e-ISSN: 2735-0703

Available online at
https://mijuitm.com.my

Mathematical Sciences and Informatics Journal 6(2) 2025, 114 – 131.

Mathematical Sciences and Informatics Journal

Responsible Procurement of AI Applications: A Risk-Based Framework for Malaysian Government Agencies

David Lau Keat[1*], Ganthan Narayana Samy[2], Fiza Abdul Rahim[2], Mahiswaran Selvanathan[3], Nurazean Maarop[2], Mugilraj Radha Krishnan[2], Sundresan Perumal[4]

[1]*Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia*
[2]*Faculty of Artificial Intelligence, Universiti, Teknologi Malaysia*
[3]*Faculty of Social Sciences and Humanity, Universiti Teknologi Malaysia*
[4]*Faculty of Science and Technology, University, Sains Islam Malaysia*

ARTICLE INFO

Article history:
Received 30 August 2025
Revised 14 September 2025
Accepted 30 September 2025
Online first
Published 31 October 2025

Keywords:
Artificial Intelligence
Risk Management
Public Sector
Large Language Model
Autonomous System

DOI:
10.24191/mij.v6i2.9172

ABSTRACT

The rapid advancement of artificial intelligence (AI), driven by innovation from technology firms and academia, has expanded its capabilities and accelerated its adoption across sectors. The integration of AI into the public sector is inevitable, as it promises greater efficiency, improved decision-making, and enhanced service delivery. However, these benefits come with new and complex risks particularly due to the emergence of generative AI and autonomous agents capable of independent decision-making. Public agencies are therefore responsible for ensuring that deployed AI systems are not only effective but also secure, ethical, and cost-efficient. Current information security frameworks, such as ISO/IEC 27001:2022, remain inadequate for addressing risks associated with large language models and agentic AI. This study proposes a risk-based framework tailored for responsible procurement of generative AI solutions within Malaysian government agencies. Employing a qualitative methodology that integrates semi-structured interviews with AI practitioners from both public and private sectors, alongside qualitative document analysis, the research identifies key risk considerations and governance requirements. The resulting framework provides a structured approach to managing AI procurement risks and aligning them with the principles of responsible AI envisioned by the Malaysian government. Future research may focus on automating elements of the framework and integrating emerging risk countermeasures from technical working groups.

[1*] Corresponding author. *E-mail address*: davidkeat@graduate.utm.my

https://doi.org/10.24191/mij.v6i2.9172

1. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has transformed both private and public sectors by enabling data-driven decision-making, automation, and enhanced service delivery. In the public sector, AI is increasingly adopted to improve operational efficiency and citizen engagement. However, the same technological capabilities introduce complex risks related to ethics, accountability, cybersecurity, and governance. As shown in Fig. 1, AI-related incidents are most frequent in the technology and government sectors, underscoring the urgency for robust governance and oversight mechanisms (Responsible AI Collaborative, 2024).

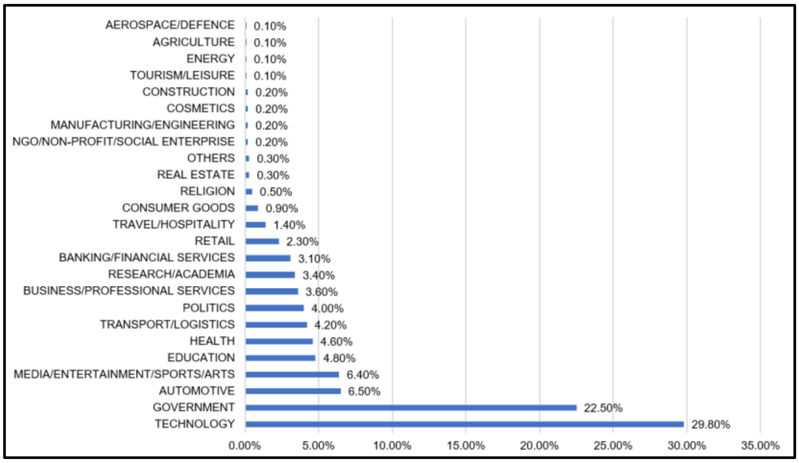


Fig. 1. Percentage of AI related incidents by sector (Responsible AI Collaborative, 2024)

Despite the existence of international standards such as ISO/IEC 27001:2022 and ISO 31000:2018, current risk management frameworks are not fully equipped to address the distinctive risks of large language models (LLMs), generative AI, and autonomous agents (McIntosh et al., 2024). These systems exhibit non-deterministic behaviors, data dependency, and adaptive learning properties that challenge traditional procurement and assurance processes (NIST, 2023) (ISO/IEC, 2022). Consequently, public agencies face difficulties in ensuring that AI systems procured from external vendors align with responsible AI principles. This includes fairness, transparency, and security (IEEE, 2025). In Malaysia, initiatives such as the Public Sector Digitalization Strategic Plan 2021–2025 (JPM, 2020) and the AI Adoption Guideline 2025 (NDD, 2025) emphasize responsible AI use in governance and public administration. However, these documents do not provide a risk-based methodology for integrating responsible AI considerations into procurement activities. Since procurement decisions largely determine how AI is designed, implemented, and monitored, there is a critical need for a framework that embeds risk management throughout the procurement lifecycle.

Accordingly, this study aims to develop a risk-based procurement framework tailored for AI applications within Malaysian government agencies. The specific objectives are to:

(i) identify the considerations and gaps in current procurement and risk management practices related to AI systems,

(ii) incorporate expert and practitioner insights to design a contextual and adaptable framework, and

(iii) provide structured guidance to enable responsible and compliant AI acquisition in the public sector.

https://doi.org/10.24191/mij.v6i2.9172

The remainder of this paper is organized as follows: Section 2 analyzes current challenges and gaps; Section 3 presents the methodology; Section 4 discusses results and the proposed framework; Section 5 provides discussion and implications; and Section 6 concludes with recommendations for future work.

2. CHALLENGES AND GAPS IN AI RISK MANAGEMENT FOR PUBLIC PROCUREMENT

In the context of information security, the severity of risk is correlated with confidentiality, integrity, and availability of information (NIST, 2023). Moreover, the prevalent identification of risks within system security encompasses the documented threats and vulnerabilities (JPM, 2024a). With regards to AI adoption, the extant data remain ambiguous concerning the probability of a risk's occurrence, particularly as it is a rapidly progressing domain that may yield divergent operational implications due to emergent advancements within its implementation (Turri & Dzombak, 2023). For instance, the process of fine-tuning a language model is known to influence the intrinsic alignment of a pre-trained model (Qi et al., 2023). This issue is aggravated when leading-edge AI technologies are entrusted with tasks that demand high levels of automation (Ferrara, 2024). Such scenarios are increasingly prevalent with the advent of agentic AI that not only leverages AI models but also integrates with other agents, memory, and tools.

