stable six-domain structure while undergoing reorganization at the sub-domain level. The initial hierarchy comprised six domains, Strategy and Governance, Organization and People, Technology, Data, Analytics, and Value Realization, with three sub-domains in five of the six domains. The refined hierarchy preserved the overall structure with more granular organization, expanding the domains Data, Analytics, and Organization and People from three to four sub-domains. The refinement process introduced several thematic enhancements. The original Strategy and Governance domain was divided into Strategy & Value Management and Governance, Risk & Compliance, separating strategic planning from governance mechanisms. The governance domain was further expanded to include Data governance, Analytics governance, and Risk, Privacy & Ethics, reflecting increasing recognition of regulatory and ethical considerations in BDA initiatives. In addition, organizational transformation became more prominent through the introduction of Change & Learning as a dedicated sub-domain and the stronger emphasis on a Data-driven culture, indicating that BDA maturity extends beyond technical capabilities. The refined hierarchy also exhibited greater BDA

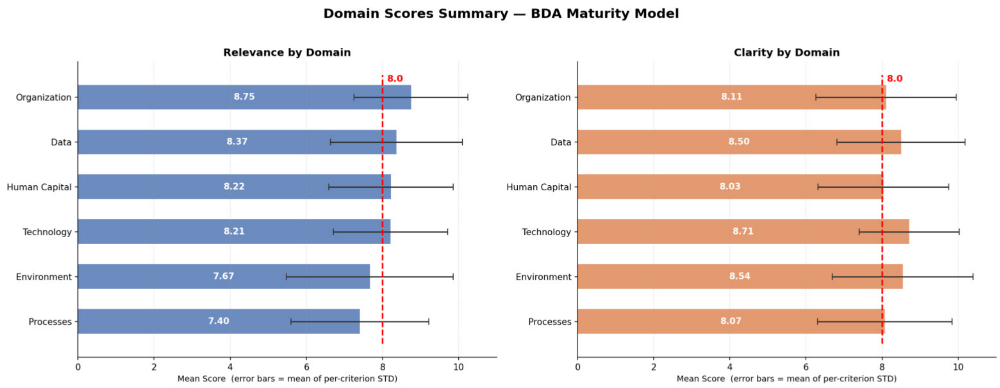

Fig. 3. Domain-Level Summary of Relevance and Clarity Scores

TABLE IV
OVERVIEW OF AI-ASSISTED CRITERIA GENERATION

	Microsoft Copilot		**SciSpace**	
	AI-generated	**AI-refined**	**AI-generated**	**AI-refined**
Number of domains	6	6	6	7
Domain names	(1) Strategy and Governance (2) Organization and People (3) Technology (4) Data (5) Analytics (6) Value Realization	(1) Strategy & Value Management (2) Governance, Risk & Compliance (3) Data & Information Architecture (4) Technology & Platform Capability (5) Analytics Capability (6) Organization, People & Culture	(1) Strategy and Governance (2) Data Management (3) Technology and Infrastructure (4) People and Organization (5) Analytics Processes and Capabilities (6) Value and Impact	(1) BDA Strategy and Management (2) Data and Analytics Governance (3) Data Management (4) Technology and Infrastructure (5) Human Capital and Organisation (6) Culture and Change (7) Analytics Processes and Capability
Number of subdomains	17	21	18	19
Number of criteria	53	64	58	63

specificity using more specialized terminology and concepts. The explicit inclusion of Big Data Analytics capability as a sub-domain distinguishes the framework from more generic business intelligence or analytics maturity models. Likewise, separating Analytics strategy from broader business strategy and introducing Portfolio management for analytics initiatives reflects specific BDA capabilities. Nevertheless, several components remained generic. For example, the Human capital and Organizational structure sub-domains describe capabilities common to many organizational maturity models without explicitly capturing competencies or roles unique to Big Data Analytics.

The SciSpace-generated hierarchy underwent a more substantial structural transformation. The initial version comprised six domains with 18 sub-domains (three per domain): Strategy & Governance, Data Management, Technology & Infrastructure, People & Organization, Analytics Processes & Capabilities, and Value & Impact. The refined hierarchy expanded to seven domains and 21 sub-domains while preserving the Technology & Infrastructure domain largely unchanged. The most significant change was the separation of Strategy & Governance into two independent domains, elevating governance to a standalone capability area. A second major transformation involved extracting Culture & Change from People & Organization into an independent domain, while in contrast, the original Value & Impact domain was integrated into new BDA Strategy & Management domain. Data Management and Analytics Processes & Capabilities domains were expanded with additional sub-domains, providing a more detailed representation of data and analytical capabilities. Several thematic enhancements emerged during refinement. Responsible AI and data governance received greater attention through the introduction of a dedicated Responsible Analytics & AI Governance sub-domain within Data & Analytics Governance domain. Analytical capabilities were described with greater granularity by distinguishing Analytical capability depth as a separate sub-domain. User empowerment became an explicit aspect of the hierarchy through Analytics Delivery & User empowerment, reflecting contemporary trends such as self-service analytics. Organizational learning also received greater emphasis through the Organizational Learning & Analytics Engagement sub-domain within Culture & Change domain. Compared with Copilot, SciSpace demonstrated a higher degree of BDA-specificity across multiple domains. The consistent use of the BDA prefix anchored the framework in the context of Big Data Analytics rather than generic analytics or information technology maturity. Furthermore, the Advanced Analytics & AI Technologies sub-domain explicitly incorporated machine learning, deep learning, natural language processing, and other AI techniques characteristic of contemporary BDA implementations. Overall, the refined SciSpace hierarchy placed stronger emphasis on governance, responsible AI, organizational transformation, and advanced analytical capabilities resulting in a more specialized framework.

The analysis of the Copilot- and SciSpace-generated criteria provides several insights into the use of GenAI for maturity model development. Both tools demonstrated a strong ability to synthesize evidence from diverse literature into hierarchical criteria structures. Despite differences in organization and terminology, both tools identified core capability domains present in the literature, such as Organization, Strategy, Governance, People, Technology, and Data. This suggests that current GenAI tools can effectively support literature synthesis and the initial conceptualization of maturity model structures. However, the generated hierarchies are largely derivative, reorganizing and consolidating existing knowledge, thereby reinforcing the role of GenAI as a tool for knowledge synthesis rather than for conceptual innovation. The comparison between the initial and refined hierarchies demonstrates the importance of iterative refinement. For both tools, refinement resulted in structures that were conceptually richer and better aligned with current Big Data Analytics practices. These improvements were driven primarily by the refinement prompts, indicating that prompt engineering plays a critical role in AI-assisted tasks. However, both tools exhibited limitations, including incomplete operationalization of certain concepts and omission of some technical and organizational aspects. Furthermore, the level of abstraction varied across domains, with some representing strategic capabilities and others describing implementation mechanisms or technologies. Human experts are required to ensure conceptual consistency, incorporate recent developments that may not yet be reflected in the underlying models, resolve structural trade-offs between simplicity and granularity, and transform conceptual domains into measurable assessment criteria. AI-generated criteria could be regarded as preliminary design artifacts that require systematic validation through expert review and empirical evaluation. The differences between Copilot and SciSpace also show that different GenAI tools exhibit different strengths. Copilot produced a structurally balanced hierarchy with a stronger emphasis on value and portfolio management, while SciSpace generated a more domain-specific framework incorporating responsible AI governance, advanced analytics capabilities, and BDA terminology. These differences suggest that AI-assisted maturity model development may benefit from hybrid approaches that combine outputs from multiple tools before human synthesis and validation.