Table 1 describes the difference between generative AI and agentic AI (Masterman et al., 2024; Ooi et al., 2025). While certain research may theorize that the likelihood of an attack is contingent upon motivation and the degree of complexity (Javaid et al., 2012), such a premise fails to address the risks that arise from non-deliberate attacks as well as the challenge of accurately assessing the complexity level for all varieties of attacks, let alone those techniques that remain unidentified (Bountakas et al., 2023). In high-stakes applications, the risks associated with the utilization of AI are linked to the degree of automation, as well as the consequences of its decisions in maintaining the safety of its intended subject (Kilian et al., 2023). Hence, the existing method of determining the risk level based on mapping of impact and probability of risk occurrence require re-examination.

Table 1. Differences between generative AI and agentic AI (adapted from Masterman et al., 2024; Ooi et al., 2025)

Aspect	Generative AI	Agentic AI
Definition	Systems that create new content such as text, images, code, or audio based on learned patterns from large datasets.	Systems that can perceive, reason, and act autonomously toward achieving specific goals, often coordinating with other agents or humans.
Core Functionality	Content generation and synthesis (e.g., summarization, translation, image creation).	Goal-oriented decision-making, planning, and execution of actions in dynamic environments.
Level of Autonomy	Limited autonomy — operates mainly on user prompts or predefined workflows.	High autonomy — capable of self-initiated actions, adaptation, and interaction without continuous human input.
Learning Mechanism	Primarily pre-trained and fine-tuned on large datasets (LLMs, diffusion models).	Combines cognitive architectures, reinforcement learning, and multi-agent coordination mechanisms.
Interaction Mode	One-directional (input → output); reactive to user prompts.	Iterative and interactive — agents communicate, reason, and take sequential decisions to reach objectives.
Example Use Cases	Text or image generation, chatbots, summarization tools, code completion.	Autonomous research assistants, automated procurement agents, multi-agent systems managing workflows.
Risk Characteristics	Bias, hallucination, intellectual property issues, data privacy, and misinformation.	Misaligned goals, cascading errors, emergent behaviours, adversarial manipulation, and loss of human oversight.

https://doi.org/10.24191/mij.v6i2.9172

Aspect	Generative AI	Agentic AI
Governance Challenges	Ensuring accuracy, transparency, and ethical content generation.	Ensuring accountability, explainability of autonomous decisions, and safe delegation of authority.
Relevant Standards/Frameworks	ISO/IEC 42001, NIST AI RMF, OECD AI Principles.	OWASP Agentic AI Threats (2025), MITRE ATLAS, ISO/IEC 22989 (automation levels).
Procurement Implications	Focus on content reliability, data licensing, and model performance benchmarks.	Requires risk assessments on autonomy levels, safety-critical use, inter-agent dependencies, and liability clauses.

Despite publication of related standards, a study that mapped the provisions in NIST NSF 2.0, COBIT 2019, ISO 27001:2022 and ISO 42001:2023 to risks of LLM indicated significant gaps in risk management (McIntosh et al., 2024). Moreover, existing frameworks such as Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege (STRIDE) (Shostack, 2014), Process for Attack Simulation and Threat Analysis (PASTA) (UcedaVelez & Morana, 2015), Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of Information, Unawareness, and Non-compliance (LINDDUN) (Deng et al., 2011), Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) (Nagar, 2017), Trike (Saitta et al., 2005) and Visual, Agile, and Simple Threat (VAST) (Bernsmed et al., 2022), while useful in many areas, leave significant gaps for Agentic AI. They do not adequately address the unique challenges arising from the autonomy, learning, and interactive nature of these systems (Chan et al., 2025).

For example, existing frameworks struggle to model the unpredictable actions of autonomous agents, which arise from their independent decision-making (Portugal et al., 2024). Specifically, these frameworks often do not adequately cover threats related to an agent's goals becoming misaligned with the intended purpose. Additionally, truncated, hallucinated, or failed responses can occur in a multi-agent system due to context overflow or deliberate attacks (Deng et al., 2025; Shavit et al., 2023). Likewise, although IEEE (2025) illustrated use cases for facial recognition, automated image processing, chatbot, and automated decision systems, there are risks specific to the use of AI agents such as that need to be considered. In this regard, a comprehensive risk taxonomy in AI such as those developed by Slattery et al. (2024) and Zeng et al. (2024) can be considered in the preparation of procurement specifications and contractual terms. Table 2 gives a comparative analysis of existing frameworks.

Notably, there are already guidelines for the procurement of cloud infrastructure for the government. However, the restrictions stipulated in the cloud computing policy on using non-panel cloud providers may also pose limitations when specialized AI capabilities are available only from niche vendors outside the Cloud Framework Agreement (CFA) (MOF, 2022). This is evident given that there are AI providers that serve different scopes in the technology layers for agentic AI systems (OWASP, 2025a). Furthermore, for AI integrated into larger digital transformation projects, the procedural requirements for designating MSPs as Nominated Sub-Contractors introduce additional contractual complexity. While it is commendable that guidelines have been provided for government agencies in terms of AI adoption (NDD, 2025) as well as AI governance and ethics (MOSTI, 2024), there is no consideration for an AI maturity model (AIMM) in an organization. Organizational readiness assessment is crucial as it influences procurement decisions.

Table 2. Comparative analysis of existing risk-based frameworks

Framework	Methodology	Strengths	Limitations	Applicable Environment
IEEE 3119:2025 (IEEE, 2025)	Procurement-oriented AI risk and governance standard	Focuses on AI procurement contracts, vendor governance, and transparency in acquisitions	Limited in scope and adoption	Public sector and enterprise AI procurement, vendor selection, contract compliance

https://doi.org/10.24191/mij.v6i2.9172

Framework	Methodology	Strengths	Limitations	Applicable Environment
ISO/IEC 42001:2023 (ISO/IEC, 2023)	AI Management System standard	Lifecycle-based AI governance, supports certification, aligns with trust and compliance requirements	High-level, potentially bureaucratic, not detailed in technical attack modelling	Organizations developing/deploying AI, especially regulated industries
NIST AI RMF (NIST, 2023)	Voluntary AI risk management framework with Govern-Map-Measure-Manage functions	Flexible, lifecycle-oriented, emphasizes trustworthiness, transparency, and resilience	Non-prescriptive, requires tailoring, adoption consistency varies	Organizations adopting AI with risk/governance overlay, adaptable to various industries
STRIDE (Ouaissa & Ouaissa, 2025)	Threat taxonomy	Simple, intuitive, easily applied in system design for identifying AI system threats	Security-only, no prioritization or privacy focus	AI application/system architecture design, especially in data flow and Application Programming Interface (API) threat analysis
PASTA (Pape & Mansour, 2024)	Attack simulation and business-driven threat modelling	Risk-centric, attacker-focused, links AI business goals and technical threats	Heavyweight, resource-intensive, requires expertise	Complex AI deployments, enterprise-scale systems, critical AI applications
LINDDUN (Deng et al., 2011)	Privacy threat modelling framework	Focused on privacy risks in AI, aligns with data protection regulations	Does not address broader security threats, prioritization challenges	AI models handling personal/sensitive data
OCTAVE (Awad et al., 2023; Nagar, 2017)	Organizational-level risk assessment of critical assets and threats	Enterprise-wide perspective, aligns AI risks with business processes	Less technical, not focused on AI-specific attack vectors	Organizational AI risk assessments, policy, and governance alignment
Trike (Masterman et al., 2024; Ooi et al., 2025)	Risk-based threat modelling with actor-asset-action approach	Detailed risk prioritization, supports linking threats to controls	Complex, less widely adopted, tooling limitations	Detailed AI system security analysis, where linking risks to mitigations is key
VAST (Bernsmed et al., 2022)	Agile, visual threat modelling framework	Scales well in Development and Operations (DevOps), lightweight, integrates with agile AI development cycles	Less detail, smaller ecosystem, lighter prioritization	Agile AI/ML development, rapid deployment pipelines, Development, Security and Operations (DevSecOps) environments

Table 3 provides a comparative overview of four major AIMM tailored to the public sector. Each model reflects different approaches to capturing how public-sector organizations grow in AI capability—from linear stage progressions (IBM, CNA) to multi-dimensional assessments (AIMM, RAI-MM). “Levels Defined” shows the stages or dimensions by which maturity is differentiated; “Method Used” summarizes how each model’s structure was derived.

Table 3. AI maturity models for public sector

AI Maturity Model	Levels Defined	Method Used
Desouza (2021)	(i) Ad Hoc (ii) Experimentation (iii) Planning & Deployment (iv) Scaling & Learning (v) Enterprise-Wide Transformation	Synthesis of academic and grey literature, plus practitioner consultations and iterative feedback on pilots to refine levels and their characteristics.
Dreyling et al. (2024)	Maturity assessed across eight dimensions without linear stages:	Design Science methodology: thematic review of existing AIMMs, coding of literature via CAQDAS, iterative expert

https://doi.org/10.24191/mij.v6i2.9172

AI Maturity Model	Levels Defined	Method Used
	Strategy, Governance, Ethics, Procurement models, Legal compliance	consultations (three feedback rounds), and questionnaire-based validation.
Willems (2025)	Maturity assessed across five dimensions: Strategy, Culture & Competences, Governance & Processes, Data & Information, Technology & Tooling	Three-round Delphi study with a diverse expert panel (academia, consultancy, public-sector practitioners), followed by case-study and expert-session validation.
CAN (2025)	Four stages for each subdomain/topic: (vi) Developing (vii) Performed (viii) Established (ix) Optimized	Comprehensive content analysis of 39 policy/strategy documents, classification into domains/subdomains/topics, internal reviews, and refinement via Small and Medium Enterprise (SME) feedback.

Based on the maturity models expounded in Table 3, an agency's maturity stage should directly inform what it procures in five key areas: technology, data, services, talent, and governance mechanisms as summarized in Table 4. Currently, Malaysian government agencies adhere to the major processes in procurement involving planning, implementation, evaluation, monitoring and project implementation as well as contract management and administration (MOF, 2013) without consideration for AI maturity model. Relatedly, the 'Procurement of Cloud Computing Services in the Public Sector' provides a structured governance framework that significantly facilitates the adoption of AI solutions within Malaysia's public sector (MOF, 2022). By mandating the use of a CFA and pre-approved panel of Cloud Service Providers (CSPs) and Managed Service Providers (MSPs), the policy ensures that AI-related cloud infrastructure and services—such as AI platforms, machine learning operations (MLOps), and high-performance computational environments—can be procured efficiently and in compliance with national ICT standards. Moreover, the CFA's emphasis on professional services, training, and integration support aligns well with the iterative and adaptive nature of AI system development, enabling government agencies to access both technical expertise and scalable cloud resources required for modern AI workloads.

Table 4. Procurement decision in different levels of AI maturity

Type of Acquisition	Low / Ad Hoc	Moderate / Experimentation	High / Transformative
Technology Acquisition	Foundational IT (cloud compute, storage, data platforms), low-risk pilot implementation, and small proof-of-concept exercises	Specialized tools & shareable platforms; modular, interoperable systems, performance-based RFP clauses, proof-of-concept contracts	Organization-wide AI platforms, multi-year strategic licenses, advanced AI-as-a-service, strong data-portability & vendor-lock-in protections
Data Partnerships & Sharing	Leverage internal/open data; consultancy to aggregate/clean data, establish basic data-ownership & sharing provisions	Inter-agency/public-private data partnerships, cooperative sharing agreements, use of inter-agency agreements for shared platforms	Open-source or co-maintained datasets/models; participation in government-wide data clouds, API/standards compliance and routine, and cross-agency data exchange
Consulting & Vendor Services	Heavy reliance on external expertise, limited AI implementation scope, vendor-led strategic road-mapping	Internalization of core skills, multi-year analytics contracts, training-curriculum and "platform as a service" procurements	In-house execution with strategic vendor partnerships, niche consulting, contracts designed for competition and continuous innovation
Workforce Development& Training	Basic AI literacy workshops, vendor-provided tutorials; appointment of Chief AI/Data Officer roles	Formal bootcamps, subscription-based learning, defined staffing plans, and rotational programs with academia	Embedded AI teams, advanced degree sponsorships and research seminars, staff augmentation for specialized projects, and contractual knowledge-transfer requirements
Risk, Governance & Ethics	Ensure basic security/privacy compliance; involve CIO/CISO/privacy officers in acquisitions	Establish AI ethics/oversight bodies, include algorithm-audit and safety requirements in RFPs, mandate impact	Governance framework or policy for periodic recertification; fairness, explain ability, and

Type of Acquisition	Low / Ad Hoc	Moderate / Experimentation	High / Transformative
		assessments and ongoing monitoring	appeal-process clauses; independent evaluations

3. METHODOLOGY

To address the research questions, this study used qualitative approaches of semi-structured interviews and qualitative document analysis. Qualitative research methodologies can generate pertinent information about risk management practices including risk prioritization, allocation of resources for risk treatment as well as identifying responsible stakeholders (Moghadasi et al., 2024). It was also chosen to describe processes and lived experiences, especially when those processes and experiences ultimately inform decisions (Carlton, 2014). For semi-structured interview, the total number of respondents was 28, which comprised of 10 females and 18 males. Most of the respondents were from the 40-44 age group (14), followed by the 45-and-above age group (8). The 30-34 and 35-39 age groups were equal in the number of respondents (3). Those who were in 40-44 age group were typically in senior positions within their organizations. Hence, they were more likely to participate in recommendation or decision-making within their organizations compared to younger workforce. However, it should be noted that there is a regulation for job rotation within the public sector, especially for certain sensitive positions and thus, the length of work experience in the organization typically does not correspond to total years of service in the public sector. Additionally, as new technology often originates from other countries, the views of foreign respondents namely: Vietnam, Germany, and the United Arab Emirates (UAE) were also obtained.

No statistical calculation was conducted to determine sample size, as qualitative research emphasizes information power rather than numerical representativeness (Malterud et al., 2016). The sample adequacy was assessed based on five dimensions: study aim, sample specificity, use of established theory, quality of dialogue, and analysis strategy. Given the study’s focused objective, purposive selection of experienced practitioners, and the richness of data obtained, 28 respondents were deemed sufficient to provide information power for meaningful and credible thematic analysis. In this regard, the identification of themes follows the guideline provided by Naeem et al. (2023) which emphasized the characteristics of reciprocal, recognizable, responsive and resourceful.