B. AI-Assisted Criteria Validation

In the second part of the AI-assisted evaluation, GenAI was used as a cognitive amplifier to support the refinement of the proposed criteria. This phase was used to critically evaluate the initial criteria derived from the literature and our experience, and identify opportunities to improve their relevance, clarity and operationalization. The prompt used for the evaluation is provided in Appendix D, while the comparison of initial and final versions of the assessment hierarchy is presented in Appendix E.

The validation was conducted using the same GenAI tools employed in the criteria generation phase, Microsoft Copilot and SciSpace, to evaluate the complete three-level hierarchy of the proposed BDAMM. The evaluation prompt instructed the tools to assess the hierarchy with respect to completeness, redundancy, mutual exclusiveness, consistency of abstraction, logical placement, terminology, and suitability for maturity assessment. Based on these instructions, the tools were requested to critically evaluate each domain, subdomain, and assessment criterion to identify conceptual overlaps, missing capabilities, inconsistencies in abstraction levels, and opportunities for improving the hierarchical organization of the model. For every proposed modification, the GenAI tools were instructed to provide a rationale supported by relevant literature and existing maturity models.

The refinement process was iterative, consisting of several validation cycles. After each AI-generated evaluation, we analyzed the proposed modifications and compared them with the referenced literature and the design objectives of the proposed BDAMM. Accepted modifications were incorporated into the hierarchy before initiating the next validation cycle. This approach enabled the progressive improvement of the assessment criteria while ensuring that all design decisions remained under our control and were consistent with the intended conceptual structure of the maturity model. Refinement refers to the systematic improvement of the assessment criteria hierarchy with respect to its completeness, relevance and suitability for maturity assessment. It included adding missing capabilities identified during the evaluation, renaming subdomains and assessment criteria to improve clarity, restructuring the hierarchy to reduce overlaps, merging or removing redundant criteria, and improving the operationalization of assessment criteria to facilitate subsequent measurement.

Table V summarizes the structural changes resulting from the refinement process, which combined recommendations derived from expert validation and AI-assisted evaluation. The six top-level domains remained unchanged, confirming the stability of the conceptual foundations established through the literature synthesis. Several subdomains, however, were renamed or reorganized to better reflect BDA-specific capabilities. Within the Human capital domain, the emphasis shifted toward analytics-specific capabilities. Similar refinements were introduced across the Technology, Processes, and External ecosystem domains, where generic concepts were replaced by more operational and BDA-oriented capability areas. Rather than autonomously redesigning the assessment criteria, GenAI was used as a cognitive amplifier to compare the proposed hierarchy with existing literature and maturity models and suggest literature-informed modifications. We subsequently evaluated these recommendations and selectively incorporated them into the model. The final assessment hierarchy reflects the combination of systematic knowledge synthesis provided by GenAI and human judgment and decision-making implemented through a human-in-the-loop refinement process.

V. Discussion

The discussion is organized around four perspectives. First, we compare the human expert and GenAI evaluations to examine the extent of agreement between the two approaches and the contributions of each approach. Then, we compare the proposed BDAMM with previously reviewed maturity models to position its conceptual structure and highlight its methodological contributions. Next, we analyze the context dependence of the assessment criteria to determine the extent to which the model requires customization for different organizational environments. Finally, we discuss the potential and limitations of GenAI as a cognitive amplifier, considering their broader implications for AI-assisted maturity model development, validation, and other knowledge-intensive research and design activities.

A. Comparison of Human and AI Validations

The refinement of the proposed assessment criteria was informed by two complementary sources: (1) the domain expert validation survey, and (2) AI-generated validations produced using Microsoft Copilot and SciSpace. Survey responses were analyzed using predefined thresholds to identify criteria with low relevance (mean relevance score < 8.0), low clarity (mean clarity score < 8.0), and high disagreement (standard deviation ≥ 2.0 for relevance or clarity). The AI evaluation consolidated the recommendations independently generated by both GenAI tools. Overall, 48 refinement proposals were identified, comprising 24 proposals originating only from the domain expert validation, 10 recommendations independently supported by both human and GenAI validations, 14 proposals generated exclusively by GenAI (4 by both tools, 6 by SciSpace and 4 by Microsoft Copilot). Table VI presents the overlapping refinement recommendations identified independently through the expert validation and the AI-assisted evaluation. The complete set of recommendations is available in the supplementary material, together with rationale, implementation status, and resulting changes. As a result, 30 recommendations were fully implemented, 11 partially, and 7 recommendations were rejected.

In response to RQ1, the comparison between human experts and GenAI-generated validations demonstrates a high degree of agreement in identifying the most important areas requiring refinement. Both sources highlighted conceptual weaknesses within External expertise, Process capability level, Institutions, and Market subdomains, leading to major structural revisions. These included replacing the Process capability level with the more BDA-specific Analytics lifecycle management subdomain and redefining the external environment subdomains to assess organizational capabilities rather than external conditions. The overlap between human and GenAI recommendations suggests that fundamental conceptual issues can be identified through different validation approaches. Human experts identified criteria with low practical relevance, insufficient clarity, or inconsistent interpretation. On the other hand, GenAI validations frequently provided literature-based justifications for restructuring the model and proposed alternative structure or additional capabilities. Expert validation proved particularly valuable for assessing the practical applicability and measurement quality of assessment criteria, while GenAI contributed more strongly to improving conceptual consistency and theoretical alignment with existing BDA maturity research.

The GenAI tools also generated several proposals that were not raised in human expert validation. These primarily involved extending the model to incorporate emerging BDA capabilities, such as Data literacy, Analytics and AI regulatory readiness, and the distinction between technical analytics competencies and business and domain competencies. These recommendations reflect more recent developments reported in the literature, illustrating the ability of GenAI to expand the conceptual scope of a maturity model. However, not all GenAI-generated recommendations were adopted. Proposals such as relocating Leadership capability from the Human capital domain to Organization or elevating Strategy to an independent top-level domain were intentionally rejected because they conflicted with the underlying conceptual design principles of the proposed BDAMM. These examples indicate that literature consensus alone is insufficient for guiding maturity model development and that AI-generated suggestions require critical expert evaluation before implementation. Individual design decisions must also remain consistent with the model's theoretical foundations and intended assessment objectives.

The refinement process illustrates the complementary roles of human experts and GenAI validations. Domain experts contributed contextual knowledge, practical experience, and assessment-oriented judgment, while GenAI contributed systematic literature synthesis, conceptual restructuring, and the identification of emerging capability areas. Neither source alone proved sufficient for developing the final assessment criteria. The highest-quality refinements emerged where human expert and GenAI evaluations converged, while AI-exclusive recommendations served as an additional source of ideas that were selectively incorporated following critical review. These observations support the central premise of this research that GenAI functions most effectively as a cognitive amplifier, augmenting expert reasoning during maturity model development. Overall, these findings answer RQ1 by demonstrating that GenAI can provide meaningful evaluations of maturity model assessment criteria that substantially overlap with expert judgment while contributing complementary insights through literature synthesis and conceptual refinement.

B. Comparison with Existing Maturity Models

To evaluate the completeness and conceptual positioning of the proposed assessment criteria, we compared them with the maturity models and capability frameworks presented in Section 2.B. The comparison encompassed the hierarchical or-

TABLE V
REFINED ASSESSMENT CRITERIA OF THE PROPOSED BIG DATA ANALYTICS MATURITY MODEL

Domains (Level 1)	Subdomains (Refined Set of Criteria)	Subdomains (Initial Set of Criteria)
1. Organization	1.1. Organizational structure 1.2. Strategy 1.3. Governance 1.4. Culture 1.5. Analytics management	1.1. Organizational structure 1.2. Strategy 1.3. Governance 1.4. Culture 1.5. Analytics management
2. Human capital	2.1. Employee competences 2.2. Analytics leadership behavior 2.3. Analytics talent management 2.4. Analytics external expertise	2.1. Employee competences 2.2. Leadership capability 2.3. Support to organizational culture 2.4. External expertise
3. Technology	3.1. Infrastructure 3.2. Architecture 3.3. Analytics technologies 3.4. Analytics environment management	3.1. Infrastructure 3.2. Architecture 3.3. Analytics tools and technologies 3.4. Technology management
4. Data	4.1. Data quality 4.2. Data sources 4.3. Data integration 4.4. Data governance 4.5. Data accessibility 4.6. Data lifecycle management 4.7. Metadata management	4.1. Data quality 4.2. Data sources 4.3. Data integration 4.4. Data governance 4.5. Data accessibility 4.6. Lifecycle management 4.7. Metadata management
5. Processes	5.1. Process automation 5.2. Analytics integration 5.3. Process design and change 5.4. Analytics adoption 5.5. Analytics lifecycle management	5.1. Process automation 5.2. Process digitalization 5.3. Process integration 5.4. Process design and change 5.5. Process scope 5.6. Process capability level
6. External ecosystem	6.1. Customer analytics engagement 6.2. External collaboration 6.3. Market positioning 6.4. Regulatory environment	6.1. Clients and users 6.2. Institutions 6.3. Market

TABLE VI
OVERLAPPING HUMAN AND GENAI REFINEMENT RECOMMENDATIONS

Human Validation	AI Validation	Rationale	Change
Redefine or merge criteria in 2.4 External expertise	Consolidate 2.4 External expertise criteria	Low relevance and clarity across the subdomain	Revised
Remove or substantially revise 5.6 Process capability level	Replace 5.6 Process capability level with Analytics lifecycle management	Low relevance, high disagreement; circularity in the maturity model; generic process maturity rather than BDA-specific	Replaced subdomain
Remove or substantially revise all criteria in 6.2 Institutions	Rename and retain only capability-based criteria	External conditions should not be scored as organizational maturity	Revised and redefined
Remove or substantially revise 6.3 Market	Rename and retain only capability-based criteria	External conditions should not be scored as organizational maturity	Revised and redefined
Reconsider 1.1.4 Centralization	Remove or refine 1.1.4 Centralization	High disagreement in relevance ratings	Redefined
Clarify criterion definition 1.2.5 Value creation potential of big data	Add Analytics value realization (replacing Value creation potential)	High clarity standard deviation; measure realized rather than expected value	Redefined
Remove 2.3.5 Trust in organizational management	Remove 2.3.5 Trust in organizational management	Low relevance; not BDA-specific	Removed
Reconsider 4.2.4 Data variety	Merge 4.2.3 Source diversity and 4.2.4 Data variety	High disagreement in relevance ratings	Merged 4.2.3 and 4.2.4
Address governance overlap between 1.3 Governance and 4.4 Data Governance	Clarify the hierarchy between 1.3 Governance and 4.4 Data Governance	Theoretical clarity; literature distinguishes the two concepts	Renamed for clarity
Add human-centricity and greater emphasis on ethics, sustainability, and privacy	Add Transparency and Explainability to Governance; add Privacy and Data Protection to Data Governance	Address emerging AI governance and responsible analytics requirements	New criteria added

ganization of assessment criteria and the conceptual coverage of individual domains and subdomains.

The comparison confirms that the proposed criteria incorporate the core capability areas consistently recognized in the literature while providing broader coverage than any of reviewed models. The four domains, Organization, Human capital, Technology, and Data, are well established in previous research, each appearing in ten or more (out of 15) reviewed models. Technology and Data are consistently represented as dedicated top-level domains, under different names, while Organization and Human capital vary across models. Despite this strong conceptual overlap, the proposed structure differs from previous work in the way these capability areas are organized. Organization is treated either as a standalone domain or alongside separate Strategy and Governance domains, whereas human-related capabilities are represented either as an independent People domain or as part of broader organizational constructs. The proposed structure adopts a different perspective by positioning Strategy and Governance as subdomains of Organization rather than as independent top-level domains. This design choice reflects the view that strategic direction and governance mechanisms are organizational capabilities that enable the effective deployment of Big Data Analytics. Similarly, Human capital is established as a separate domain to distinguish organizational arrangements from the competencies, leadership behaviors, talent management practices, and external expertise required to develop and sustain BDA capabilities. These design decisions represent a deliberate departure from the dominant structures found in the literature to provide greater diagnostic granularity for organizational maturity assessment. Second distinguishing characteristic is the introduction of Processes and External ecosystem as dedicated top-level domains. Although process-related capabilities appear in approximately half of the reviewed models, they are usually embedded within organizational or technology domains and primarily address process management or business process integration. The proposed criteria extend this perspective by including process automation, analytics adoption, and analytics lifecycle management as subdomains. External ecosystem domain represents the most distinctive contribution to the proposed criteria. External factors such as regulatory requirements and customer orientation appear in several existing maturity models but embedded within organizational or governance domains. By establishing a dedicated top-level domain, the proposed model acknowledges that organizational BDA maturity depends not only on the development of internal capabilities, but also on an organization's ability to interact with and respond to its external environment. The closest conceptual equivalents are provided in [22], [31], but neither treats these capabilities as standalone maturity domain.