For document analysis, the government circulars and grey literature by technology working groups selected include: JPM (2024a), JPM (2024b), MOSTI (2024), NDD (2025), NDD (2023a), NDD (2023b), OWASP (2025b), OWASP (2025a) and NIST (2024). The criteria for selections of these nine documents were authenticity, credibility, representativeness and meaning (He et al., 2015). As this study used semi-structured interviews and qualitative document analysis, identification of similarities, differences and gaps were sought as an attempt to triangulate the results. The core idea of triangulation is to use multiple strategies to converge data, methods, or investigators to validate findings and ensure that the results are robust and reliable (Carter, 2014; Vivek et al., 2023). In this study, data source triangulation was used to consolidate the findings from semi-structured interview and qualitative document analysis (Meydan & Akkaş, 2024). The merits of employing triangulation in this study include:

(i) Enhanced credibility and validity: By using multiple methods or data sources, triangulation increases the credibility of the findings and reduces the risk of bias. This is particularly important in qualitative research, where the subjective nature of data collection can sometimes lead to skepticism about the validity of the results (Abdalla et al., 2018; Meydan & Akkaş, 2024).

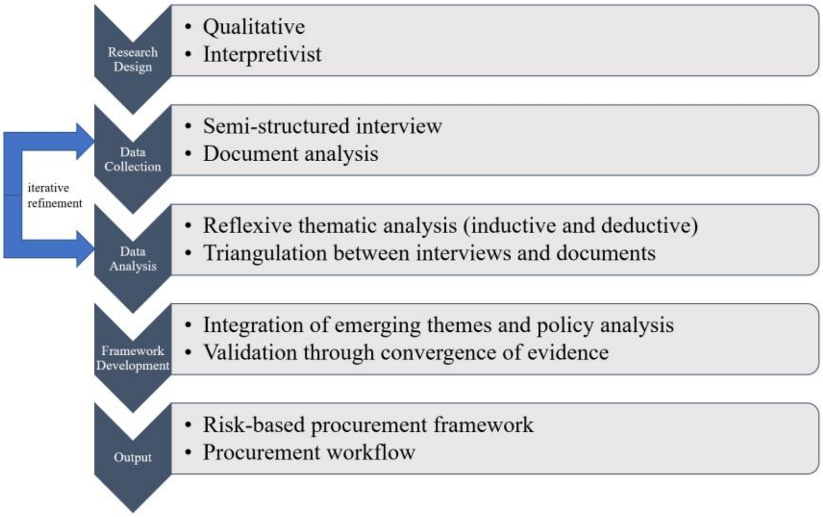
(ii) Comprehensive understanding: Triangulation provides a more holistic understanding of the research phenomenon by capturing different perspectives and experiences. This is especially useful in complex social phenomena where a single method or data source may not be sufficient to capture the full picture (Meydan & Akkaş, 2024; Vivek et al., 2023).

https://doi.org/10.24191/mij.v6i2.9172

(iii) Increased rigor: Triangulation enhances the rigour of qualitative research by providing a systematic approach to data collection and analysis. This is particularly important in addressing the criticism that qualitative research lacks the rigour of quantitative methods (Abdalla et al., 2018; Morgan, 2024).

(iv) Flexibility and adaptability: Triangulation facilitates researchers in adapting their methods and approaches as the study progresses. This flexibility is particularly useful in qualitative research, where the research design often evolves in response to emerging findings (Vivek et al., 2023; Wood et al., 2020).

Subsequently, a framework is developed considering the emerging themes and policy analysis. It is validated through convergence of evidence. The resultant output is a risk-based procurement framework. In addition, a workflow is presented to guide the procurement practice in applying the framework. Fig. 2 illustrates the research process.



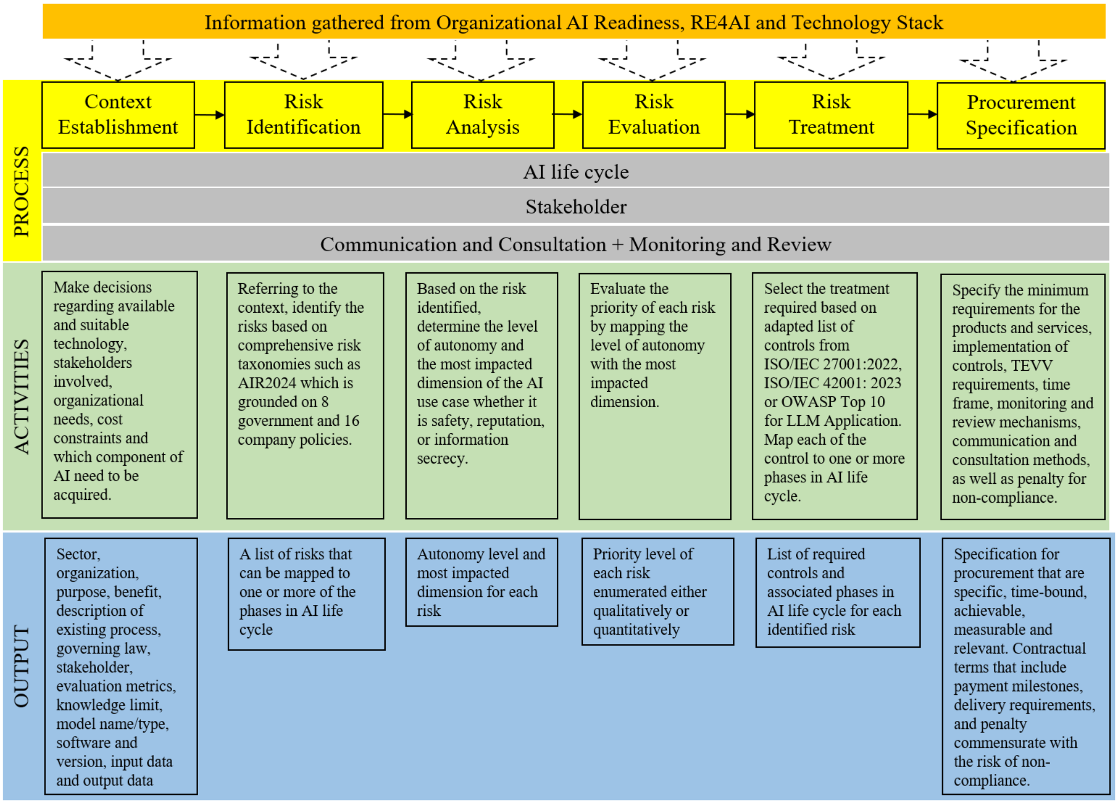
- Qualitative
- Interpretivist
- Semi-structured interview
- Document analysis
- Reflexive thematic analysis (inductive and deductive)
- Triangulation between interviews and documents
- Integration of emerging themes and policy analysis
- Validation through convergence of evidence
- Risk-based procurement framework
- Procurement workflow