A more detailed comparison at the subdomain level reveals that, although equivalent concepts are present in reviewed models, the coverage of many capability areas remains partial. The proposed criteria extend domains by distinguishing capabilities that are aggregated or omitted in existing models. Human capital is expanded with inclusion of talent management and external expertise, compared to reviewed models that only included employee competencies and less frequent leadership. Similarly, the Technology domain includes architecture as separate subdomain, which is part of relatively few models, compared to infrastructure and analytics technologies that are among the most frequently represented capabilities. Within the Data domain, Metadata management represents another capability with only limited coverage in the reviewed literature. The least represented subdomains are associated with the Processes and External ecosystem domains. Capabilities such as customer analytics engagement, external collaboration, market positioning, process automation, process design and change, and analytics lifecycle management appear in a very small number of existing models, representing areas that have received little attention in previous maturity research. Their inclusion extends the assessment beyond traditional organizational and technological capabilities toward the operational and environmental factors that influence the successful adoption and exploitation of Big Data Analytics. Finally, the comparison highlights one methodological difference in the hierarchical organization of assessment criteria. Reviewed models employ one- or two-level structures in which broad capability areas directly contain assessment criteria. The proposed criteria are organized into three-level hierarchy consisting of domains, subdomains, and measurable assessment items.

This hierarchical structure provides the foundation for implementing the assessment instrument using the Logic Scoring of Preference (LSP) method [32]. LSP supports both the evaluation of a single alternative and the comparison of multiple alternatives. In the context of the proposed BDAMM, this enables the assessment of an individual organization's BDA maturity, the comparison of multiple organizations, and, more importantly, the evaluation and prioritization of alternative improvement scenarios. Based on Graded Logic [33], LSP models decision-making using observable and measurable human logical reasoning and provides three complementary aspects of explainability. In the proposed BDAMM, LSP can be used not only to produce an overall maturity score but to transparently explain how individual assessment criteria contribute to subdomain, domain, and overall maturity score, why a specific organization or improvement scenario attains or fails to attain particular degree of BDA maturity, and why one improvement scenario is preferred over another. This level of explainability enhances the interpretability of assessment results and provides decision makers with clear justification for identifying capability gaps and prioritizing improvement initiatives.

C. Context Dependence of Assessment Criteria

To evaluate the generalizability of the proposed assessment criteria, we further classified each criterion according to its degree of context dependence. The objective was to distinguish criteria that can be applied uniformly across organizations from those that require adaptation when the model is customized for different contexts, such as industry sectors.

The proposed classification comprises four categories: (1) universal - criterion is applicable regardless of industry or size, its meaning, scope, and assessment logic remain unchanged, requiring no contextual adaptation; (2) context-adaptive - criterion is relevant across organizations, but its measurement or importance should be calibrated to the organizational context; (3) domain-conditional - applicability of the criterion depends on the presence of specific organizational capability (e.g., reliance on external expertise), if these conditions are absent, the criterion may be of limited relevance; and (4) industry-specific - criterion primarily applies to organizations operating within particular industries, its interpretation and measurement depend on sector-specific regulations or business practices and therefore require substantial adaptation before being transferred to other domains. Table VII presents the distribution of the proposed criteria across the defined context dependence categories. Individual assessment criteria are listed unless all criteria within a subdomain belong to the same category, in which case the subdomain is shown instead. The category Universal is not included in the table for readability because it contains a large number of proposed assessment criteria.

Most of the assessment criteria are universal (62 of 115), suggesting that the proposed model possesses a high degree of transferability, with its overall domain structure and most assessment criteria remaining applicable regardless of context. The largest group of non-universal criteria is context-adaptive category (41 of 115), representing the primary customization layer of the model. When adapting the model to specific industry sector, these criteria require adjustments for measurement scales or benchmark thresholds. Context-adaptive criteria are concentrated in capability areas that are influenced by organizational characteristics, technological complexity, and external conditions, such as governance, employee competencies, infrastructure, analytics technologies, data management, process automation, and market positioning. For example, Cloud computing infrastructure (criterion 3.1.2) measures the extent to which an organization leverages cloud computing platforms to support its analytics capabilities. While a universally relevant capability, its implementation depends on industry-specific and operational requirements. Banks often adopt hybrid cloud architecture, keeping sensitive financial and risk data on-premise while executing less sensitive workloads in the cloud. Healthcare organizations may follow a cloud-first strategy but must comply with data sovereignty requirements, restricting the storage and processing of patient data across jurisdictions. In contrast, software companies are frequently cloud-native with fewer industry-specific constraints. Similarly, the assessment of Data storage and retention (criterion 4.6.1) depends on industry-specific requirements. Financial institutions are typically required to retain trading records for five to seven years, while healthcare organizations must comply with GDPR [34] and national healthcare regulations, with clinical trial data often retained for 15–25 years after study completion. In contrast, manufacturing companies generally have no mandatory sector-specific data retention requirements.

A small number of criteria (10 of 115) were classified as domain-conditional, indicating that their relevance depends on whether an organization exhibits specific characteristics or performs particular activities. These criteria are assessed if the underlying capability is relevant to the organization and primarily concern the use of external expertise, collaboration with external partners, customer analytics engagement, and customer-facing processes. For example, Access to external expertise (criterion 2.4.1) measures the extent to which an organization has established mechanisms for obtaining analytics expertise from external sources. This criterion is relevant for organizations in the early stages of analytics adoption or organizations lacking specialized analytics capabilities, where external expertise represents an important enabler of BDA maturity. In contrast, large technology companies or analytics-native organizations that have developed comprehensive in-house capabilities may have little or no need for external analytics support. Similarly, Customer-facing processes (Criterion 5.4.3) is applicable only to organizations whose operations involve direct interaction with external customers. In customer-oriented businesses, analytics can support personalization, customer engagement, and service optimization. However, organizations such as regulatory agencies or government institutions may have no customer-facing processes. Additionally, the concept of a customer varies. For example, B2C retailers interact with large numbers of individual consumers, while B2B manufacturers serve a smaller number of long-term industrial clients, resulting in different applications of analytics within customer-facing processes.

The industry-specific category contains only two criteria, both related to regulatory and standards compliance. Unlike other organizational capabilities, these criteria are determined directly by sector-specific standards and compliance requirements. As a result, they must be redefined for each industry-specific implementation of the model. For example, Regulatory and legal compliance (criterion 6.4.1) measures an organization's ability to comply with the legal and regulatory framework governing its analytics and data operations. Some regulations, such as GDPR, apply across multiple sectors, but primary compliance requirements are industry specific. Banking organizations, for example, must comply with regulations such as BCBS 239 governing risk data aggregation and reporting, while healthcare organizations must comply with national healthcare legislation like HIPAA in the United States. Similarly, Industry standards compliance (criterion 6.4.2) is inherently sector-specific and cannot be meaningfully assessed using a common framework. For example, software organizations may evaluate compliance with standards such as CMMI, ITIL or ISO/IEC 27001.

The classification shows that contextual adaptation is required primarily at the measurement level rather than at the conceptual level. The domain hierarchy and most assessment criteria remain universally applicable, while a subset requires contextual tailoring to ensure valid and meaningful maturity assessment. This classification also aligns well with the planned LSP-based assessment approach. Measurement of

TABLE VII
CLASSIFICATION OF CRITERIA BASED ON CONTEXT DEPENDENCE

Type	Criteria
Context-Adaptive	1.1.2. Specialization of analytics function, 1.2.4. Analytics business value measurement 1.3.3. Analytics governance policies, 1.3.4. Ethics and responsible analytics, 1.3.5. Transparency and explainability 2.1.1. Technical analytics competencies, 2.1.2. Business and domain competencies 2.3.1. Analytics employee satisfaction, 2.3.2. Analytics employee motivation, 2.3.3. Analytics talent retention 3.1.1. On-premise computing infrastructure, 3.1.2. Cloud computing infrastructure, 3.1.3. Network and data transfer capacity 3.1.5. Analytics infrastructure resilience, 3.2.3. Security architecture, 3.2.4. Architectural standards 3.3.2. Data processing technologies, 3.3.4. Data visualization and reporting tools, 3.3.5. ML and AI technologies 4.1.1. Data quality monitoring, 4.2.1. External data acquisition, 4.2.3. Data source and type diversity 4.3.2. External data integration, 4.3.3. Data harmonization and modeling, 4.3.4. Master data management 4.4.4. Privacy and data protection, 4.5.4. Data sharing and exchange 4.6.1. Data storage and retention, 4.6.2. Data update and change management, 4.6.3. Data archiving and deletion 4.7.3. Data lineage and traceability, 5.1.1. Workflow automation, 5.1.2. Decision automation 5.2.3. External analytics integration, 5.4.1. Operational processes 6.2.2. Industry collaboration capability, 6.2.4. Technology ecosystem participation 6.3.1. Analytics-driven competitive positioning, 6.3.2. Industry analytics benchmarking 6.3.3. Market analytics responsiveness, 6.4.3. Analytics and AI regulatory readiness
Domain-Conditional	2.4. Analytics external expertise, 5.4.3. Customer-facing processes, 6.1. Customer analytics engagement 6.2.1. Government support utilization, 6.2.3. Research collaboration capability
Industry-Specific	6.4.1. Regulatory and legal compliance, 6.4.2. Industry standards compliance

context-dependent criteria can be adapted while the model structure remains the same. Universal criteria are evaluated using common measurement scales, while context-adaptive, domain-conditional, and industry-specific criteria can be measured through context-specific evaluation functions while preserving their position in the hierarchy. Furthermore, LSP allows the relative importance of assessment criteria to be adjusted through weighting and provides a wide range of aggregation functions that capture different logical relationships among criteria. In this way, heterogeneous context-specific measurements can be aggregated into a single explainable maturity score, enabling organizations from different industries and operating environments to be evaluated with a common conceptual model while preserving contextual relevance and supporting meaningful comparisons.

D. Possibilities and Limitations of AI as Cognitive Amplifier

Responding to RQ2, the findings of this research support the view that AI can function as a cognitive amplifier in knowledge-intensive tasks. The observed level of cognitive amplification can be characterized as moderate. AI substantially facilitates tasks such as literature synthesis and information processing, while continuing to require extensive human oversight. Researchers are required to verify AI-generated outputs, resolve inconsistencies and ensure methodological soundness. Consequently, AI reduced the cognitive effort associated with routine analytical activities while simultaneously increasing the importance of expert judgment during evaluation and decision-making. The AI-generated hierarchies consistently captured the core capability areas established in the BDA maturity literature, demonstrating that current GenAI tools are effective at synthesizing existing knowledge. At the same time, the results indicate that cognitive amplification arises from the interaction between AI capabilities and human expertise. The AI-generated criteria remained largely derivative, indicating that the main contribution of current GenAI lies in knowledge synthesis rather than conceptual innovation. The tools reorganized existing concepts and proposed terminology more closely aligned with recent literature, rather than introducing new ones. However, the refinement stage produced more coherent and internally consistent hierarchies than the initial AI-generated outputs, demonstrating that cognitive amplification is achieved through iterative human–AI interaction. This finding also supports the importance of prompt engineering as a mechanism for guiding AI reasoning. Consistent with previous research, the quality of GenAI output was influenced by contextual information provided and iterative feedback, making effective prompting an interactive dialogue in which human expertise progressively guides the model toward conceptually stronger solutions. The effectiveness of this human–AI interaction depends critically on the quality of human input. This observation supports the argument of [35] that AI becomes a cognitive amplifier only when embedded within workflows that require active user engagement. Consequently, domain expertise remains essential. Researchers must be able to recognize hallucinations and misleading recommendations, provide targeted instructions, and critically evaluate AI-generated outputs. In this research, we rarely encountered hallucinations. Requiring the tools to justify each recommendation with supporting literature substantially reduced the risk of fabricated outputs. In most cases, the recommendations were accompanied by references to relevant publications, many of which we were already familiar with, facilitating the verification process. When recommendations concerned parts of the maturity model that intentionally differed from the existing literature, the tools generally acknowledged the suggestions as speculative. For example, recommendations concerning the Architecture subdomain recognized that its conceptualization differed from common structures reported in the literature and noted that the proposed design might represent a deliberate modeling choice. However, occasional errors still occurred. When evaluating the overall criteria hierarchy, the GenAI tools sometimes failed to distinguish between hierarchical levels, resulting in incorrect counts of subdomains or assessment criteria. Ultimately, the quality of the outcome depends on both the capabilities of the AI system and the user's ability to critically assess the value of its recommendations.

Several observations from the refinement process further illustrate these findings. Both AI tools occasionally produced inconsistent recommendations. For example, criteria proposed for removal in one iteration were later recommended for reinstatement after the hierarchy had been modified. Similarly, comparable design decisions received contradictory evaluations across different iterations. These observations demonstrate that GenAI produces context-dependent rather than stable judgments, adapting its recommendations to the conversational context instead of consistently applying methodological principles. This behavior is reinforced by the tendency of AI tools to be overly agreeable, readily validating user suggestions even when they are methodologically questionable [12]. Consequently, AI-generated outputs should not be interpreted as objective or authoritative judgments but as recommendations requiring continuous critical assessment by domain experts. AI should be viewed as a collaborative reasoning partner that complements expert judgment. Another important implication concerns cognitive offloading. Throughout the model development and validation process, GenAI substantially reduced the effort required to review the literature, synthesize existing knowledge, organize concepts, and identify potential improvements. This reduction in analytical workload shifted the primary focus from information gathering to activities, such as critical evaluation, reasoning, validation, and design decision-making. In this sense, GenAI supported cognitive offloading for routine analytical tasks while simultaneously increasing the importance of human judgment. While AI accelerated knowledge synthesis and generated recommendations, the final responsibility for selecting appropriate recommendations and ensuring the theoretical and practical validity of the maturity model remained with the researchers, as we expected at the beginning of this research.