Fig. 2. Research process

4. RESULTS

The framework was developed after formation and refinement of themes as well as triangulation between the two qualitative approaches. This structured approach addresses technical, ethical, and operational risks in AI procurement. It emphasizes collaboration, continuous monitoring, and readiness alignment, allowing organizations to navigate the complexities of AI adoption responsibly and effectively. In addition, the framework is composable as each process can refer to contemporary findings and requirements, given that AI is a rapidly developing field. For context establishment, the properties in Table 5 are required. For risk identification, interested readers can refer to the taxonomy developed by Zeng et al. (2024). Referring to Fig. 3, the framework incorporates elements from risk management, including context establishment, risk identification, risk assessment, and risk treatment before arriving at procurement specification. Also, the concurrent tasks of monitoring and review as well as communication and consultation are retained as required practices.

https://doi.org/10.24191/mij.v6i2.9172



Information gathered from Organizational AI Readiness, RE4AI and Technology Stack

<table>
<tr><td rowspan="4">PROCESS</td><td>Context Establishment</td><td>Risk Identification</td><td>Risk Analysis</td><td>Risk Evaluation</td><td>Risk Treatment</td><td>Procurement Specification</td></tr>
<tr><td colspan="6">AI life cycle</td></tr>
<tr><td colspan="6">Stakeholder</td></tr>
<tr><td colspan="6">Communication and Consultation + Monitoring and Review</td></tr>
<tr><td>ACTIVITIES</td><td>Make decisions regarding available and suitable technology, stakeholders involved, organizational needs, cost constraints and which component of AI need to be acquired.</td><td>Referring to the context, identify the risks based on comprehensive risk taxonomies such as AIR2024 which is grounded on 8 government and 16 company policies.</td><td>Based on the risk identified, determine the level of autonomy and the most impacted dimension of the AI use case whether it is safety, reputation, or information secrecy.</td><td>Evaluate the priority of each risk by mapping the level of autonomy with the most impacted dimension.</td><td>Select the treatment required based on adapted list of controls from ISO/IEC 27001:2022, ISO/IEC 42001: 2023 or OWASP Top 10 for LLM Application. Map each of the control to one or more phases in AI life cycle.</td><td>Specify the minimum requirements for the products and services, implementation of controls, TEVV requirements, time frame, monitoring and review mechanisms, communication and consultation methods, as well as penalty for non-compliance.</td></tr>
<tr><td>OUTPUT</td><td>Sector, organization, purpose, benefit, description of existing process, governing law, stakeholder, evaluation metrics, knowledge limit, model name/type, software and version, input data and output data</td><td>A list of risks that can be mapped to one or more of the phases in AI life cycle</td><td>Autonomy level and most impacted dimension for each risk</td><td>Priority level of each risk enumerated either qualitatively or quantitatively</td><td>List of required controls and associated phases in AI life cycle for each identified risk</td><td>Specification for procurement that are specific, time-bound, achievable, measurable and relevant. Contractual terms that include payment milestones, delivery requirements, and penalty commensurate with the risk of non-compliance.</td></tr>
</table>

Fig. 3. Risk-based Procurement Framework

Requirements engineering for AI (RE4AI), assessment of organizational maturity and examination of the technology stack may be implemented previously or specifically prior to the intended procurement exercise. These activities provide input required to subsequent processes. Then, the procurement team needs to delineate the precise scope of acquisition. This involves identifying which components—ranging from computational infrastructure to algorithmic services and human expertise—fall within the boundary of external sourcing. Such delineation must be informed by earlier insights into the technological environment and organizational capabilities, ensuring that the external procurement complements internal strengths rather than duplicating or misaligning them. In parallel, agencies must conduct a comprehensive role-mapping exercise in which internal personnel are temporarily assumed to perform all functions. This baseline mapping allows for a more granular understanding of institutional capacity, interdependencies, and potential gaps that may require outsourcing or capacity building.

To support the processes in risk management process related to AI, augmentations are necessary in the processes of context establishment, risk identification, risk analysis, risk evaluation and risk treatment. For context establishment, government agencies must consider the governing law for the sector, stakeholders involved in the acquisition and implementation, benchmarks, evaluation metric, knowledge limit, model name and type, architectural preference, and external integration on top of the common information associated with software acquisition. Table 5 elucidates the requirements for context establishment.

https://doi.org/10.24191/mij.v6i2.9172

Table 5. Requirements for context establishment

No.	Property	Description
1.	Sector	Industry in which the organization is operating.
2.	Organization name	Name of the organization.
3.	Purpose	Purpose of considering the use of AI.
4.	Benefit	The potential benefits using AI.
5.	Description of work process without AI adoption	Information on the existing workflow or method. This is required to gauge the necessity for AI adoption.
6.	Governing law	The legislation that governs the business area or the use of related equipment.
7.	Stakeholder	The internal and external stakeholders that may be affected by the use of AI.
8.	Minimum score for required benchmark (if any)	Benchmarks assist in gauging the intended outcome from the model such as reduction of hallucination, bias or translation accuracy.
9.	Configuration parameters (if relevant)	Specifying these parameters support optimization for tasks to be performed by the model.
10.	Evaluation metric	Types of available metrics used to measure the performance for acceptability of AI.
11.	Knowledge limit	The boundary or scope of information contained by the AI model.
12.	Model name/type	The name of the model and type of algorithm used for the model, such as XGBoost, Convolutional Neural Network (CNN), Recursive Neural Network (RNN).
13.	Architectural preference (if any)	This includes the use of vector databases, multimodal embeddings, RAG and fine-tuned model.
14.	External integration (if any)	Any integration with existing system or external providers via API.
15.	Software and version	The names and versions of tools, databases or orchestration framework for AI to run which would form a compatible list of software.
16.	Input data	The data to be entered or received by AI.
17.	Output data	The expected output for the AI.

For risk analysis, organizations need to consider the area of risk impact and define the severity level of each core aspect. Table 6 shows a sample of severity levels defined for risk impact in confidentiality, safety, and reputation. In this instance, five levels in increasing severity are defined for each area risk impact.

Table 6. Severity Level for Area of Risk Impact

Severity Level	Area of Risk Impact		
	Confidentiality	Safety	Reputation
Level 1	Status of information is ‘Open’	Potential to cause minor injury	Minimal damage to the reputation of the organization which do not require compensation
Level 2	Status of information is ‘Limited’	Potential to cause severe injury to less than 2 persons that do not require hospitalization	Minor damage to the reputation of the organization which require compensation
Level 3	Status of information is ‘Confidential’	Potential to cause severe injury to more than 2 persons that require hospitalization	Considerable damage to the reputation of the organization with potential lawsuits to be filed against the organization or affecting competitive advantage of the organization
Level 4	Status of information is ‘Minor Secret’	Potential to cause loss of lives to less than 2 persons	Major damage to the reputation of the organization such that would affect the main function of the organization
Level 5	Status of information is ‘Major Secret’	Potential to cause loss of lives to more than 2 persons	Catastrophic damage to the reputation of the organization such that the main client no longer trust the organization

https://doi.org/10.24191/mij.v6i2.9172

Consequently, risk evaluation for AI should consider the mapping between areas of risk impact and the automation level. In this regard, the risk level pertaining to the use of AI is dependent on its use case and can be represented by:

$$R = f(A, X), \text{ where } X = \begin{cases} P, & if\ reputation - driven \\ S, & if\ safety - driven \\ I, & if\ information - driven \end{cases}$$

The risk level (R) of an AI system or component is a function of automation level (A) with one of the areas of risk impact which may be reputation-driven (P), safety-driven (S) or information-driven (I). The number of levels for automation level should match the severity level of the risk area for enumeration of risk level. In this regard, Table 7 provides the description for each automation level (ISO/IEC, 2022). This is followed by Table 8 which illustrates the results of intersections between 'Confidentiality' as the area of risk impact and automation level. Consequently, Table 9 shows the range of values for the resultant risk level.