However, the application of GenAI in this research differs from many previously reported use cases. Rather than generating text, software code, or isolated recommendations, we used GenAI to evaluate and iteratively refine a conceptual artifact whose components are highly interdependent. Although the quality of individual recommendations was important, preserving conceptual consistency across the hierarchy and maintaining coherence throughout the refinement process proved even more critical. Many recommendations were well justified, but contradictory recommendations occasionally emerged across iterations, requiring expert assessment of their consistency with previous design decisions and the overall conceptual structure. Furthermore, unlike tasks with objectively correct solutions, such as software development, maturity model design permits multiple valid options. As a result, human judgment remained essential for determining which design best satisfied the theoretical foundations and intended assessment objectives of the model. In this regard, the interaction more resembled a peer-review process than conventional content generation: GenAI critically evaluated an existing conceptual artifact and justified its recommendations using literature, while the researchers assessed, accepted, modified, or rejected those recommendations.

Overall, the findings suggest that the greatest value of GenAI is not in generating more ideas but in stimulating better human reasoning. Although AI can rapidly synthesize literature, analyze conceptual relationships, and recommend improvements, it cannot determine which recommendations are methodologically rigorous or contextually suitable. These decisions remain the responsibility of human researchers. As argued by [36], trust in AI-assisted research should be grounded in transparent reasoning and methodological justification rather than empirical performance alone. Accordingly, the most effective approach is a human-in-the-loop process that combines AI-assisted literature synthesis and human expertise and judgement. Under these conditions, GenAI functions as a cognitive amplifier that accelerates knowledge synthesis, facilitates human understanding and supports more informed research decisions while preserving human responsibility. These findings answer RQ2 by demonstrating that the effectiveness of GenAI as a cognitive amplifier depends on iterative human–AI collaboration rather than autonomous AI capabilities.

VI. Conslusion

In this paper, we evaluated assessment criteria for a Big Data Analytics maturity model and explored the potential of generative artificial intelligence as a cognitive amplifier through this use case. The research produced two outcomes: a refined hierarchy of assessment criteria that provides the foundation for subsequent BDA maturity model development and empirical evidence on the roles of human expertise and GenAI in knowledge-intensive design activities.

In response to RQ1, the findings demonstrate that GenAI can provide meaningful evaluations of maturity model assessment criteria that substantially overlap with expert judgment. Both human experts and GenAI independently identified several conceptual weaknesses in the initial assessment criteria. However, their contributions were complementary rather than identical. Expert validation addressed practical relevance, clarity, and operationalization of the assessment criteria, while GenAI proved effective in identifying conceptual overlaps and incorporating emerging BDA capabilities from literature. At the same time, AI-generated recommendations represented mostly reorganizations or extensions of existing concepts rather than novel contributions, confirming the role of GenAI as a tool for knowledge synthesis rather than conceptual innovation.

In response to RQ2, the findings show that the effectiveness of GenAI as a cognitive amplifier depends fundamentally on the interaction between AI capabilities and human expertise. GenAI proved valuable for facilitating literature synthesis, supporting the iterative refinement of assessment criteria, therefore reducing the cognitive effort required for routine analytical tasks. At the same time, we identified important limitations, such as context-dependence and occasionally inconsistent recommendations, as well as a tendency to produce overly agreeable responses. These findings suggest that effective use of GenAI requires domain expertise and continuous human oversight, where AI supports reasoning while humans retain responsibility for methodological rigor and decision-making. Cognitive amplification should be understood as an outcome of effective human–AI collaboration rather than an inherent property of the AI system itself.

Beyond answering the research questions, we make two broader contributions. First, we propose a refined and evaluated assessment criteria for the subsequent development and empirical validation of the Big Data Analytics maturity model. Second, we contribute to the research on AI-assisted approaches by providing empirical evidence of the conditions under which GenAI can function as a cognitive amplifier in knowledge-intensive design activities. However, this research has several limitations. The AI-assisted evaluation was conducted using two GenAI tools and therefore reflects their capabilities at the time of the research. As foundation models continue to evolve rapidly, their capabilities and overall performance are likely to improve. In addition, the expert validation involved a relatively small number of participants, which may limit the generalizability of the findings.

A. Future Work

Future research will focus on validating the refined assessment hierarchy across a broader range of organizations and industries, further investigating the context dependence of assessment criteria and the proposed context classification, and developing measurement scales for the refined criteria. In subsequent research phase, the LSP assessment instrument will be implemented and validated, including the definition of criterion weights, aggregation functions, and explainability mechanisms. In addition, future research could examine how newer generations of GenAI models influence maturity model development and other knowledge-intensive design activities.

APPENDIX

A. Survey Instrument

The following questions were used to collect information about the respondents, their professional roles, experience, organizational context, and expertise in Big Data Analytics.

Q1: What is your current position/role?

Technical role (e.g., data scientist, data engineer, ML/AI engineer)
Manager (business / non-technical)
Manager (analytics / IT-related)
Consultant / external expert
Researcher / academic
Other:

Q2: How many years of professional experience do you have?

Less than 3 years
3–5 years
6–10 years
More than 10 years

Q3: How many years of experience do you have in Big Data Analytics or related fields?

None
Less than 1 year
1–2 years
3–5 years
6–10 years
More than 10 years

Q4: In which areas do you have professional experience?

Data engineering / data infrastructure
Data analytics / business intelligence
Machine learning / artificial intelligence
Data governance / data management
Business analysis / decision-making
IT systems / software development
Other:

Q5: In which industry does your organization primarily operate?

Information technology / software
Finance / banking / insurance
Healthcare / pharmaceuticals
Telecommunications
Manufacturing / production
Retail / e-commerce
Public sector / government
Education / research
Other:

Q6: What is the approximate size of your organization?

Fewer than 50 employees
50–249 employees
250–499 employees
500–999 employees
1000 or more employees

Q7: To what extent are you involved in analytics-related activities in your organization?

1 – Not involved
2 – Slightly involved
3 – Moderately involved
4 – Highly involved
5 – Directly responsible / leading analytics activities

Q8: How would you rate your level of expertise in Big Data Analytics?

1 – No expertise
2 – Basic
3 – Intermediate
4 – Advanced
5 – Expert

B. Prompt: Generating a Hierarchical Criteria Tree for the Big Data Analytics Maturity Model

You are acting as a research assistant developing a Big Data Analytics (BDA) maturity model. Your task is to synthesize existing academic knowledge and propose a structured hierarchy of assessment criteria. Base your analysis on literature concerning Big Data maturity models, Big Data Analytics maturity models, data maturity models, analytics capability and maturity models, and existing surveys and systematic literature reviews on BDA maturity.

Objective
Construct a three-level hierarchical criteria tree suitable for assessing organizational Big Data Analytics maturity. The hierarchy should consist of:

Level 1: maturity domains,
Level 2: maturity subdomains,
Level 3: assessment criteria, representing specific organizational capabilities that can later be operationalized as measurable indicators.

Requirements
The hierarchy should represent concepts reported across the literature. Merge synonymous concepts appearing under different names in different maturity models. Each Level 3 criterion should represent one distinct capability and should not combine multiple concepts. Do not define measurement scales or maturity levels.
For every criterion, provide:

criterion name,
a short definition of one or two sentences,
examples of terminology used in different papers, where applicable.

Output format
Present the hierarchy using the following structure:

Level 1 Domain → Level 2 Subdomain → Level 3 Criterion + Definition

C. Prompt: Evaluating the AI-Proposed Criteria Hierarchy

Evaluate the proposed hierarchy as if you were a panel of experts in Big Data Analytics, and maturity model development.

Identify:

missing criteria,
redundant criteria,
inconsistent granularity,
misplaced criteria,
overlapping domains,
criteria that should be merged or split,
domains that are too broad or too narrow.

Justify every suggested modification with reference to the maturity model literature.

D. Prompt: AI-Generated Critical Evaluation of the Big Data Analytics Maturity Model Criteria Tree

You are acting as expert in Big Data Analytics refining a Big Data Analytics (BDA) maturity model. I will provide a hierarchical criteria tree of assessment criteria based on a literature review of Big Data and Big Data Analytics maturity models. The hierarchy consists of three levels: Level 1 (domains), Level 2 (subdomains), and Level 3 (assessment criteria).

Your task is to critically evaluate the conceptual structure of this hierarchy using the existing scientific literature on Big Data, Big Data Analytics, analytics capability, analytics maturity, and related organizational capability frameworks.

Review the hierarchy according to the following principles:
Completeness: Identify important capabilities that are supported in literature but missing from hierarchy. Explain why they should be considered.
Redundancy: Identify duplicate or overlapping criteria. Recommend whether they should be merged, separated, or removed.
Mutual exclusiveness: Identify criteria that overlap conceptually across multiple subdomains. Suggest clearer boundaries between subdomains.
Consistency of abstraction: Verify that all Level 1 domains represent concepts of similar scope. Verify that Level 2 subdomains have comparable granularity. Verify that Level 3 criteria represent individual organizational capabilities rather than combinations of multiple concepts.
Logical placement: Evaluate whether each criterion belongs to its assigned subdomain and domain. Suggest alternative placements where appropriate.
Terminology: Identify inconsistent naming. Recommend terminology that is commonly used in academic literature. Point out synonymous concepts that could be unified.
Suitability for maturity assessment: Evaluate whether each Level 3 criterion represents a capability that could later be operationalized into measurable assessment items. Identify criteria that are too broad, too vague, or difficult to measure.

Requirements
For every recommendation:

describe the issue,
propose modification,
justify the recommendation using published literature.

Important instructions Critically evaluate the model rather than attempting to confirm it. Distinguish between well-supported recommendations and speculative suggestions. Preserve the focus on assessing organizational BDA maturity rather than general digital maturity.

E. Comparison of the Initial and Refined Assessment Criteria for the Proposed BDAMM

	Initial		Refined	
Domain	**Subdomains**	**Criteria**	**Subdomains**	**Criteria**
Organization	5	25	5	22
Human capital	4	18	4	16
Technology	4	19	4	20
Data	7	28	7	26
Processes	6	20	5	17
Environment	3	11	4	14
Total	**29**	**121**	**29**	**115**

Acknowledgment

This research was supported by the University of Belgrade – Faculty of Organizational Sciences and, in part, by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia through institutional funding (grant number: 200151).

Generative AI Disclosure Statement

Generative AI tools were used as part of the research methodology to generate and evaluate BDAMM assessment criteria, consistent with the objectives of this paper. All research ideas, methodological decisions, analyses, interpretations of the results, and conclusions are the work of the authors, who critically evaluated all AI-generated outputs and take full responsibility for the final content of the publication.

Supplementary Materials

Additional analyses supporting the findings presented in this paper, including detailed results and extended versions of the tables, are available in the online repository (link: https://github.com/m-djukic/BDAMM).

References

[1] J.-M. Sahut, D. Schweizer, and M. Peris-Ortiz, "Technological forecasting and social change introduction to the VSI: Technological innovations to ensure confidence in the digital world," *Technological Forecasting and Social Change*, vol. 179, p. 121680, 2022. doi: https://doi.org/10.1016/j.techfore.2022.121680

[2] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. J. Patil, and D. Barton, "Big data: The management revolution," *Harvard Business Review*, vol. 90, no. 10, pp. 61–67, 2012.

[3] S. F. Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How "big data" can make big impact: Findings from a systematic review and a longitudinal case study," *International Journal of Production Economics*, vol. 165, pp. 234–246, 2015. doi: https://doi.org/10.1016/j.ijpe.2014.12.031

[4] M. Gupta and J. F. George, "Toward the development of a big data analytics capability," *Information & Management*, vol. 53, no. 8, pp. 1049–1064, 2016. doi: https://doi.org/10.1016/j.im.2016.07.004

[5] D. Zhang, S. L. Pan, J. Yu, and W. Liu, "Orchestrating big data analytics capability for sustainability: A study of air pollution management in china," *Information & Management*, vol. 59, no. 5, p. 103231, 2022. doi: https://doi.org/10.1016/j.im.2019.103231

[6] S. F. Wamba, A. Gunasekaran, S. Akter, S. J.-f. Ren, R. Dubey, and S. J. Childe, "Big data analytics and firm performance: Effects of dynamic capabilities," *Journal of Business Research*, vol. 70, pp. 356–365, 2017. doi: https://doi.org/10.1016/j.jbusres.2016.08.009

[7] C. Adrian, R. Abdullah, R. Atan, and Y. Y. Jusoh, "Towards developing strategic assessment model for big data implementation: A systematic literature review," *International Journal of Advanced Soft Computing and Its Applications*, vol. 8, no. 3, pp. 173–192, 2016.

[8] M. Rosemann and T. de Bruin, "Towards a business process management maturity model," in *Proceedings of the European Conference on Information Systems*, Regensburg, Germany, 2005.