Table 7. AI Automation Level (ISO/IEC, 2022)

Level of Automation	Description
1-Assistance	The system assists an operator.
2-Partial automation	Some sub-functions of the system are fully automated while the system remains under the control of an external agent.
3-Conditional automation	Sustained and specific performance by a system, with an external agent being ready to take over when necessary.
4-High automation	The system performs parts of its mission without external intervention.
5-Full automation	The system is capable of performing its entire mission without external intervention.

Table 8. Risk level determination for information-driven use case

Impact Level		Confidentiality				
		Open (1)	Limited (2)	Confidential (3)	Minor Secret (4)	Major Secret (5)
Automation Level	Assistive (1)	very low (1)	very low (2)	very low (3)	very low (4)	low (5)
	Partial (2)	very low (2)	very low (4)	low (6)	low (8)	medium (10)
	Conditional (3)	very low (3)	low (6)	low (9)	medium (12)	high (15)
	High (4)	very low (4)	low (8)	medium (12)	high (16)	very high (20)
	Full (5)	low (5)	medium (10)	high (15)	very high (20)	very high (25)

Table 9. Risk level derivation from risk value

Risk Value (Automation * Confidentiality)	Risk Level	
1-4	very low	1
5-9	low	2
10-14	medium	3
15-19	high	4
20-25	very high	5

https://doi.org/10.24191/mij.v6i2.9172

Building upon the contextual groundwork, the procurement team must initiate a structured risk appraisal tailored to the AI domain. Drawing from established frameworks and national guidelines, the agency identifies potential sources of harm across the AI lifecycle. These may include algorithmic bias, model drift, misuse of autonomous features, or vulnerabilities to adversarial inputs. Risks are then prioritized to guide allocation of resources for established countermeasures as highlighted in established standards such as ISO/IEC 27001:2022 and ISO/IEC 42001: 2023. These requirements are translated into vendor obligations and stipulated in procurement documents. The options for risk treatment can be mapped to one or more phases in AI life cycle to facilitate project implementation and monitoring.

To further clarify the responsibilities of all parties, the role of stakeholders is delineated not only internally and externally but also between impacted parties and duty holders. This role division forms the foundation of the procurement specification, which must not only detail performance deliverables and timelines but also incorporate mechanisms for continuous evaluation, penalties for underperformance, and flexibility for mid-course corrections. The outcome is a procurement document that is technically robust, context-aware, and strategically aligned with the long-term governance objectives of public sector AI deployment. Fig. 4 shows the workflow based on the proposed framework that supports and complements existing procurement framework. The steps illustrated encompass all the procurement processes of planning, implementation, evaluation, monitoring and project implementation as well as contract management and administration, with emphasizes given to managing risks of AI.

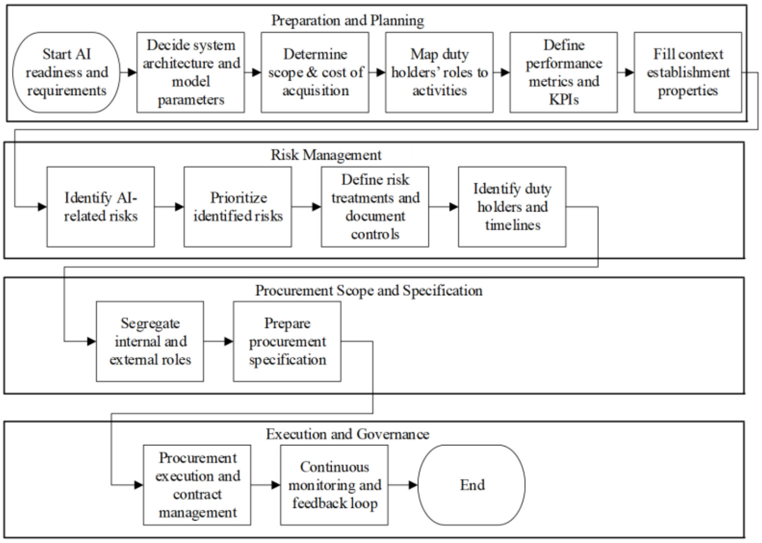

Preparation and Planning
Start AI readiness and requirements
Decide system architecture and model parameters
Determine scope & cost of acquisition
Map duty holders' roles to activities
Define performance metrics and KPIs
Fill context establishment properties
Risk Management
Identify AI-related risks
Prioritize identified risks
Define risk treatments and document controls
Identify duty holders and timelines
Procurement Scope and Specification
Segregate internal and external roles
Prepare procurement specification
Execution and Governance
Procurement execution and contract management
Continuous monitoring and feedback loop
End

Fig. 4. Procurement workflow in line with the proposed framework

5. DISCUSSION

This study developed a risk-based procurement framework that complements existing procurement and risk management processes practiced by Malaysian government agencies. The proposed framework provides a cohesive and structured mechanism for integrating AI-specific risk considerations into the aspects of technical, organizational, physical and people, like the existing practice in Information Security Management System (ISMS). However, it also blends the processes in risk management into each procurement phase, from planning to contract management. Moreover, it prepares government agencies in adopting the framework by assessing the organizational AI readiness, RE4AI, and required technology

https://doi.org/10.24191/mij.v6i2.9172

stack for successful AI implementation. This framework is modular in nature as the various processes can be adapted or updated as new discovery and development is made in AI.

While international standards such as ISO/IEC 27001:2022 and NIST AI RMF (2023) provide general guidance on information security and AI risk management, they do not specifically address the procurement dimension of AI adoption in the public sector. ISO/IEC 27001:2022 primarily focuses on ISMS and emphasizes protection of confidentiality, integrity, and availability of data. However, it does not provide mechanisms for integrating AI-specific risk assessment such as model drift, bias propagation, or adversarial vulnerabilities into procurement specifications. Furthermore, these standards remain voluntary and provide limited guidance on operationalizing risk controls during procurement, especially in government contracting contexts that require predefined deliverables and accountability mechanisms. Notably, the IEEE Standard for the Procurement of Artificial Intelligence and Automated Decision Systems (IEEE 3119:2025) addresses the various aspects of procurement but does not consider the current practices in Malaysian government agencies.

Hence, a mandatory mechanism in the form of regulation is established for the European Union with the ratification of the EU AI Act (2024). It adopts a risk-tiered approach that classifies AI systems into unacceptable, high, and low-risk categories. Although this approach establishes important regulatory baselines, it primarily functions as a compliance instrument rather than a procurement framework. While the act does not consider the existing controls established in an organization, it mandates conformity assessments and transparency obligations for high-risk AI systems. In addition, the act stipulates the creation of an AI office that is charged with coordination of enforcement, issuance of guidance and harmonization of implementation across member states. This framework is aligned with this requirement in that it provides guidance for procurement as well as specification of stakeholders' responsibilities in various phases of the AI lifecycle.

In short, embedding AI risk management principles into the procurement lifecycle transforms AI acquisition from a transactional process into a governance mechanism for responsible innovation. It enables agencies to demonstrate due diligence, align with global best practices, and maintain compliance with both information security and AI ethics requirements. The framework thus bridges the existing gap between technical risk management and public procurement governance, offering a scalable reference model for responsible AI implementation in the public sector.

6. CONCLUSION

Practically, the proposed framework offers government agencies a structured, context-sensitive tool for navigating the procurement of AI technologies, which are often characterized by opacity, rapid evolution, and ethical ambiguity. By prioritizing risk management, the framework enables agencies to make informed decisions that account for technical feasibility, vendor reliability, data governance, and societal impact. It supports procurement professionals in identifying and mitigating risks across the lifecycle of AI acquisition. The framework also has the potential to standardize procurement practices across agencies, enhance transparency, and foster public trust in AI deployment. Ultimately, it equips policy makers with a pragmatic instrument to align procurement decisions with strategic objectives and regulatory mandates in an era of digital governance. Future work will involve converting the framework to an online system with capabilities to validate any conflicting requirements with existing Malaysian regulations.

7. ACKNOWLEDGEMENTS/FUNDING

This research is sponsored by the Public Services Department of Malaysia.

https://doi.org/10.24191/mij.v6i2.9172

8. CONFLICT OF INTEREST STATEMENT

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

9. AUTHORS' CONTRIBUTIONS

David Lau Keat conceived the study and led the project administration. Ganthan Narayana Samy designed the methodology and supervised the technical aspects. Fiza Abdul Rahim coordinated data collection and curation. Mahiswaran Selvanathan conducted the analysis and prepared figures/tables. Nurazean Maarop contributed to policy review and interpretation of findings. Mugilraj Radha Krishnan verified the results and supported visualization. Sundresan Perumal reviewed the manuscript for important intellectual content. David Lau Keat drafted the manuscript; all authors revised it critically, approved the final version, and agree to be accountable for the work.

REFERENCES

Abdalla, M. M., Oliveira, L. G. L., Azevedo, C. E. F., & Gonzalez, R. K. (2018). Quality in qualitative organizational research: Types of triangulation as a methodological alternative. *Administração: Ensino e Pesquisa (RAEP), 19*(1), 66–98. https://doi.org/10.13058/raep.2018.v19n1.578

Awad, A. I., Shokry, M., Khalaf, A. A. M., & Abd-Ellah, M. K. (2023). Assessment of potential security risks in advanced metering infrastructure using the OCTAVE Allegro approach. *Computers and Electrical Engineering, 108*, 108667. https://doi.org/10.1016/j.compeleceng.2023.108667

Bernsmed, K., Cruzes, D. S., Jaatun, M. G., & Iovan, M. (2022). Adopting threat modelling in agile software development projects. *Journal of Systems and Software, 183*, 111090. https://doi.org/10.1016/j.jss.2021.111090

Bountakas, P., Zarras, A., Lekidis, A., & Xenakis, C. (2023). Defense strategies for adversarial machine learning: A survey. *Computer Science Review, 49*, 100573. https://doi.org/10.1016/j.cosrev.2023.100573

Carlton, E. L. (2014). Answering the call for integrating population health: Insights from health system executives. In *Population health management in health care organizations* (pp. 115–138). Emerald Group Publishing.

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum, 41*(5), 545–547. https://doi.org/10.1188/14.ONF.545-547

Chan, A., Wei, K., Huang, S., Rajkumar, N., Perrier, E., Lazar, S., Hadfield, G. K., & Anderljung, M. (2025). *Infrastructure for AI agents* (arXiv:2501.10114). arXiv. https://arxiv.org/abs/2501.10114

CNA (Center for Naval Analyses). (2025, May 1). *CNA's Artificial Intelligence (AI) Maturity Model for Government Agencies*. https://www.cna.org/analyses/2025/05/artificial-intelligence-maturity-model

https://doi.org/10.24191/mij.v6i2.9172

Deng, M., Wuyts, K., Scandariato, R., Preneel, B., & Joosen, W. (2011). A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering, 16*(1), 3–32. https://doi.org/10.1007/s00766-010-0115-7

Deng, Z., Guo, Y., Han, C., Ma, W., Xiong, J., Wen, S., & Xiang, Y. (2025). AI agents under threat: A survey of key security challenges and future pathways. *ACM Computing Surveys, 57*(7), 1–36. https://doi.org/10.1145/3716628

Desouza, K. C. (2021). *Artificial intelligence in the public sector: A maturity model*. IBM Center for The Business of Government.

Dreyling, R., Lemmik, J., Tammet, T., & Pappel, I. (2024). An artificial intelligence maturity model for the public sector: A design science approach. *TalTech Journal of European Studies, 14*(2), 217–239. https://doi.org/10.2478/bjes-2024-0023

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union, L* (12 July 2024). https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science, 7*(1), 549–569. https://doi.org/10.1007/s42001-024-00250-1

He, M. F., Schultz, B. D., & Schubert, W. H. (2015). *The SAGE guide to curriculum in education*. SAGE Publications.

IEEE Standards Association. (2025). *IEEE Std 3119-2025: Standard for the procurement of artificial intelligence and automated decision systems*. IEEE.

ISO. (2018). *ISO 31000:2018—Risk management—Guidelines*. International Organization for Standardization.

ISO/IEC. (2022). *ISO/IEC 22989:2022—Information technology—Artificial intelligence—Concepts and terminology*. International Organization for Standardization.

ISO/IEC. (2022). *ISO/IEC 27001:2022—Information security, cybersecurity and privacy protection—Information security management systems—Requirements*. International Organization for Standardization.

ISO/IEC. (2023). *ISO/IEC 42001:2023—Information technology—Artificial intelligence—Management system*. International Organization for Standardization. https://www.iso.org/standard/81230.html

Javaid, A. Y., Sun, W., Devabhaktuni, V. K., & Alam, M. (2012, November 13–15). Cyber security threat analysis and modeling of an unmanned aerial vehicle system. In *2012 IEEE Conference on Technologies for Homeland Security (HST)* (pp. 585–590). https://doi.org/10.1109/THS.2012.6459914

JPM. (2020). *Pelan Strategik Pendigitalan Sektor Awam 2021–2025*.

JPM. (2024a). *SPA Bil. 3 Tahun 2024—Garis Panduan Pengurusan Risiko Keselamatan Maklumat Sektor Awam*.