[9] J. Becker, R. Knackstedt, and J. Pöppelbuß, "Developing maturity models for IT management: A procedure model and its application," *Business & Information Systems Engineering*, vol. 1, no. 3, pp. 213–222, 2009. doi: https://doi.org/10.1007/s12599-009-0044-5

[10] M. Đukić and I. Luković, "Exploring big data maturity models: Findings from a systematic literature review," in *European Conference on Advances in Databases and Information Systems*. Cham: Springer, 2026, pp. 264–279.

[11] M. Đukić and I. Luković, "Developing initial assessment criteria for a big data analytics maturity model," in *Proceedings of the IEEE International Conference on Evolving and Adaptive Intelligent Systems*, 2026, in press.

[12] T. An, "AI as cognitive amplifier: Rethinking human judgment in the age of generative AI," *arXiv preprint arXiv:2512.10961*, 2025. doi: https://doi.org/10.48550/arXiv.2512.10961

[13] Z. A. Al-Sai, M. H. Husin, S. M. Syed-Mohamad, R. Abdullah, R. A. Zitar, L. Abualigah, and A. H. Gandomi, "Big data maturity assessment models: A systematic literature review," *Big Data and Cognitive Computing*, vol. 7, no. 1, p. 2, 2023. doi: https://doi.org/10.3390/bdcc7010002

[14] B. Langer, "Understanding data & analytics maturity: A systematic review of maturity model composition," *Schmalenbach Journal of Business Research*, vol. 77, no. 2, pp. 205–227, 2025. doi: https://doi.org/10.1007/s41471-024-00205-2

[15] R. Marohn and Y. Li, "Data analytics capability maturity models for small and medium enterprises: A systematic literature review," 2024, publication details not provided.

[16] R. B. Sadiq, N. Safie, A. H. Abd Rahman, and S. Goudarzi, "Artificial intelligence maturity model: A systematic literature review," *PeerJ Computer Science*, vol. 7, p. e661, 2021. doi: https://doi.org/10.7717/peerj-cs.661

[17] G. Reichl and R. Gruenbichler, "Maturity models for the use of artificial intelligence in enterprises: A literature review," in *Proceedings of the 19th International Scientific Conference on Industrial Systems*, Nov. 2023, pp. 486–502.

[18] K. Ariansyah, A. B. Setiawan, A. Hikmaturokhman, A. Ardison, and D. Walujo, "Big data readiness in the public sector: An assessment model and insights from indonesian local governments," *Journal of Science and Technology Policy Management*, vol. 16, no. 2, pp. 252–278, 2025. doi: https://doi.org/10.1108/JSTPM-01-2023-0010

[19] N. Areerakulkan and W. A. Pongpech, "A dempster–shafer big data readiness assessment model," in *Proceedings of the International Conference on Enterprise Information Systems*, vol. 2, 2021, pp. 581–585.

[20] M. Comuzzi and A. Patel, "How organisations leverage big data: A maturity model," *Industrial Management & Data Systems*, vol. 116, no. 8, pp. 1468–1492, 2016. doi: https://doi.org/10.1108/IMDS-12-2015-0495

[21] A. Corallo, A. M. Crespino, V. Del Vecchio, M. Gervasi, M. Lazoi, and M. Marra, "Evaluating maturity level of big data management and analytics in industrial companies," *Technological Forecasting and Social Change*, vol. 196, p. 122826, 2023. doi: https://doi.org/10.1016/j.techfore.2023.122826

[22] C. Dremel, S. Overhage, S. Schlauderer, and J. Wulf, "Towards a capability model for big data analytics," in *Wirtschaftsinformatik 2017 Proceedings*, 2017, pp. 1141–1155.

[23] B. Farah, "A value-based big data maturity model," *Journal of Management Policy and Practice*, vol. 18, no. 1, pp. 11–18, 2017.

[24] M. Farzaneh, F. Mozaffari, S. P. Ameli, M. Karami, A. Mohamadian, and E. Arianyan, "Designing an organizational readiness framework for big data adoption," in *Proceedings of the 9th International Symposium on Telecommunications*, 2018, pp. 387–391.

[25] R. Fornasiero, L. Kiebler, M. Falsafi, and S. Sardesai, "Proposing a maturity model for assessing artificial intelligence and big data in the process industry," *International Journal of Production Research*, vol. 63, no. 4, pp. 1235–1255, 2025. doi: https://doi.org/10.1080/00207543.2024.2372840

[26] I. Hausladen and M. Schosser, "Towards a maturity model for big data analytics in airline network planning," *Journal of Air Transport Management*, vol. 82, p. 101721, 2020. doi: https://doi.org/10.1016/j.jairtraman.2019.101721

[27] S. Limpeeticharoenchot, N. Cooharojananone, T. Chavarnakul, N. Charoenruk, and K. Atchariyachanvanich, "Adaptive big data maturity model using latent class analysis for small and medium businesses in thailand," *Expert Systems with Applications*, vol. 206, p. 117965, 2022. doi: https://doi.org/10.1016/j.eswa.2022.117965

[28] S. Mouhib, H. Anoun, M. Ridouani, and L. Hassouni, "Global big data maturity model and its corresponding assessment framework results," *IAENG International Journal of Applied Mathematics*, vol. 53, no. 1, 2023.

[29] C. M. Olszak and M. Mach-Król, "A conceptual framework for assessing an organization's readiness to adopt big data," *Sustainability*, vol. 10, no. 10, p. 3734, 2018. doi: https://doi.org/10.3390/su10103734

[30] H. Sulaiman, Z. C. Cob, and N. A. Ali, "Big data maturity model for malaysian zakat institutions to embark on big data initiatives," in *Proceedings of the 4th International Confer-*

ence on Software Engineering and Computer Systems, 2015. doi: https://doi.org/10.1109/ICSECS.2015.7333084 pp. 61–66.

[31] M. Willetts and A. S. Atkins, "Big data analytics maturity model for SMEs," *International Journal of Information Technology and Computer Science*, vol. 16, no. 2, pp. 1–15, 2024. doi: https://doi.org/10.5815/ijitcs.2024.02.01

[32] J. Dujmović, *Soft Computing Evaluation Logic*. John Wiley & Sons, 2018. ISBN 9781119258561

[33] J. Dujmović, *Graded Logic*. Springer, 2025. ISBN 9783031882374

[34] European Union, "Regulation (EU) 2016/679 of the european parliament and of the council: General data protection regulation (GDPR)," Official Journal of the European Union, 2016.

[35] F. O. Alubthane, "Amplifier or substitute? a systematic review of generative AI's impact on higher-order cognitive skills among university students," *Frontiers in Psychology*, vol. 17, p. 1863931, 2026.

[36] M. Kwiatkowska and X. Zhang, "When to trust AI: Advances and challenges for certification of neural networks," in *Proceedings of the 18th Conference on Computer Science and Intelligence Systems*. IEEE, Sep. 2023, pp. 25–37.