JPM. (2024b). *SPA Bil. 8 Tahun 2024—Garis Panduan Pengurusan dan Pengendalian Rahsia Rasmi dalam Perkhidmatan Awam*.

Kilian, K. A., Ventura, C. J., & Bailey, M. M. (2023). Examining the differential risk from high-level artificial intelligence and the question of control. *Futures, 151*, 103182. https://doi.org/10.1016/j.futures.2023.103182

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Masterman, T., Besen, S., Sawtell, M., & Chao, A. (2024). *The landscape of emerging AI agent architectures for reasoning, planning, and tool calling: A survey* (arXiv:2404.11584). arXiv. https://arxiv.org/abs/2404.11584

McIntosh, T. R., Susnjak, T., Liu, T., Watters, P., Liu, D., Nowrozy, R., & Halgamuge, M. N. (2024). From COBIT to ISO 42001: Evaluating cybersecurity frameworks for opportunities, risks, and regulatory compliance in commercializing large language models. *Computers & Security, 144*, 103964. https://doi.org/10.1016/j.cose.2024.103964

Meydan, C. H., & Akkaş, H. (2024). The role of triangulation in qualitative research: Converging perspectives. In *Principles of conducting qualitative research in multicultural settings* (pp. 98–129). IGI Global.

MOF. (2013). *Garis Panduan Mengenai Pengurusan Perolehan Information Telecommunication Technology (ICT) Kerajaan*. Kementerian Kewangan Malaysia.

MOF. (2022). *Perolehan Perkhidmatan Pengkomputeran Awan (Cloud) Sektor Awam*. https://ppp.treasury.gov.my/sub-topik/fail/221/muat-turun

Moghadasi, N., Valdez, R. S., Piran, M., Moghaddasi, N., Linkov, I., Polmateer, T. L., Loose, D. C., & Lambert, J. H. (2024). Risk analysis of artificial intelligence in medicine with a multilayer concept of system order. *Systems, 12*(2), 47. https://doi.org/10.3390/systems12020047

Morgan, H. (2024). Using triangulation and crystallization to make qualitative studies trustworthy and rigorous. *The Qualitative Report, 29*(7), 1844–1856.

MOSTI. (2024). *The National Guidelines on AI Governance and Ethics for Responsible and Inclusive AI*. https://mastic.mosti.gov.my/publication/the-national-guidelines-on-ai-governance-ethics/

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods, 22*, 16094069231205789. https://doi.org/10.1177/16094069231205789

Nagar, S. (2017). Introduction to Octave. In *Introduction to Octave: For engineers and scientists* (pp. 1–16). Springer. https://doi.org/10.1007/978-3-319-63754-6_1

https://doi.org/10.24191/mij.v6i2.9172

NDD. (2023a). *KRISA—Kejuruteraan Sistem Aplikasi Sektor Awam*. https://sqa.jdn.gov.my/index.php/ms/garis-panduan/garis-panduan-pembangunan-aplikasi-krisa

NDD. (2023b). *PPrISA—Garis Panduan Pengurusan Projek Sektor Awam 2.0*. https://sqa.jdn.gov.my/index.php/ms/pendahuluan

NDD. (2025). *Garis Panduan Pengadaptasian AI Sektor Awam*. https://www.jdn.gov.my/garis-panduan-pengadaptasian-ai-sektor-awam/

NIST. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

NIST. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., & Lee, V.-H. (2025). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems, 65*(1), 76–107. https://doi.org/10.1080/08874417.2023.2261010

Ouaissa, M., & Ouaissa, M. (2025). Analyzing and mitigating attacks in IoT smart home using a threat modeling approach-based STRIDE. *International Journal of Interactive Mobile Technologies (iJIM), 19*(2). https://doi.org/10.3991/ijim.v19i02.52377

OWASP GenAI Security Project. (2025a). *Agentic AI—Threats and mitigations*. https://genaisecurityproject.com/resource/agentic-ai-threats-and-mitigations

OWASP GenAI Security Project. (2025b). *LLM and GenAI security solutions landscape—Q1 2025*. https://genaisecurityproject.com/resource/llm-and-generative-ai-security-solutions-landscape-q12025/

Pape, N., & Mansour, C. (2024). PASTA threat modeling for vehicular networks security. In *2024 7th International Conference on Information and Computer Technologies (ICICT)* (pp. 474–478). https://doi.org/10.1109/ICICT62343.2024.00083

Portugal, I. D. S., Alencar, P., & Cowan, D. (2024). An agentic AI-based multi-agent framework for recommender systems. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 5375–5382). https://doi.org/10.1109/BigData62323.2024.10825765

Qi, X., Zeng, Y., Xie, T., Chen, P.-Y., Jia, R., Mittal, P., & Henderson, P. (2023). *Fine-tuning aligned language models compromises safety, even when users do not intend to!* (arXiv:2310.03693). arXiv. https://arxiv.org/abs/2310.03693

Responsible AI Collaborative. (2024, July 8). *AI Incident Database (AIID): June 2024 incident roundup*. AI Incident Database Blog. https://incidentdatabase.ai/blog/incident-report-2024-june

Saitta, P., Larcom, B., & Eddington, M. (2005). *Trike v1 methodology document* [Draft white paper].

Shavit, Y., Agarwal, S., Brundage, M., Adler, S., O'Keefe, C., Campbell, R., Lee, T., Mishkin, P., Eloundou, T., & Hickey, A. (2023). *Practices for governing agentic AI systems*. OpenAI. https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf

Shostack, A. (2014). *Threat modeling: Designing for security*. Wiley.

Slattery, P., Saeri, A. K., Grundy, E. A., Graham, J., Noetel, M., Uuk, R., Dao, J., Pour, S., Casper, S., & Thompson, N. (2024). *The AI Risk Repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence* (arXiv:2408.12622). arXiv. https://arxiv.org/abs/2408.12622

Turri, V., & Dzombak, R. (2023). Why we need to know more: Exploring the state of AI incident documentation practices. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)* (pp. 576–583). https://doi.org/10.1145/3600211.3604700

UcedaVelez, T., & Morana, M. M. (2015). *Risk centric threat modeling: Process for attack simulation and threat analysis*. Wiley.

Vivek, R., Nanthagopan, Y., & Piriyatharshan, S. (2023). Beyond methods: Theoretical underpinnings of triangulation in qualitative and multi-method studies. *Southeast European University Review, 18*(2), 137–149. https://doi.org/10.2478/seeur-2023-0088

Willems, W. J. M. (2025). *From principles to practice: A tailored maturity model for responsible AI in public sector organisations* (Master's thesis). University of Twente.

Wood, L. M., Sebar, B., & Vecchio, N. (2020). Application of rigour and credibility in qualitative document analysis: Lessons learnt from a case study. *The Qualitative Report, 25*(2), 456–470.

Zeng, Y., Klyman, K., Zhou, A., Yang, Y., Pan, M., Jia, R., Song, D., Liang, P., & Li, B. (2024). *AI risk categorization decoded (AIR 2024): From government regulations to corporate policies* (arXiv:2406.17864). arXiv. https://arxiv.org/abs/2406.17864

https://doi.org/10.24191/mij.v6i2.9172